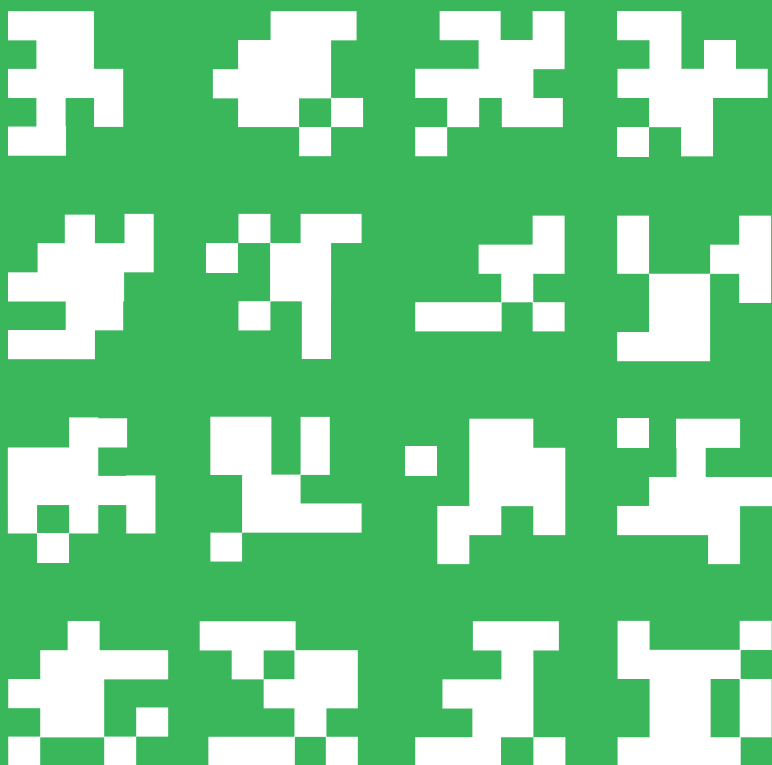


navigating the
digital age

an in-depth
exploration
into the
intersection
of modern
technologies
and societal
transformation

edited by:
ljubiša bojić
simona žikić
jörg matthes
damian trilling



Navigating the Digital Age

Book Series

DISPUT

Reviewers

**Kevin LaGrandeur, Full Professor, New York Institute of
Technology, New York**

**Robert Braun, Senior Research Fellow, Institute for Advanced
Studies, Vienna**

**Boris Delibašić, Full Professor, Faculty of Organizational
Sciences, University of Belgrade, Belgrade**

**Navigating the
Digital Age**

**An In-Depth
Exploration into
the Intersection
of Modern
Technologies
and Societal
Transformation**

**Edited by:
Ljubiša Bojić
Simona Žikić
Jörg Matthes and
Damian Trilling**



**INSTITUTE
FOR PHILOSOPHY
AND SOCIAL
THEORY**



**universität
wien**

Department of Communication

Navigating the Digital Age:

An In-Depth Exploration into the Intersection of Modern Technologies and Societal Transformation

Edited by

Ljubiša Bojić, Senior Research Fellow, Institute for Philosophy and Social Theory, Digital Society Lab, University of Belgrade; The Institute for Artificial Intelligence Research and Development of Serbia, Novi Sad ORCID: 0000-0002-5371-7975

Simona Žikić, Research Fellow/Associate Professor, Institute for Philosophy and Social Theory, University of Belgrade; Faculty for Media and Communications, Singidunum University, Belgrade ORCID: 0000-0001-6093-1665

Jörg Matthes, Full Professor, Faculty of Social Sciences, University of Vienna ORCID: 0000-0001-9408-955X

Damian Trilling, Full Professor, Faculty of Humanities, Vrije Universiteit Amsterdam ORCID: 0000-0002-2586-0352

Published by

Institute for Philosophy and Social Theory, University of Belgrade

Kraljice Natalije 45, Belgrade

Department of Communication, Faculty of Social Sciences, University of Vienna

Währinger Straße 29; Kolingasse 14-16, Vienna

For publishers

Gazela Pudar Draško, Director

Sabine Einwiller, Department Chair

Design and typesetting

Nadežda Todorović and Tijana Radovanović Marković (JedninaDuo)

Print run

300

Printed by

Alta Nova, Beograd

ISBN 978-86-82324-85-0

2024

This edited book was realised with the support of the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, according to the Agreement on the realisation and financing of scientific research 451-03-66/2024-03/200025.

This publication has been accomplished with the support and collaboration from the COST Action Network CA21129—What are Opinions? Integrating Theory and Methods for Automatically Analyzing Opinionated Communication (OPINION)— <https://www.opinion-network.eu/>, and supported by the TWON (project number 101095095), a research project funded by the European Union, under the Horizon Europe framework (HORIZON-CL2-2022-DEMOCRACY-01, topic 07). More details about the project can be found on its official website: <https://www.twon-project.eu/>.

Contents

Preface

Ljubiša Bojić, Simona Žikić, Jörg Matthes
and Damian Trilling _____ 11

PART I: COMMUNICATION AND TECHNOLOGY

Blockchain and Journalism: Opportunities
and Challenges Ahead
Walid Al-Saqaf _____ 21

Innovating in Journalism with Newsgames:
An Exploratory Study in Portugal
Bruno Frutuoso Costa and Bruno Contreiras Mateus _____ 43

Digital Technologies, Individualisation,
and the Rise of Digital Populism
Dominic Spada _____ 83

Emotional Dynamics in Online Videos:
Integrating Machine Learning and Content Analysis
Sara Major and Aleksandar Tomašević _____ 128

The Power of Images: Visual Framing of the Ukraine Conflict on Instagram	
Gheorghe Anghel	153

PART II: THE DIGITAL TRANSFORMATION OF SOCIETY

True Investment: Post-Authentic Honesty on Social Media	
Hans-Georg Moeller	199

Fill In, Accept, Submit, and Prove that You Are not a Robot: Ubiquity as the Power of the Algorithmic Bureaucracy	
Mikhail Bukhtoyarov and Anna Bukhtoyarova	220

A Balanced Approach to Artificial Intelligence in Government Decision-Making Processes	
Aleksandar Jazić and Ljubiša Bojić	244

The Business of Umbilical Cord Biobanking in Serbia: Biocapital(ism) and Symbolic Geographies of Health	
Ljiljana Pantović and Jelena Čeriman	267

PART III:
DEMOCRACY, SECURITY AND TECHNOLOGY

**Standardization for High Impact Technologies: Why
Citizens Should Care**

Claudia Roda, Rebecca Stone and Antonio Kung _____ 305

**Technology-Facilitated Gender-Based Violence and
the Novel: Realizing Women's Rights**

Sekai Zhou _____ 347

**Cybersecurity Culture from the Perspective of Social
Cognitive Theory: Examining the Relevance of the
Reciprocal Determinism Model**

Sladana Ćurčić _____ 377

**Unsettling Future of AI-Powered Police-Human
Interaction: A Shadow of a Doubt over Moral Sensitivity**

Srdan T. Korać _____ 407

**Strategies for Cybersecurity Innovation and Crisis
Communication in Digital Democracy Era**

Mokhtarrudin Ahmad and Kavitha Balakrishnan _____ 455

PART IV:
ART AND NEW MEDIA

The Mechanical Turkness: Tactical Media Art and the Critique of Corporate AI	
Dejan Grba	509
Desktop Films: Posthuman Gaze and Technogenesis in Contemporary Cinema	
Jelena Mišeljić	540
Just Look like Us, or Look just like Us: The Measure of (Un)canniness	
Predrag Krstić	562
Art after AI: The Impact of Generative AI on the Artworld	
Jelena Novaković and Jelena Guga	584

Navigating the Digital Age explores how modern technologies impact our societies and our lives. Throughout history, technologies and media, from Roman roads to today's 5G and nanotechnologies, have influenced our society and reshaped our behavior and lifestyle. Compiled by a group of international academic authors, this book describes numerous ways technology is reshaping our world, the environments, the communities we belong to, and our understanding of them. It is hard to imagine a technology that only affects society/culture, or a society/culture that only affects technology. Rather, we observe a reciprocal process: On the one hand, technology develops from the needs of society, while on the other hand, technology changes society.

The world we live in is a world of communications and media. Media culture, hyperproduction, and the hypersphere prevail. Divided into four distinctive parts, the book re-examines the spheres of communication, the digital transformation of society, the intersections of democracy, security, and technology, and finally, the connection between art and new media. From the impact of artificial intelligence in journalism to the role of blockchain in government decision-making, from the perspective of cyber security culture to the implications of art generated by artificial intelligence—this book covers a wide landscape of new media.

Navigating the Digital Age encourages a rigorous discourse on the interactions between technology and society. It provides a mediological approach and interpretation of new technology in society; it contemplates not only new media and technologies but also how new media and technology affect society and the communities we belong to. It is not just a reflection on our current circumstances but also a projection of future possibilities in our evolving digital ecosystem.

This preface serves as a gateway into the thematic and intellectual realm that this edited book seeks to explore. It lays the groundwork for the subsequent exploration of the intricate and multifaceted relationship between technology, society, and the individual. This book, through its diverse contributions, aims to offer a panoramic view of the current digital age, characterized by rapid technological advancements and their interaction with human life and societal structures. At its core, this book is an invitation to engage with critical perspectives on the digital transformation that permeates every aspect of our lives. It challenges readers to consider not only the technological but also the ethical, cultural, and social dimensions of this transformation. By bringing together a wide array of voices and insights from various disciplines, the book endeavors to foster a holistic understanding of the digital landscape that shapes our contemporary world. We would like to emphasize the importance of interdisciplinary dialogue in addressing the complex challenges and opportunities presented by the digital age. In this preface, we aim at setting the stage for a journey that navigates through themes of communication, identity, privacy, and the evolving dynamics of power and resistance within digital spaces. We invite the reader to embark on a reflective and insightful exploration of the digital age, and hope that this book encourages a deep and nuanced engagement with the questions and debates that

define our time, highlighting the significance of understanding technology not merely as a tool but as a fundamental force that shapes our reality. Through this exploration, the book aspires to contribute meaningfully to the ongoing discourse on the role of technology in society, offering perspectives that enlighten, challenge, and inspire.

Information and data in the online ecosystem are new driving forces of humanity in the 21st century, sometimes even called the new oil and gold of today. In the contemporary, fluid digital age that is rapidly developing, current technologies are advancing at a great pace, affecting our ability to understand, assimilate, and implement technological advancements. Along with this surge in innovation comes a significant shift in society, as we grapple with the implications of these technological interventions on our normative structures. The development and transformation of new technologies and media do not occur in the same way across all geographical and cultural spaces. We, as editors, have attempted to connect different perspectives on this transformation from a broad spectrum of authors covering three continents and ten countries.

Divided into four different, yet closely related parts, this book explores multiple angles and perspectives on the complex interaction between technology and society, creating a synergy of communication, digital transformation, democracy and security, art, and new media.

The first part, titled “Communication and Technology”, explores how technologies, especially blockchain, have the potential to reshape the world of journalism and the creation of digital content. Humans are primarily social beings, and communication is a practice without which an individual cannot sustain themselves. Communication regulates social relations. With the digital transformation and

the development of the fourth industrial revolution, civilization has gained a galaxy of accessible channels and tools through which people can communicate. The digital era has changed the way information is distributed and consumed, opening the door to new models of journalism and media production. The further development of the fourth industrial revolution will continue to fluidly change the use of new technologies in many aspects. The first part of the book unites discussions of communication approaches about new trends such as news-games, digital populism, emotional dynamics in online videos, and the power of visual framing in conflicts.

The second part deals with the “Digital Transformation of Society”. It is hard to imagine the world today without the internet and digitalization. The contribution of internet technology to the development of humanity is immeasurable and rightly stands as a leading achievement of the human mind. The chapters in this part focus on the changes brought about by the internet and digitalization across different sectors of society and all aspects of human life. New information technology surpasses Gutenberg’s galaxy and the revolution in the accessibility of the written word, so today we cannot compare Gutenberg’s galaxy with the era of the present, which marks the 21st century with the emergence of globally available, attainable, received, achievable, and accessible digital means of information, the volume of Big Data, AI, VR, Blockchain. From fascinating interactions of online sincerity on social media to the intricate dynamics of power in the age of algorithmic bureaucracy, each chapter offers thought-provoking insights into the digital revolution.

This section also addresses the implementation of advanced technologies such as artificial intelligence in government decision-making processes and explores new areas of bio-capital such as cord blood

banking. Huge amounts of data stored in digital spaces are analyzed daily, and groups of government officials analyze the statistics of the same to make the right decisions on many topics and challenges the world faces.

In the third part, we delve into the deep and challenging themes of “Democracy, Security, and Technology”. The emergence of the internet was presented as an optimistic promise and belief that it would finally become a truly democratic medium and tool. With the enormous amount of information accessible to everyone at the speed of light and resistance to censorship, it symbolized a true democratic tool for communication and awareness. Unlike traditional media and technology, where the audience is mainly a passive consumer and observer, today’s media users can increasingly be seen as not only active information seekers, but also creators and disseminators. This part focuses on how technology relates to gender rights, cyber security culture, and interactions between the police and people driven by artificial intelligence. Issues of privacy, data security, and ethics become key in this context, as technological tools become more integrated into everyday life. The book also considers how technological advancements can impact social justice and equality, especially in the context of access to technology and digital division. In an era of rapid technological progress and fluidity, it is important that democratic processes keep pace with and respond to changes in the technological landscape. The chapter calls for social engagement in the high-impact area of technological standardization, clarifying the importance of citizen participation in shaping the future of technology.

Finally, in Part IV, “Art and New Media”, we explore the intersection of emerging technology with artistic expression. Desktop movies,

corporate AI critique through tactical media art, the (un)scary impact of artificial intelligence on art, and the emergence of AI generators are some of the avant-garde topics examined that reveal new avenues for creative expression made possible through advanced technologies. The use of artificial intelligence in art represents a broad and fascinating field of exploration. Generative AI is increasingly used to create artworks, from paintings and music to literature and performances, from commercial advertisements to film scenes. This part investigates how AI is changing the landscape of art, how it affects the perception and interpretation of artworks, and how it shapes the very practice of creating art with the use of new tools and media.

New technologies and media are the leading phenomena of today and play a significant role in modern society by providing, in various ways, a wide range of information and communication channels. Through this journey composed in eighteen chapters, the book invites both readers and scholars to delve into the discourse around the emergence and potential of new technology and its significant impact on society. Guided by McLuhan's thesis on media and technology as extensions of the human body, new technologies represent much more than the extension that was in the 20th century. The use of AI technology in all aspects of the human environment is rapidly advancing, and we as a society must keep up with the transformation. The insightful perspectives embedded in these pages aim to stimulate critical thinking and foster an understanding of the technological landscape and its prospects. Accordingly, it is crucial for society to be educated, to develop digital and algorithmic literacy, and to analytically and critically approach and keep up with current and upcoming technological transformations.

History has shown that it is in human nature to strive for better, to evolve, and this precisely leads to the expansion of the development of both science and technology. However, as with everything, every question, challenge, or stance can be viewed from multiple different perspectives. We live in a world of binary oppositions, and surely one group of people will be against the development of technology and further advancement of artificial intelligence and information technologies, while another group will advocate for the further development of technology. As technology continues to develop at an unprecedented pace, society is somewhat wandering into unknown spaces and territories of transformation. In this new fluid-virtual reality, understanding the intersection between new technologies and society becomes crucial. We hope this book will contribute to enlightening the discourse around this critical crossroads. Let's think together and be part of the civilizational transformation!

Part I

Communication and Technology

Blockchain and Journalism: Opportunities and Challenges Ahead

Walid Al-Saqaf

Introduction

The early days of the internet in the 1990s were met with skepticism. Clifford Stoll, an astronomer, was a notable critic who dismissed the transformative potential of the internet in a *Newsweek* editorial (Stoll, 1995). Today, as we reflect on the evolution of the internet and its monumental impact on various aspects of our lives, including journalism, it's clear that Stoll's prediction was off the mark (O'Sullivan, Heinonen, 2008). However, Stoll was correct in cautioning against the spread of disinformation, a problem that has become prevalent in the era of social media (Sanders et al., 2000). His foresight illustrates the challenges of predicting the trajectory and societal impact of emerging technologies.

My own experience as a computer engineering student in the early 1990s involved experimenting with the then-novel internet. I recall the fascination of browsing files on the World Wide Web (WWW) using

Walid Al-Saqaf, Associate Professor, Doha Institute for Graduate Studies, Doha
ORCID 0000-0002-1993-5696, email: walid.alsaqaf@dohainstitute.edu.qa

DOI: <https://doi.org/10.5281/zenodo.14199822>

a graphical interface for the first time. As time passed, the growth and evolution of technology—from computers to mobile phones—have been remarkable. The example of smartphones is particularly illustrative. What started as bulky devices akin to “bricks” has evolved into globally ubiquitous tools that have spurred the emergence of ‘mobile journalism’ (Perreault, Stanfield, 2019). This narrative underscores the journey that most innovations undergo, from their inception to mass adoption (Rogers, Singhal, Quinlan, 2008). Predicting the precise course and impact of such technologies on our lives, particularly in fields such as journalism, remains a complex task.

Blockchain, a financial technology (fintech) innovation, is currently navigating its unique evolutionary path. Despite facing challenges ranging from high energy consumption to market manipulation and misuse by cybercriminals, blockchain holds the potential to bring about significant change (Valdeolmillos et al., 2020). However, it is essential to avoid premature and unfounded hype surrounding blockchain as the definitive “savior” of journalism; all technologies can reflect the biases and belief systems of their creators (Teng, 2022). Moreover, blockchain may evolve into something drastically different from what it is today. Some envision its true potential in integrating with the existing web, marking the advent of Web3—a system that leverages blockchains, cryptocurrency, and non-fungible tokens (NFTs) to empower end users by emphasizing data ownership and privacy (Voshmgir, 2020). Cryptocurrency markets, including the NFT market, have experienced volatile swings in recent years; at one point, the NFT market capitalization fell by an unprecedented 97%, from USD 17 billion in January 2022 to USD 466 million in September of the same year (Shukla, 2022).

While these fluctuations have drawn attention to the monetary aspects of blockchain, it's important to consider the non-fintech dimensions of this technology and its potential societal impact. This chapter aims to highlight the social implications of blockchain, particularly in relation to digital journalism. The goal is to inspire journalism scholars and practitioners to explore the relevance of blockchain for journalism and to collaborate across disciplines in this complex, technical field.

Blockchain as an Emerging Technology

The inception of blockchain technology, characterized as a series of transaction blocks linked together, was first proposed in a 2008 whitepaper titled “Bitcoin: A Peer-to-Peer Electronic Cash System,” authored by the pseudonymous Satoshi Nakamoto (Nakamoto, 2008). The impetus behind this innovation was the absence of a decentralized mechanism for internet users to transfer currency without relying on traditional third-party financial institutions. Bitcoin marked the first globally tested, reliable method for enabling the direct exchange of virtual money among internet users, bypassing the need for intermediary trust.

This intent is encapsulated in the message embedded in the genesis block of the Bitcoin blockchain by Satoshi Nakamoto: “The Times 03/ Jan/2009 Chancellor on brink of second bailout for banks.” This refers to a Times article detailing the UK Chancellor of the Exchequer’s intent to rescue failing banks by infusing “billions more into the economy” (Elliott, 2009). The underlying notion here is that reliance on central

authorities, such as banks, governments, or individuals, inherently involves trust, as citizens authorize these entities to access their funds.

In the wake of the 2008 financial crisis, public trust in centralized systems and institutions plummeted, leading to increased interest in cryptocurrencies (Ito et al., 2017). The crisis, regarded by some economists as the severest since the Great Depression, shook people's faith in financial institutions (Uslaner, 2010). In the aftermath of the global Covid-19 recession, inflation-stricken countries like Venezuela (Ellsworth, 2021) and Lebanon (Azhari, 2021) witnessed their citizens adopt cryptocurrency as a reliable means of receiving payments and as a store of value. Similarly, during the 2022 Russian-Ukrainian conflict, when financial services in Ukraine were partially or entirely suspended, Bitcoin and other cryptocurrencies emerged as effective alternatives for receiving donations outside the traditional financial system (Criddle, Oliver, 2022).

Another significant advantage of cryptocurrencies is their potential to extend banking services to the unbanked billions in developing nations (Larios-Hernández, 2017: 865–874). The internet's decentralized nature, which underpins blockchain, allows it to easily reach a vast number of users without any central authority's control, censorship, or the potential for destruction on a global scale. This characteristic is also inherent to blockchain-based services, which can adapt to varying conditions, allowing users to continue accessing their virtual funds via proxies or Virtual Private Network (VPN) services if direct access becomes limited.

In essence, blockchains are financial, distributed, immutable, open, and programmable database systems that distinguish them from previous technologies. Bitcoin was the first practical implementation of

blockchain technology, followed by several other blockchains offering various enhancements, such as decentralized applications (dApps) and custom cryptocurrencies (tokens), fostering the rise of decentralized finance (DeFi) and non-fungible tokens (NFTs) (Christodoulou et al., 2022). A notable example is Ethereum, the second-largest blockchain, which launched its more energy-efficient and faster Ethereum 2.0 in 2022 (Arslanian, 2022: 91-98). However, this chapter will not delve into the specifics of any particular blockchain but will instead broadly discuss the technology's relevance and potential applications within the journalism industry.

Blockchains are fundamentally financial, as Bitcoin's initial use case was a secure peer-to-peer money transfer system. The value of a native cryptocurrency unit is typically subject to traditional market dynamics of supply and demand, which often lead to significant volatility due to the largely unregulated nature of the space. This volatility has resulted in the creation of stablecoins like Tether (USDT), which are pegged to stable assets such as the USD, to mitigate such erratic market swings (Bullmann, Klemm, Pinna, 2019). While the fintech aspect of blockchain is a potent revenue magnet, especially if DeFi becomes more pervasive, it also carries substantial risk. This risk arises from the potential compromise of private keys of a crypto wallet or hacking attacks against custodial wallet services on centralized exchanges, as experienced by FTX in 2022 (Kharpal, 2022) and Mt. Gox in 2013 (Feder et al., 2017).

Blockchain's distributed nature ensures that its data is spread across numerous nodes on the internet, preventing any single entity from controlling or manipulating the system. Instead of central financial bodies that traditionally maintained ledgers, a blockchain system stores exact ledger replicas across all nodes. These nodes follow a

well-established and automatically executed set of instructions to reach consensus on which transaction blocks to validate. Thus, blockchain, a distributed ledger technology (DLT), is a digital system that records transactions synchronously using a consensus mechanism. In a public proof-of-work blockchain like Bitcoin, any new node can join without permission, incentivized by a built-in reward system that pays miners transaction fees and mining rewards for any specific block they mine. The more nodes a blockchain has, the more secure it becomes.

Blockchains are also characterized by their immutability. The structure of the blockchain, wherein transactions are chained into blocks that depend on preceding valid blocks, makes altering content stored in any block virtually impossible without changing all subsequent blocks. Therefore, data written on the blockchain is generally considered permanent and stored indefinitely. This permanence is crucial for the security and usability of blockchains, particularly for payments, and the level of security is positively correlated with the total number of participating distributed nodes.

Openness is another defining feature of blockchains. It necessitates open-source code, standards, and access, allowing anyone to read all transactions and analyze their connections to verify provenance by tracing any transaction back to its origin. This openness fosters interoperability, enabling the creation of open standards, preserving transparency, and allowing new nodes to participate in transaction block validation. The predictable and reliable structure of blockchains also allows anyone with basic technical data analysis skills to extract and analyze blockchain data with relative ease.

Lastly, blockchains are programmable, thanks to a key component known as smart contracts. These allow for the creation of executable

code that can be activated by predefined triggers, such as the sending or receiving of a transaction or any other automated or manual event, effectively creating programmable money. This feature is utilized in dApps and serves as a crucial pillar in the formation of decentralized autonomous organizations (DAOs), which operate not through any central governing authority but through a democratically predefined set of rules encoded in the blockchain. This feature, unique to blockchains, significantly reduces the human element in decision-making in favor of algorithmic instructions that are tamper-proof. It opens up numerous opportunities for enforcing contracts automatically without the need for intermediaries, potentially enhancing accountability and transparency (Al-Saqaf, Seidler, 2017).

Blockchain's ability to facilitate peer-to-peer value exchange represents a significant advancement in building the Web3 framework, potentially signaling a return to the original vision of a truly decentralized internet. However, for wide-scale adoption, it needs to prove its utility against existing solutions. This can be achieved if all the necessary pieces, including identity verification, security, privacy, and incentives for different stakeholders, fall into place.

Exploring Blockchain Applications in Journalism

In the early 1990s, the internet and mobile phones bore little connection to journalism. However, as these technologies became ubiquitous, they profoundly influenced the journalism landscape. Similarly,

although blockchain might seem unrelated to journalism at present, with time and imaginative thinking, it could revolutionize the field in unprecedented ways (Al-Saqaf and Seidler, 2017; Ito et al., 2017; Saini, 2021; Samaniego et al., 2016). By harnessing various blockchain features, we can imagine at least five potential applications for journalism: 1) Provenance Verification, 2) Investigative Journalism, 3) Content Monetization and Dissemination, 4) Intellectual Property and Archiving, and 5) Freedom of Expression. Each area is ripe for exploration, with several relevant projects already underway.

Provenance Verification

Blockchain's immutability lends itself to storing original, authentic content from news media institutions. This unalterable storage ensures the authenticity of the content, making blockchain an appealing tool to combat disinformation. Initiatives like the News Provenance Project, a collaboration between the New York Times and IBM, utilize the Hyperledger Fabric open-source blockchain technology for this purpose (DiCicco, Agarwal, 2020). Similarly, Project Origin, a partnership between Microsoft and three media publishers (BBC, CBC, and New York Times), and Cogency Lab are harnessing the power of permissioned blockchains to authenticate media content.

In addition, blockchain could facilitate alternative revenue streams such as cryptocurrency micropayments, thus bolstering newsrooms' financial stability and reducing susceptibility to external influences. The now-defunct Civil project employed this approach, attempting to combat disinformation by improving journalistic content quality using

a democratic, Ethereum-based, multi-layered governance model. Other initiatives like DNN, Trive, Defudger, and Primas are also attempting to leverage blockchain to fight disinformation (Al-Saqaf, Picha, 2019). Beyond news media, companies like Truepic are providing resources for journalists, using the Bitcoin blockchain to verify photos and videos and combat deepfakes (Hasan, Salah, 2019).

Investigative Journalism

Blockchain's traceability can be a goldmine for data journalists, offering insights directly tied to news events. Companies like Chainalysis specialize in blockchain analytics, providing real-time cryptocurrency intelligence for advanced investigations. Journalists can also independently tap into this resource, as demonstrated by NYT reporter Keith Collins, who tracked ransomware victims' payments in 2017 (Roush, 2017). Blockchain's openness enables anyone with technical skills to access and analyze a wealth of accurate transaction data.

However, Bitcoin's pseudonymous nature means that while transaction data is readily available, linking it to individuals or entities requires additional investigative skills (Kamphausen, Werse, 2019). This was highlighted when Twitter users used open-access blockchain data to expose the FTX crypto trading and exchange platform's collapse in 2022 (Sigalos, 2022).

Content Monetization and Distribution

The combination of blockchain's financial and other attributes opens up new possibilities for content distribution and monetization. For example, blockchain could enable real-time micropayments between readers and news content providers. Platforms like Steemit incentivize users to publish, evaluate, and rate content through such micropayments, although the revenue generated is lower than that of central social media platforms (Thelwall, 2018: 736–751). Using blockchain rather than a centralized system could offer news content providers and readers enhanced security and reduced risk of data misuse, given blockchain's transparency and accountability. Furthermore, with blockchain, readers maintain full rights over their data and can customize their privacy preferences via smart contracts.

Intellectual Property and Archiving

Blockchain's immutability ensures the preservation and traceability of content, making it an ideal tool for archiving and intellectual property protection. Once information is stored on the blockchain, it cannot be altered or deleted, thus safeguarding the original work of journalists and other content creators. This feature can be invaluable for copyright protection, enabling content creators to prove their work's originality and publication date. Several projects are leveraging blockchain for content archiving and intellectual property

protection. For instance, **Po.et** is a decentralized protocol for content ownership, discovery, and monetization in media built on the Bitcoin blockchain. It allows creators to timestamp their works onto the Bitcoin blockchain, thereby ensuring the immutability and provability of the content.

Freedom of Expression

Blockchain can also potentially impact freedom of expression. By creating decentralized platforms for content distribution, blockchain can reduce censorship and promote freedom of speech. As blockchain is inherently resistant to censorship, it can be used to build platforms where content can be published freely without interference from third parties. Blockchain's transparency can serve as a mechanism for holding news organizations accountable, enhancing trust and credibility in journalism. Given the public nature of blockchain transactions, news organizations can use this technology to openly share their funding sources, editorial decisions, and other operational aspects. This could promote transparency and build trust with their audience.

As this section demonstrates, blockchain technology offers promising opportunities for reshaping the journalism industry. As with any disruptive technology, the ultimate impact will depend on how it is adopted and implemented. However, by providing a new level of transparency, immutability, and decentralization, blockchain could potentially revolutionize journalism, making it more trustworthy, secure, and resistant to manipulation. However, one should remember

that the technology remains in its early stages of development and therefore, several challenges need to be overcome before its full benefits could be reaped as described in the next section.

Challenges to Overcome

Despite the potential benefits of blockchain technology in journalism, there are several significant challenges that must be addressed before this potential can be fully realized. The following section highlights these key challenges, including technological complexity, scalability issues, regulatory hurdles, privacy concerns, and the need for widespread adoption.

Technological Complexity

Perhaps the most immediate challenge facing the integration of blockchain technology in journalism is its inherent complexity (Tapscott, Tapscott, 2018). The implementation of blockchain systems requires a deep understanding of distributed ledger technology, cryptography, and other complex technical concepts. This poses a challenge for journalists and newsrooms that may not have the necessary technical expertise. Moreover, the development of blockchain-based solutions for journalism also requires substantial resources, including time and funding, which many newsrooms may not have.

Even when newsrooms manage to implement blockchain technology, the complexity of these systems may also pose challenges for users. For example, engaging with blockchain-based platforms often requires users to manage cryptographic keys, which can be a daunting task for those unfamiliar with the technology (Zuboff, 2019). Therefore, there is a need for more user-friendly interfaces and simplified procedures to encourage broader participation in blockchain-based journalism platforms.

| Scalability Issues

Scalability is another significant challenge for the use of blockchain technology in journalism (Böhme et al., 2015). Current blockchain systems, such as Bitcoin and Ethereum, can process only a limited number of transactions per second. This limitation could potentially hinder the application of blockchain technology in journalism, where the timely dissemination and verification of information are crucial.

While various solutions have been proposed to address the scalability issue, such as *sharding* and off-chain transactions (Croman et al., 2016), these solutions are still in the experimental stages and may introduce other complexities. Additionally, the energy consumption associated with blockchain, particularly proof-of-work systems, raises environmental concerns that could impede scalability efforts (Flintham et al., 2018).

Regulatory Hurdles

Regulation represents another substantial challenge to the adoption of blockchain technology in journalism (Hileman, Rauchs, 2017). Blockchain operates in a legal grey area in many jurisdictions, with laws and regulations struggling to keep up with the rapid development of the technology. This uncertain regulatory environment can pose risks for journalists and newsrooms seeking to leverage blockchain technology. For example, the use of cryptocurrencies for micropayments, a potential application of blockchain in journalism, is subject to varying and often unclear regulations across different jurisdictions.

Moreover, the decentralized and borderless nature of blockchain also presents regulatory challenges. It may be difficult for newsrooms to comply with different regulatory standards across various jurisdictions, particularly in areas such as data protection and privacy (Finck, 2018).

Privacy Concerns

Despite the potential of blockchain technology to enhance privacy through decentralization and cryptography, it also raises significant privacy concerns (Zyskind, Nathan, Pentland, 2015). One concern is that while transactions on the blockchain are pseudonymous, they are also public and traceable. This could potentially lead to de-anonymization, where individuals' identities are revealed by analyzing their transaction patterns (Meiklejohn et al., 2013). Moreover, once

data is stored on the blockchain, it is immutable and cannot be deleted. This immutability could potentially conflict with privacy rights such as the “right to be forgotten,” which allows individuals to request the deletion of their personal data under certain conditions (Ausloos, 2012: 143–152).

Establishing Trust for Widespread Adoption

Finally, for blockchain technology to revolutionize journalism, it needs to be widely adopted by journalists, newsrooms, and consumers. However, the adoption of blockchain technology is currently limited by the challenges mentioned above, as well as a general lack of trust, awareness, and understanding of the technology (Hughes et al., 2019). Many people are still unfamiliar with blockchain and its potential applications, contributing to a hesitancy to adopt and engage with blockchain-based platforms.

Blockchain technology’s success in journalism is inherently dependent on network effects, where the value of the system increases with the number of users (Parker, Van Alstyne, 2005). In the context of journalism, this means that both content producers and consumers must adopt the technology for it to be effective. For instance, a blockchain-based platform for verifying the provenance of news would be less effective if only a small number of newsrooms use it. Similarly, a blockchain-based system for content monetization would require widespread adoption among consumers to generate significant revenue for journalists and newsrooms. Trust is a critical factor

in the adoption of blockchain technology in journalism. Journalists, newsrooms, and consumers must trust that the technology is secure, reliable, and beneficial. This is particularly challenging given the skepticism and misconceptions surrounding blockchain and cryptocurrencies (Mougayar, 2016).

In short, while blockchain technology holds significant potential to transform journalism, there are substantial challenges that need to be overcome. Addressing these challenges will require concerted efforts from technologists, journalists, regulators, and society at large. The journey towards blockchain-powered journalism is likely to be a complex one, but with careful navigation of the challenges, the potential rewards are significant.

Conclusion

A reflection on the initial skepticism that surrounded the Internet—often dismissed as slow, impractical, and inadequate—offers a valuable lesson about the challenges of predicting the trajectory and potential impact of emerging technologies on journalism. The true potential of the Internet, particularly its capacity to revolutionize how news is produced and consumed, was largely underestimated. This past miscalculation should motivate us to approach blockchain technology with a more open mindset, rather than passing premature judgment.

This chapter asserts that, despite its challenges, blockchain technology carries significant potential to reshape journalism in ways

currently unattainable. While the current state of blockchain operations—characterized by slow speed, limited scalability, high cost, and significant energy consumption—faces substantial hurdles, these difficulties are not insurmountable. Just as the Internet evolved, it is plausible that future, more efficient blockchain models will emerge and seamlessly integrate into the future web. Once this transition occurs, blockchain's influence on journalism could be far-reaching, touching upon various aspects such as content monetization, archiving, intellectual property protection, press freedom enhancement, and investigative journalism. Even though numerous projects have already explored these areas with limited success, their pioneering efforts are paving the way for future innovations.

However, the challenges—ranging from the digital divide, privacy concerns, energy consumption, to trust issues—must not be downplayed. We must acknowledge and address these hurdles if blockchain's unique attributes, such as its immutability, transparency, global distribution, and programmability, are to be fully harnessed. The financial implications associated with cryptocurrencies and blockchain should not overshadow these other attributes. Instead, a balanced view should be maintained, recognizing both the potential benefits and the challenges. Encouraging further research into the myriad applications at the intersection of blockchain and journalism will be a positive contribution to the field of digital journalism. It will help the profession avoid a lag in embracing technological advancements, ensuring journalism remains resilient and adaptable in the face of continuous change.

References

- Al-Saqaf, W., & Edwardsson, M. P. (2019). Could blockchain save journalism? An explorative study of blockchain's potential to make journalism a more sustainable business. In *Blockchain and Web 3.0*. Routledge.
- Al-Saqaf, W., & Seidler, N. (2017). Blockchain technology for social impact: Opportunities and challenges ahead. *Journal of Cyber Policy*, 2(3), 338–354. <https://doi.org/10.1080/23738871.2017.1400084>
- Arslanian, H. (2022). Ethereum. In H. Arslanian (Ed.), *The book of crypto: The complete guide to understanding Bitcoin, cryptocurrencies and digital assets* (pp. 91–98). Springer International Publishing. https://doi.org/10.1007/978-3-030-97951-5_3
- Ausloos, J. (2012). The 'right to be forgotten'—Worth remembering? *Computer Law & Security Review*, 28(2), 143–152. <https://doi.org/10.1016/j.clsr.2012.01.006>
- Azhari, T. (2021, September 21). FEATURE—Young Lebanese driving crypto “revolution” after banks go bust. *Reuters*. <https://www.reuters.com/article/lebanon-crypto-currency-youth-idUSL8N2QH1MW>
- Böhme, R., Christin, N., Edelman, B., & Moore, T. (2015). Bitcoin: Economics, technology, and governance. *Journal of Economic Perspectives*, 29(2), 213–238. <https://doi.org/10.1257/jep.29.2.213>
- Bullmann, D., Klemm, J., & Pinna, A. (2019). In search for stability in crypto-to-assets: Are stablecoins the solution? *SSRN Scholarly Paper*. <https://doi.org/10.2139/ssrn.3444847>
- Christodoulou, K., Katelaris, L., Themistocleous, M., Christodoulou, P., & Iosif, E. (2022). NFTs and the metaverse revolution: Research perspectives and open challenges. In M. C. Lacity & H. Treiblmaier (Eds.), *Blockchains and the token economy: Theory and practice* (pp. 139–178). Springer International Publishing. https://doi.org/10.1007/978-3-030-95108-5_6
- Criddle, C., & Oliver, J. (2022, March 1). How Ukraine embraced cryptocurrencies in response to war. *Financial Times*.

- Croman, K., Decker, C., Eyal, I., Gencer, A. E., Juels, A., Kosba, A., Miller, A., Saxena, P., Shi, E., Gün Sirer, E., Song, D., & Wattenhofer, R. (2016). On scaling decentralized blockchains. In J. Clark, S. Meiklejohn, P. Y. A. Ryan, D. Walach, M. Brenner, & K. Rohloff (Eds.), *Financial cryptography and data security* (pp. 106–125). Springer. https://doi.org/10.1007/978-3-662-53357-4_8
- DiCicco, K. W., & Agarwal, N. (2020). Blockchain technology-based solutions to fight misinformation: A survey. In K. Shu, S. Wang, D. Lee, & H. Liu (Eds.), *Disinformation, misinformation, and fake news in social media: Emerging research challenges and opportunities* (pp. 267–281). Springer International Publishing. https://doi.org/10.1007/978-3-030-42699-6_14
- Elliott, F. (2009, January 3). Chancellor Alistair Darling on brink of second bailout for banks. *The Times*. <https://www.thetimes.co.uk/article/chancellor-alistair-darling-on-brink-of-second-bailout-for-banks-n9l382mn62h>
- Ellsworth, B. (2021, June 22). As Venezuela's economy regresses, crypto fills the gaps. *Reuters*. <https://www.reuters.com/technology/venezuelas-economy-regresses-crypto-fills-gaps-2021-06-22/>
- Feder, A., Gandal, N., Hamrick, J. T., & Moore, T. (2017). The impact of DDoS and other security shocks on Bitcoin currency exchanges: Evidence from Mt. Gox. *Journal of Cybersecurity*, 3(2), 137–144. <https://doi.org/10.1093/cybsec/tyx012>
- Finck, M. (2018). Blockchains and data protection in the European Union. *European Data Protection Law Review*, 4(1), 17–35.
- Flintham, M., Karner, C., Bachour, K., Creswick, H., Gupta, N., & Moran, S. (2018). Falling for fake news: Investigating the consumption of news via social media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–10). <https://doi.org/10.1145/3173574.3173950>
- Hasan, H. R., & Salah, K. (2019). Combating deepfake videos using blockchain and smart contracts. *IEEE Access*, 7, 41596–41606. <https://doi.org/10.1109/ACCESS.2019.2905689>
- Hileman, G., & Rauchs, M. (2017). 2017 global cryptocurrency benchmarking study. SSRN Scholarly Paper No. 2965436. <https://doi.org/10.2139/ssrn.2965436>

Hughes, A., Park, A., Kietzmann, J., & Archer-Brown, C. (2019). Beyond Bitcoin: What blockchain and distributed ledger technologies mean for firms. *Business Horizons*, 62(3), 273–281. <https://doi.org/10.1016/j.bushor.2019.01.002>

Ito, J., Narula, N., & Ali, R. (2017, March 8). The blockchain will do to the financial system what the internet did to media. *Harvard Business Review*. <https://hbr.org/2017/03/the-blockchain-will-do-to-banks-and-law-firms-what-the-internet-did-to-media>

Kamphausen, G., & Werse, B. (2019). Digital figurations in the online trade of illicit drugs: A qualitative content analysis of darknet forums. *International Journal of Drug Policy*, 73, 281–287. <https://doi.org/10.1016/j.drug-po.2019.04.011>

Karim, S., Lucey, B. M., Naeem, M. A., & Uddin, G. S. (2022). Examining the interrelatedness of NFTs, DeFi tokens and cryptocurrencies. *Finance Research Letters*, 47, Article 102696. <https://doi.org/10.1016/j.frl.2022.102696>

Kharpal, A. (2022, November 21). Bitcoin drops to lowest in more than a week, ether slides as FTX collapse ripples through crypto market. *CNBC*. <https://www.cnbc.com/2022/11/21/bitcoin-btc-ether-eth-fall-as-ftx-collapse-ripples-through-market.html>

Larios-Hernández, G. J. (2017). Blockchain entrepreneurship opportunity in the practices of the unbanked. *Business Horizons*, 60(6), 865–874. <https://doi.org/10.1016/j.bushor.2017.07.012>

Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G. M., & Savage, S. (2013). A fistful of bitcoins: Characterizing payments among men with no names. In *Proceedings of the 2013 Conference on Internet Measurement Conference* (pp. 127–140). <https://doi.org/10.1145/2504730.2504747>

Mougayar, W. (2016). *The business blockchain: Promise, practice, and application of the next Internet technology*. John Wiley & Sons.

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. <https://bitcoin.org/bitcoin.pdf>

O'Sullivan, J., & Heinonen, A. (2008). Old values, new media. *Journalism Practice*, 2(3), 357–371. <https://doi.org/10.1080/17512780802281081>

Parker, G. G., & Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. *Management Science*, 51(10), 1494–1504. <https://doi.org/10.1287/mnsc.1050.0400>

Perreault, G., & Stanfield, K. (2019). Mobile journalism as lifestyle journalism? *Journalism Practice*, 13(3), 331–348. <https://doi.org/10.1080/17512786.2018.1424021>

Rogers, E. M., Singhal, A., & Quinlan, M. M. (2008). Diffusion of innovations. In D. W. Stacks & M. B. Salwon (Eds.), *An integrated approach to communication theory and research* (2nd ed., pp. 418–434). Routledge.

Roush, C. (2017, May 15). Quartz reporter creates bot tracking hacking payments. *Talking Biz News*. <https://talkingbiznews.com/they-talk-biz-news/quartz-reporter-creates-bot-tracking-hacking-payments/>

Saini, K. (2021). Blockchain foundation. In K. Saini., P. R. Chelliah, & D. K. Saini (Eds.), *Essential enterprise blockchain concepts and applications* (pp. 1-14). CRC Press. <https://doi.org/10.1201/9781003097990-1>

Samaniego, M., Jamsrandorj, U., & Deters, R. (2016). Blockchain as a service for IoT. In *2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom), IEEE Cyber, Physical and Social Computing (CPSCoM), and IEEE Smart Data (SmartData)* (pp. 433–436). <https://doi.org/10.1109/iThings-GreenCom-CPSCoM-SmartData.2016.102>

Sanders, C. E., Field, T. M., Miguel, D., & Kaplan, M. (2000). The relationship of Internet use to depression and social isolation among adolescents. *Adolescence*, 35(138), 237–242.

Shukla, S. (2022, September 28). NFT trading volumes collapse 97% from January peak. *Bloomberg*. <https://www.bloomberg.com/news/articles/2022-09-28/nft-volumes-tumble-97-from-2022-highs-as-frenzy-fades-chart>

Sigalos, M. (2022, November 11). Sam Bankman-Fried steps down as FTX CEO as his crypto exchange files for bankruptcy. *CNBC*. <https://www.cnbc.com/2022/11/11/sam-bankman-frieds-cryptocurrency-exchange-fts-files-for-bankruptcy.html>

Stoll, C. (1995, February 26). The Internet? Bah! Hype alert: Why cyberspace isn't, and will never be, nirvana. *Newsweek*. <https://www.newsweek.com/clifford-stoll-why-web-wont-be-nirvana-185306>

Tapscott, D., & Tapscott, A. (2018). *Blockchain revolution: How the technology behind Bitcoin and other cryptocurrencies is changing the world* (1st ed.). Portfolio/Penguin.

Teng, Y. (2022). What does it mean to trust blockchain technology? *Metaphilosophy*, 54(1), 145–160. <https://doi.org/10.1111/meta.12596>

Thelwall, M. (2018). Can social news websites pay for content and curation? The SteemIt cryptocurrency model. *Journal of Information Science*, 44(6), 736–751. <https://doi.org/10.1177/0165551517748290>

Uslaner, E. M. (2010). Trust and the economic crisis of 2008. *Corporate Reputation Review*, 13(2), 110–123. <https://doi.org/10.1057/crr.2010.8>

Valdeolmillos, D., Mezquita, Y., González-Briones, A., Prieto, J., & Corchado, J. M. (2020). Blockchain technology: A review of the current challenges of cryptocurrency. In J. Prieto, A. K. Das, S. Ferretti, A. Pinto, & J. M. Corchado (Eds.), *Blockchain and applications* (pp. 153–160). Springer International Publishing. https://doi.org/10.1007/978-3-030-23813-1_19

Voshmgir, S. (2020). *Token economy: How the Web3 reinvents the internet* (2nd ed.). BlockchainHub.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power* (Paperback ed.). Profile Books.

Zyskind, G., Nathan, O., & Pentland, A. S. (2015). Decentralizing privacy: Using blockchain to protect personal data. In *2015 IEEE Security and Privacy Workshops* (pp. 180–184). <https://doi.org/10.1109/SPW.2015.27>

Innovating in Journalism with Newsgames: An Exploratory Study in Portugal

Bruno Frutuoso
Costa
Bruno Contreiras
Mateus

Introduction

The Fourth Industrial Revolution, also referred to as digital transformation, is presented as a paradigm in which cutting—edge technologies are utilized in automation and data exchange to enhance the quality, efficiency, and productivity of companies' production processes (Gilchrist 2016; Skilton, Hovsepian, 2018). Early on, the automotive, technology, and biologics industries took the lead in this transformation (Coelho, 2016).

In the creative industries, particularly media, innovation emerged in the United Kingdom at the end of the 20th century, encompassing

Bruno Frutuoso Costa, Research Fellow, Iscte – University Institute of Lisbon, CIES-Iscte, Lisbon ORCID: 0000-0003-3023-8960, email: bruno.frutuoso.costa@hotmail.com

Bruno Contreiras Mateus, Guest Assistant Professor, Iscte – University Institute of Lisbon and IADE – European University, Lisbon ORCID: 0009-0007-0458-8947, email: bruno.goncalo-mateus@iscte-iul.pt

DOI: <https://doi.org/10.5281/zenodo.14203373>

sectors that combine creativity and communication technologies. The aim was to create economic and social value for companies (Granado, Silva, Vicente, 2020). By the beginning of the 21st century, innovation became closely associated with the rise of the Internet (Bonixe, 2020: 153–169).

Blogs and social media platforms were two forms of online communication initially adopted by individual users seeking to share opinions, thoughts, and ideas, as well as connect with friends and others. They were quickly integrated into news organizations, driven by user feedback and participation, leading to the development of new multimedia features and functionalities (Dogruel, 2014: 52–69).

Although traditional news companies have been compelled to migrate to digital platforms at the risk of compromising their revenues, Bonixe (2020) notes that this does not necessarily imply that the innovation processes in the media began solely with this transition. He provides management and production process innovation as examples.

Innovation in the media industry varies significantly across countries and companies and is influenced by financial factors, organizational culture, and government regulation (de-Lima-Santos, Ceron 2021; Meier et al., 2022; Pérez-Seijo, Vicente, 2022).

From 2010 to 2020, Meier studied and compared five European countries—namely Germany, Austria, Spain, Italy, and Switzerland—and identified the most relevant journalistic innovations, which include the use of social media and networks, the creation of new multimedia formats, the implementation of data analysis tools, and the adoption of digital business models (Meier et al., 2022).

In the area of artificial intelligence, it has been employed to automate repetitive tasks, such as report writing, and to personalize content relevant to news websites. These actions are supported by large data sets that reflect users' interests and behaviors in the digital environment (de-Lima-Santos, Ceron, 2021).

Data analysis to identify trends and patterns, automatic fact-checking, and the creation of automatic captions and transcripts for videos are underutilized in the news industry (*ibid.*), except in countries such as the U.S. and China.

In Portugal, the adaptation of the business model of news companies occurs slowly and involves few forms of innovation (Cardoso et al., 2019; Crespo et al., 2018). Newsgame projects are a journalistic innovation defined in the literature as the creation of games based on real events, intended to convey complex news information in a more attractive and engaging way to increasingly fragmented audiences across platforms (Bogost, Ferrari, Schweizer, 2012). Thus, this study aims to contribute to the production of scientific knowledge on a subject that is little studied in Portugal by identifying the challenges of creating newsgame projects and exploring professionals' perceptions regarding the integration of innovation in the media industry.

Industry 4.0 and News Media

Industry 4.0 encompasses more than the digitization of equipment, procedures, data collection, and IoT networking. Digital transforma-

tion is built on interconnectivity, transparency of information, decentralized decision-making, and technical assistance (Hermann, Pentek, Otto, 2016), all of which require companies to embrace speed, comprehensiveness, and depth of change (Schwab, 2017). This perspective complements the vision introduced by Takeuchi and Nonaka (2008/2004), who argue that sustainable companies are those that create knowledge, disseminate it throughout the organization, and rapidly incorporate it into new technologies and products.

Technological evolution has made scalable cloud and service solutions available, such as Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Hybrid Cloud. The combination of these solutions with cyber-physical systems, algorithms (which integrate implicit business logic), and cybersecurity mechanisms in a mixed infrastructure allocated in public and private data centers enables the development of integrated information management solutions. In turn, artificial intelligence empowers these infrastructures with the ability to learn and produce accurate, complex, and sophisticated predictions that support decision-making and the definition of new business directions (Iansiti, Lakhani 2020; Skilton, Hovsepian, 2018).

The false belief that Industry 4.0 is synonymous with complex, expensive tools only accessible to multinational companies has prevented many organizations from investing in the digitization of key operations within their value chains and research and development (RandD) capabilities for value-added products and services to address market challenges and opportunities (Gilchrist, 2016).

Media, together with performing arts, visual arts, music, film and video, design, advertising, and leisure software, constitute the so-called creative industries (Granado, Silva, Vicente, 2020). The diversification

of journalistic products, new business models, and the competition for space with digital platforms are inseparable dimensions of the investment made by the news industry in emerging technologies.

These initiatives can be analyzed under the 4Ps model introduced by Francis and Bessant (2005), which segments innovation into: 1) the ability to introduce or improve products; 2) the ability to introduce or improve processes; 3) the ability to define or redefine the positioning of the company or its products; 4) the ability to define or redefine the dominant paradigm within the company (Francis, Bessant, 2005).

According to Dogruel, innovation in media has eight characteristics, in terms of products and processes, that distinguish it from other industries: 1) its multidimensional nature involves technological, economic, social, and cultural dimensions and elements; 2) there is a high level of uncertainty and unpredictability; 3) as innovations are targeted at users, they seek to align with user feedback and participation; 4) they have a significant impact on society and culture in terms of social norms, values, and behaviors; 5) complex interactions occur between various actors and stakeholders; 6) they require new forms of collaboration and cooperation among different actors and entities to be successful; 7) they necessitate new forms of regulation and governance; and 8) they presuppose new business models to remain sustainable over time (Dogruel, 2014: 52-69).

Challenges Associated with Innovation in the News Media

Not being a recent practice in journalism, innovating is associated in recent years essentially with the development of apps (Fante, 2018), betting on storytelling strategies (Granado, Silva, Vicente, 2020), such as gamification, augmented reality, infographic representation (Gehlen, Sousa 2018), and binaural sound (Vicente, Pérez-Seijo, 2022), as well as redefining the organization's positioning or its business model (Bonixe, 2020 153–169).

The migration of news media to digital platforms has generated new challenges in terms of routines, languages, deontology, and training of professionals. Major media events are seen, for example, by news companies as an opportunity to test new ways of reporting events. In Bonixe's study (*ibid.*), the papal visits to Portugal were used by radio stations to launch 360-degree videos, live and continuously.

In Spain, one of the most significant innovations is the adoption of digital technologies for the production, distribution, and consumption of content. This has allowed news organizations to reach wider audiences, engage with readers in new ways, and test new ways of storytelling. For example, El País launched a virtual reality app that allows audiences to experience news in immersive 360-degree environments (García-Avilés et al., 2019).

The study by Meier (2022) found significant differences between the five European countries studied in terms of the innovations adopted

(Meier et al., 2022). While Germany and Switzerland focused on the creation of new multimedia formats, Spain and Italy focused more on the use of social media. In parallel, Austria was the country that most implemented data analysis tools in its journalistic work. Although most European media outlets have adopted social media as a tool for news reporting, some countries, such as Spain and Germany, have been more successful in effectively integrating these platforms into journalistic work.

Innovation and its integration are a complex process that requires a combination of resources, collaborative leadership, and a predisposed organizational culture (García-Avilés et al., 2019; Meier et al., 2022; Pérez-Seijo and Vicente, 2022). Many organizations have limited resources to invest in expensive innovation, which can make it difficult to adopt new technologies and infrastructure (García-Avilés et al., 2019; Meier et al., 2022; Pérez-Seijo and Vicente, 2022).

Journalists and other professionals may be more resistant to using new technologies because they feel that they undermine their primary mission of reporting events or because of the perception that they may threaten their work, but also because they lack the technical skills to implement them effectively (García-Avilés et al., 2019; Meier et al., 2022; Pérez-Seijo and Vicente, 2022).

Meier et al. (2022) point out that sustainable business models are scarce. Several innovations have been shown not to be economically viable in the long term, which can make them difficult to adopt and cause companies to fear them.

The need to find a balance between innovation and traditional journalistic values takes on special importance in the use of artificial

intelligence in newsrooms (Meier et al., 2022; Pérez-Seijo and Vicente, 2022). Studies such as de-Lima-Santos and Ceron (2021) and Pérez-Seijo and Vicente (2022) have identified that, in addition to the lack of financial resources to invest in these technologies and the difficulties in finding qualified professionals, there is a clear concern about ethics and data privacy associated with the lack of understanding of how artificial intelligence can be used in the service of journalism.

To explore these factors and how they influence media innovation, García-Avilés et al. (2019) interviewed twenty journalists who were managing innovation projects in Spain. Faced with declining revenues, the companies implemented new business models, such as sponsored content and hosting events, as well as new content production, such as virtual reality and interactive infographics.

The factors that conditioned the implementation of the innovation were the availability of resources and support from management (García-Avilés et al., 2019). Journalists who had access to training and other resources were more available than those who did not. In addition, having the support of management was crucial to overcoming resistance to change and ensuring that innovations were integrated into the culture of the organization.

Overview of Journalism Innovation in Portugal

In a survey of 236 Portuguese journalists, Cardoso et al. (2019: 25) identify that 64.8% have an academic background in communication

sciences. Journalists face significant challenges in their work, such as the pressure to produce content quickly and the lack of resources that can constrain innovation in the industry.

The OberCom report (Cardoso et al., 2019) also provides, albeit little, information on journalists' perceptions of innovation in newsrooms. Most respondents (60.5%) consider the use of new technologies important to improve journalistic work.

The study by Crespo et al. (2018) reinforces that news companies in Portugal are slowly adapting to new practices, favoring the use of different available narratives, and editing with new publishing tools. Legacy media were at an early stage of searching for new sources of revenue beyond traditional advertising. The rate of innovation in startups' business models was even slower since they did not have economic shortages. In natively digital media, the search for new revenue streams was very much on the board (Crespo et al., 2018: 100).

During the Covid-19 pandemic, journalists adopted new routines, such as working remotely, conducting interviews by phone or video conferencing (Miranda et al., 2021). To cope with information overload and adapt to constantly changing news related to the pandemic, some journalists chose to focus on reliable and verified sources, while others used monitoring tools to filter relevant information (Miranda et al., 2021: 298).

Looking at the Portuguese case through case studies, Bonixé notes that broadcasters have sought to progressively adopt "measures at the level of their organizational structure that stem from their presence on the Internet. (...) These measures are of various levels: convergence between traditional radio and digital platforms, adoption of

multimedia tools, brand expansion and professional profiles” (Bonixe 2020: 158).

In affirming radio on the aesthetic, creative and informational planes of the digital medium, static and animated infographics began to be used in the construction of stories, as well as photography, photo galleries, videos, 360 degrees, streaming broadcasts, graphic information, creation of spaces for listeners to participate in debates through comments left on the website and other digital platforms, geolocation and timelines for chronological events (ibid.: 158).

As for the dimension of capitalizing the online presence with the creation of new products, Bonixe highlights the creation of newspapers in pdf and a radio for the deaf, with translation into Portuguese Sign Language of the sound broadcast. The sound reports mostly use binaural technology (Vicente and Pérez-Seijo, 2022).

The study by Fante (2018) reports that some companies are developing apps. The apps, two from the press (Expresso and Público) and one from the digital media (Observador), have features such as notification, geolocation, and content customization to offer a more engaging and interactive experience to customers. However, the author identifies that the multimedia character is little diversified (Fante, 2018).

The four journalists interviewed in the study by Gehlen and Sousa (2018), who work in the field of data journalism, in Expresso, Público, and O Jogo newspapers and Renascença radio station, mention that data can be worked to produce multiplatform reporting on any topic, from politics, economics, sports to the environment. However, the task of working with data implies a rigorous process of analysis,

treatment, and multimedia editing of the information into infographics. The need for investment in training and capacity building of professionals or the hiring of human resources are the main obstacles to its consolidation in Portuguese journalism.

As for innovation in news coverage of Covid-19 in Portugal, the press continued to focus on storytelling formats and data journalism, with immersive, empathetic, and explanatory narratives. Some Portuguese newspapers, such as Público and Expresso, created specific sections on their websites dedicated to Covid-19 coverage. These sections included up-to-date news, the visualization of data through interactive infographics and maps, and information on prevention and treatment measures (Silva and Granado, 2021: 12).

Gamification of news

In 2013, the most popular New York Times story was not an article but rather an interactive story called “How Y’all, Youse, and You Guys Talk”. Considering that the newsgame was launched on December 21 of that year, it took 11 days to generate more hits than any other linear or multimedia journalistic text (Meyer, 2014). It is increasingly common to see regular newsgame launches, particularly in the largest international media outlets. The study by Ferrer-Conill et al. (2020) offers a panoply of examples, from The Financial Times’ award-winning “Uber” project to “A World without Chocolate”, an independent collaboration between Spain and Italy that was simultaneously published in El Period, Al Jazeera, Tages Anzeiger, and La Repubblica.

Cattani and Amaro (2018) argue that newsgames can serve as a complement or constitute new business models for traditional journalism by attracting new types of advertising and new customers, because of increased visibility through shares on social networks and traffic generated on websites. Regarding the investment required, the authors point out that many international projects are crowdfunded. This practice of collective funding can involve many people interested in supporting innovative journalistic projects.

As a form of innovation, it allows users to interact with important topics and better understand the context of the knowledge one wants to convey through a news game. Newsgames can be defined as games created from real events, making information more accessible and interesting (Bogost et al., 2012; Cattani & Amaro, 2018). Events are at the basis of the narrative and gameplay of this form of interactive journalism that has the dual purpose of informing and educating the “players” about those topics. Often the stories would be complex and difficult to understand by other traditional journalistic means (Bogost et al., 2012).

Newsgames can also be a way to promote media literacy, helping journalistic audiences develop critical skills to filter information in the digital age (Cattani & Amaro, 2018). On the other hand, these are used in the higher education of students at the level of journalistic writing and the development of interviewing techniques (Aoki, 2020). The potential to amplify the diversity of voices and perspectives in journalism allows different social groups to be represented and heard (Cattani & Amaro, 2018).

Newsgames can also be considered tools that break barriers between journalists and audiences, if they are used by news companies with

the collaborative goal or to stimulate participation (Aoki, 2020). In the first dimension, audiences can intervene in the construction of the game through open codes. In the second, they allow them to experience the news in a more engaging way, whose participation enables organizations to collect and analyze data from participants.

In the gamification of news, Bogost et al. (2012) stated that the great challenges for newsrooms lie in finding the right balance between gameplay and information, ensuring the transmission of information that is accurate and relevant from the world and therefore understandable to a wide variety of audiences, as well as ensuring that “players” are not led by the game narratives to wrong or biased conclusions.

As in Portugal the adaptation of the business model of journalistic companies occurs slowly and there are no scientific studies on news-games, an exploratory interview was conducted on April 12, 2023, to structure the questionnaire survey that is presented and discussed in the following sections.

We interviewed Diogo Queiroz de Andrade, director of innovation at Global Media Group, one of the largest media groups in Portugal and with presence in the press, radio, and digital sectors. He has over 25 years of experience, making him one of the most knowledgeable people in the digital journalism market in Portugal.

Andrade (2023) framed newsgames as part of the diversification of digital news formats, which have been “one of the elements that most contributes to attracting new audiences” and to respond to the industry crisis, as it “forces us all (rightly) to understand how we can better serve our customer.”

His definition of newsgames highlights not only the immersive context in which information is conveyed to the reader in a subtle way, but also the narrative context “that appeals to emotional factors”. He suggests that this appeal is at the genesis of the interaction between the consumer and the product, and that the nature of the format does not change the essential journalistic criteria.

When asked about the main challenges of this innovation process, Andrade (2023) mentioned the creation of a product that makes sense, conveys information, involves the reader, and triggers an emotional response. In this context, not all news topics are likely to be transformed, which represents an inherent limitation to the viability of the model.

In terms of production, the difficulties relate to the conservative mentality dominant in organizational cultures in Portugal and the need for programming and design skills, something that the current human resources do not possess.

Contrary to what one might think, journalism is a rather conservative profession where innovative practices enter very slowly and against a lot of resistance. The gamification of news is already a reality, the shift towards greater adoption of this and other formats has precisely to do with the need to be more open to innovation.

(Andrade, 2023)

Methodological Choices

For the present study, we are interested in starting from the vision that innovation can be seen as a way for the news industry to find new business models and respond to the challenges imposed by the digital public space. Especially due to the democratization of the process of creating and sharing content to any user, which contributes to capturing attention, since this method sustains the business model of news companies. In this context, the difficulty of journalism to engage the public in an increasingly fragmented and personalized environment between digital platforms emerges (Pérez-Seijo and Vicente, 2022).

Although several innovative strategies have been identified in Portugal, it is necessary to understand how these projects are effectively integrated into newsroom routines in order to consolidate their use (Bonixe, 2020).

This paper aims to explore the implementation of newsgames as a form of journalistic innovation in Portugal and to understand what challenges are faced by organizations. The research is guided by the following research questions:

RQ1: To what extent does declining revenue promote the creation of newsgames?

RQ2: What were the challenges associated with news gamification?

RQ3: How does innovation enable media to respond to changes impacting the business model?

The survey has been considered as a method that “provides a quantitative or numerical description of the trends, attitudes, or opinions of a population by studying a sample of that population” (Babbie cited in Creswell, 2009: 29). García-Avilés et al. (2019) demonstrated that practitioners’ experiences provide insight into the context and factors that shape the process of journalistic innovation.

Using a mixed methodology that allowed flexibility in the vertical analysis of the key aspects of innovation in the media (administration of evidence), we opted to construct a questionnaire survey consisting of 15 open questions in the form of a semi-structured dialog (Appendix 1). The survey sent by email was applied during the time window from April 28 to May 31, 2023. The selection of the non-probability convenience sample (Bryman, 2012) was designed to reflect all innovative media of different typologies (press, radio, digital, and television) and to consist only of professionals who have participated in the creation of newsgames.

The sample, presented in Figure 1, is composed of 5 male and 5 female individuals, working in the press (70%), digital journalism (10%), television (10%), and radio (10%). All Portuguese media that have launched newsgames are represented, namely *Expresso*, *Observador* and *Público*. This means that the two participants from *Rádio Renascença* and *CNN* have worked previously in these media.

Figure 1: Media outlets of the sample (n = 10)

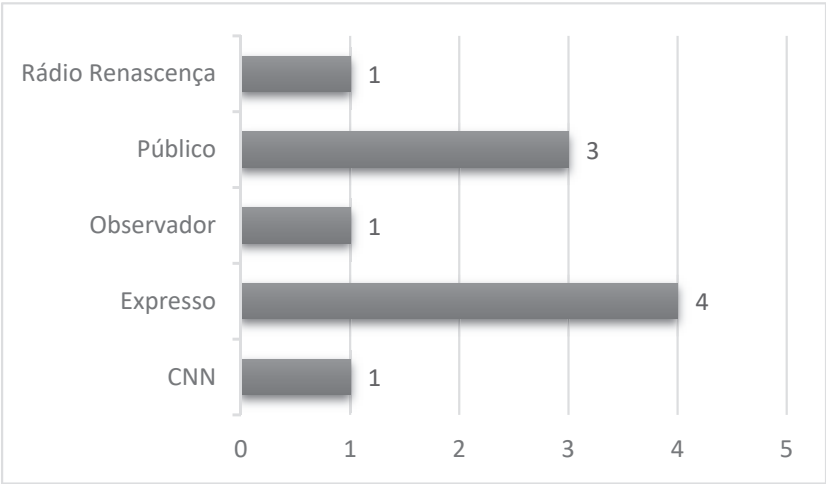


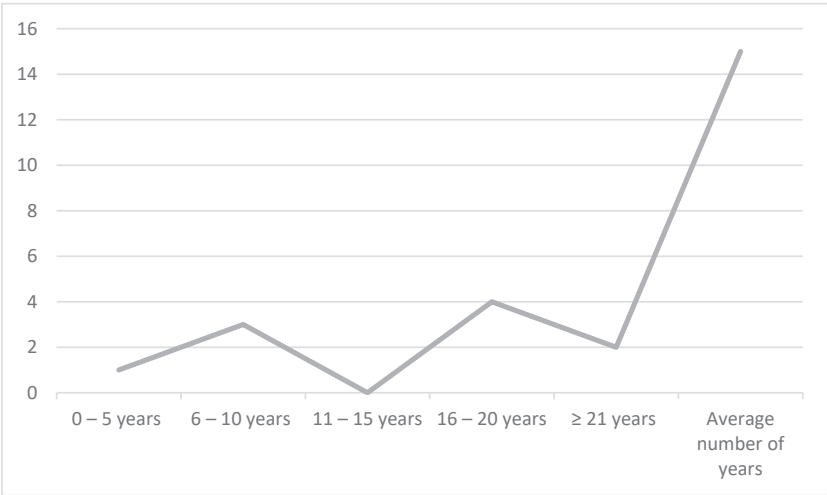
Table 1: Roles in the newsroom

Role	Absolut Frequency	Percentage
Multimedia Coordinator	2	20%
Web Designer	2	20%
Editor-in-Chief	1	10%
Deputy Director	1	10%
Online Editor	1	10%
Journalist	1	10%
Data Journalist	1	10%
Infographic Journalist	1	10%
	n = 10	100%

As can be seen in Table 1 and Figure 2, the sample is diverse in the positions they hold in newsrooms and reveals long longevity and professional experience. The average number of years that the participants

have been in journalism is 15 years. Directors, editors, coordinators, and newsroom chiefs represent 50% of the sample, who joined the newspaper industry at least 16 years ago. In addition to management and leadership positions, the other half correspond to positions as web designer, journalist, data journalist, and infographic journalist.

Figure 2: Professional Experience



This study is the result of a qualitative descriptive analysis of the data from the survey, with primarily expository objectives for the information collected and some simple quantitative observations. The data presented below is not intended to exhaust the approach to the reality of Portuguese journalism in terms of newsgames.

The empirical content analysis tool was used with the aid of MAXQDA software to combine qualitative and quantitative analysis techniques to interpret and determine some patterns and meanings of the content conveyed through the textual responses (Bardin, 2016/1977). The categorization matrix was constructed by deductive and inductive

methods, from the literature and the data under analysis, respectively (Elo and Kyngäs, 2008; Mayring, 2000).

Thus, in the first phase of coding, we characterized the participations according to the following variables: gender of the participant, media outlet, position, and professional experience.

In the second phase, we analyzed the dimensions and respective categories, with only one sublevel, identified from the literature and coded according to each description (category memo): newsgames concept (audience participation and not referencing audiences), genesis of the first project (migration to digital, revenue drop, new business model, content complement, international projects, multimedia experiences, new audiences, and boosting engagement), participation in projects (number of projects and function performed), creation process (stages and journalistic criteria), innovation challenges (playability-information balance, leadership, organizational culture, government regulation, financial factors, technical skills, human resources, time spent, financial return, application structure and speed of development), audience receptivity (good, poor, incipient, can't respond), ways to measure return (social media metrics, website metrics, and new customers), journalists' perceptions of innovation (harms journalistic mission, threatens work, response to changes affecting business model, ethical issues, informational relevance, attracting new audiences, and active audience engagement), and other ongoing innovations (social media, new multimedia formats, data analysis tools, adoption of digital business models, artificial intelligence, doesn't know how to respond, and digital transformation).

In this way, the questionnaire represents the unit of analysis of the database, while the unit of analysis of the corpus or record is the answer

to each open question. During the analytical process, the dimensions were not coded, as they correspond to nominal variables. The category characterizes the content, which is not the case with the dimensions or concepts.

The absolute frequency values shown in Tables 2 and 3 are the result of one of the features of the MAXQDA software. Each reference, even synonyms or expressions that point to the same value, was counted when it referred to one of the previous categories. The categorization process was carried out once, discussed and repeated a second time by the researchers in June 2023, in Portugal, to minimize subjectivities arising from the coding process (Lupton, 2006). At the request of some participants, anonymity will be granted in the reporting of their statements. For this reason, all identifications are merely fictitious names.

Discussion of Results

The snapshot of the documents (Figure 3) makes it possible to compare how the codes were assigned across all the paragraphs in the documents. As this visualization is proportional to the size of each document, most of the coded segments belong to the dimensions “innovation challenges” (78, 35%), “participation in projects” (37, 17%), and “journalists’ perceptions of innovation” (25, 11%). The disparate relative proportions of the categories under analysis show that there is no fixed sequence of themes mentioned by the participants, even though the questionnaire survey was constructed in a semi-structured way.

Figure 3: Document comparison chart

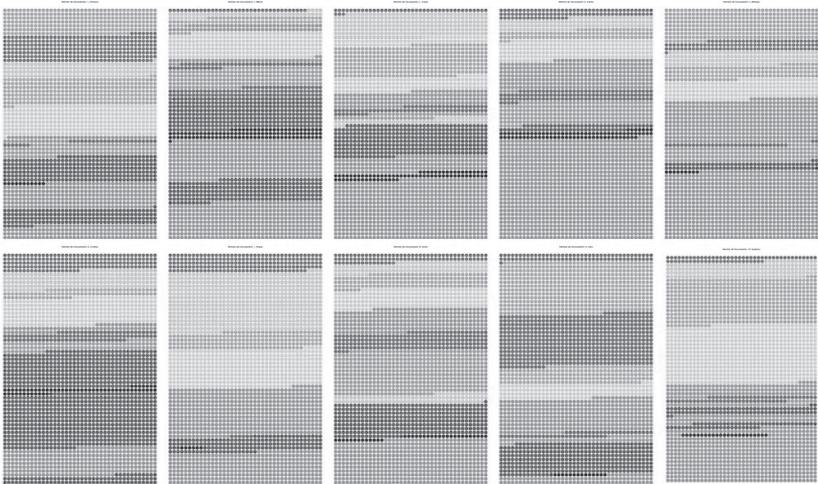


Figure 3: Document comparison chart

- Figure legend:**
- > ● CREATION PROCESS
 - > ● PARTICIPATION IN PROJECTS
 - > ● GENESIS OF THE FIRST PROJECT
 - > ● INNOVATION CHALLENGES
 - > ● JOURNALISTS' PERCEPTIONS OF INNOVATION
 - > ● NEWSGAMES CONCEPT
 - > ● OTHER ONGOING INNOVATIONS
 - > ● AUDIENCE RECEPTIVITY
 - > ● WAYS TO MEASURE RETURN

In the dimension “newsgames concept”, it is identified that 70% of the newsgames definitions presented by the journalists do not mention the importance of this innovation to stimulate the active participation of the public, either in the news consumption or in the process of news gamification itself. Gustavo, fictitious name, explains that it is a “diffuse concept” that sometimes generates diverse and misunderstood interpretations.

Manuel’s definition was the one that best fit the initial conceptualization of Bogost et al. (2012) and supplemented with inputs from Cattani and Amaro (2018). See his explanation.

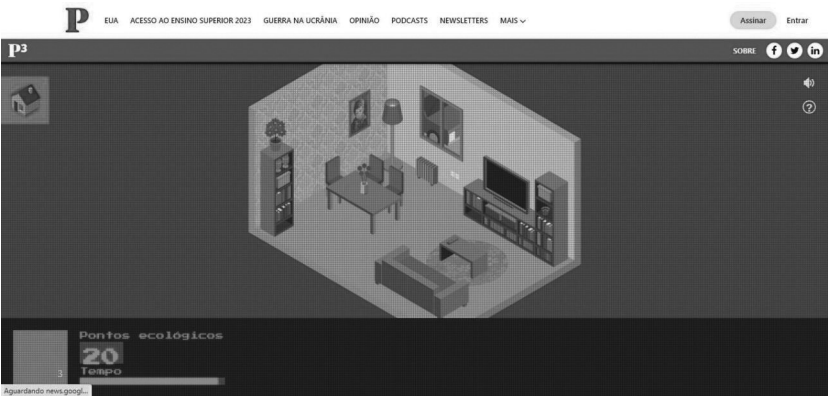
Some people in the field will define newsgames as a type of game built with journalistic principles and whose goal is to inform readers by putting them at the center of—or as part of—a story (usually complex or best explained in the first person). I prefer a reverse and broader definition: they are a form of gamification of news or information, with the goal of engaging and motivating readers in a story or topic. (Manuel, 2023)

In recent years, some newsgames projects have been sporadically launched in Portugal. Examples are “Eco’a Breca” from Público in September 2019, “Consegues Sobreviver como um estudante universitário?” from Expresso in September 2022, and “Costa Strike” from Observador in May 2023.

“Eco’a Breca” (Figure 4) was one of the first newsgames in Portugal, and it allowed users to explore ecological challenges inside their homes, related, for example, to water or energy consumption in everyday tasks. The interface is simple, with images of the interior of a house representing several rooms, but very functional from the point of view of gamification and the information it intends to pass on to users, who collect ecological points until they reach a goal of 200.

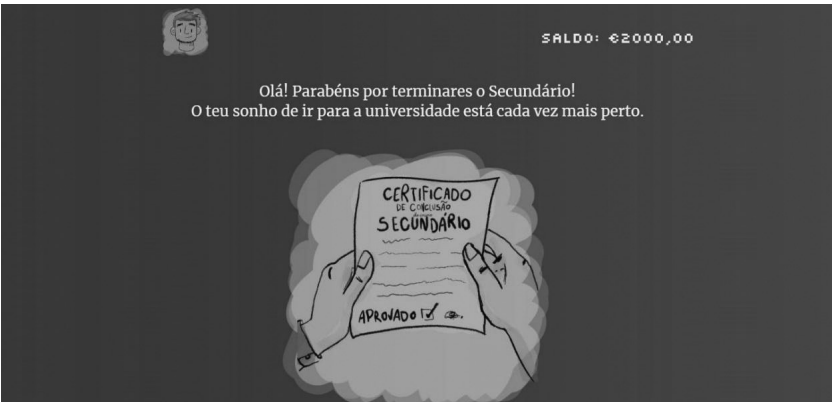
“Consegues Sobreviver como um estudante universitário?”, that means “Can You Survive as a University Student?”, (Figure 5) came at a critical moment in Portugal, when student room prices rose by up to 70% from one year to the next, which led to many having to look for new accommodation solutions or even giving up studying. This game explores the management of a budget, so the user can put themselves in the shoes of a student far from home and manage their money

Figure 4: Application structure of “Eco’a Breca”



according to the challenges of everyday life. This narrative gets users thinking about the challenges students face and their dilemmas.

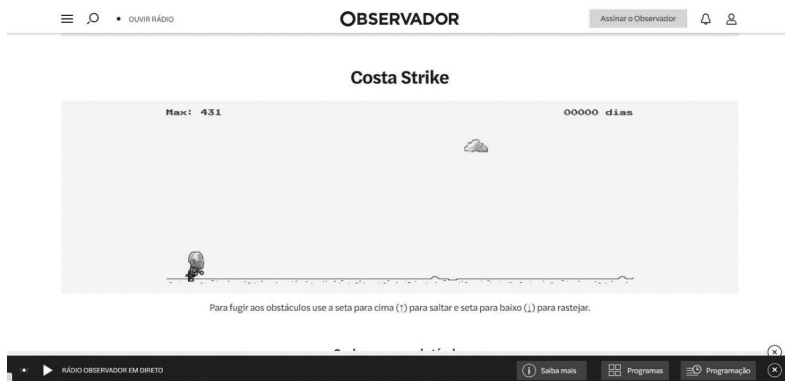
Figure 5: Application structure of “Consegues Sobreviver como um estudante universitário?”



In the third example, the newsgame “Costa Strike” (Figure 6) comes amid controversy in national politics. Costa is the prime minister. And he is in trouble due to a succession of controversial events in one of

the most important ministries, including the minister, a stolen computer with confidential information, and scenes of violence in ministerial offices. The game is more entertaining than informative, but, as in all political crises, it puts Costa face-to-face with all the players in this political game.

Figure 6: *Application structure of “Costa Strike”*



The participants refer that newsgames arise essentially to test new multimedia formats and add value to content. At no time was mentioned the goal of constituting a new business model, the drop in revenue or innovation resulting from the migration of media to digital. For example, at Expresso newspaper and Rádio Renascença, newsgames were born to test new multimedia formats and add value to news content.

As for the role played by the participants, one can see that, given the lack of human and economic resources of newsrooms in Portugal, there are no well-defined roles. In the absence of multidisciplinary teams and with little technical knowledge of programming, Diana mentioned that, while in the first project she focused on prototyping,

in the second project she was asked to play the role of journalist and web developer, having had the collaboration only for the illustrations. Gustavo, data journalist, pointed out:

I end up being responsible for almost all the work – even though it is always collaborative work with the editorial staff. Usually, I am the author of the ideas, or if it is just a theme (the typical “I should do something about [the theme],” I am the person who is responsible for coming up with a more creative solution. (Gustavo, 2023)

Although there is no common newsgames creation process for all newsrooms, it is possible to identify in the participants’ responses several good practices that can be replicated and applied by other media companies, namely:

- *Research and theme discussion*: initial meeting between journalists, designers, editors, editorial coordinators or heads, and directors to discuss ideas, possible paths and reach the best way to tell the story based on the type of information that at the outset is thought to be needed. Example of questions used in this phase: What story do we want to tell? Is the angle of the story still the same? What is the amount of data? Will it be a more multimedia game? Is it interesting to incorporate videos or other elements?
- *Conception and Design*: in this editorial production phase, after research and information gathering, the journalists create the script where the concept is defined, the places and challenges

proposed in each section are identified, and the form and punctuation scale are defined, as well as the textual elements to be included.

- *Programming and design.* The first phase is the idealization of the prototype, sketching wireframes or data-driven (codes for the latter case), and the delineation of the functionalities that the game will have by the programmers and web designers. Then operationalization consists of designing the environments and developing the program schema. Next, programming begins and the creation of illustrations and other graphic solutions to incorporate the editorial production.
- *Testing and bug fixing phase:* because it is an iterative process of research and development, as soon as it is possible to test minimally, the programmers share with journalists to adjust, changes, and incorporate definitive texts, already chosen images, among others.
- *Publication, distribution and promotion* on the homepage, social networks, newsletters, among others.

All participants mention that the journalistic and ethical criteria for newsgames should be the same as for all other written or image content published in the media, to which can be added those of a technical nature, namely functionality, usability, image.

Proximity, relevance, and novelty were the most prominent news criteria. Still, there are themes with more potential for readers' involvement, through discussion, exchange of perspectives and sharing on social networks.

Gustavo pointed out that proximity is defined by several authors as a news factor. However, in newsgames this proximity is emotional in addition to geographical. “With some of these features, it is possible for us to give personalized information based on the user’s geography – which, by its nature, is clearly more relevant. But at its genesis, the editorial and journalistic criteria are always the same.” For you, news-games are just a way of building the narrative.

As we can see in Table 2, among the challenges associated with innovation, human resources are the most invoked by the interviewees for this research. In second place is the time spent. And then there are technical skills. The CNN Portugal participant points out that newsrooms are still too focused on fast news.

“At this moment, I think that newsgames in Portugal are still isolated experiences, given the time and resources needed to develop them. In newsrooms increasingly consumed and dominated by the production of fast news and with more and more economic constraints, I think there is still a long way to go before we can say that newsgames are part of the informative strategy of a national media organization.” (CNN Portugal, 2023)

The Renascença participant corroborates and complements this idea. “Especially the development, in the projects that are ours from scratch, because we don’t have a dedicated team of developers. And time. They are usually time-consuming and complex projects that involve several iterations”.

About the shortage of technical skills in the newsrooms, one of the respondents of *Espresso* believes that administrations should allocate part of the budget to professionals specialized in technological areas.

The proximity of the IT sector to the newsroom; the ability to hire people with both journalistic and IT skills, which are in short supply; administrations must open up a budget gap that allows for long-term investments, so that these jobs stand out. (*Espresso*, 2023)

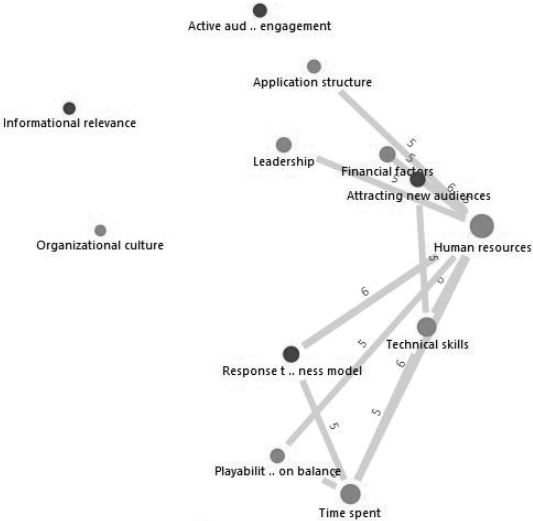
In fact, financial issues are another challenge that many point out as being associated with innovation. With less frequency in the answers but with great strategic implications for media companies, leadership and organizational culture are two aspects focused on. Financial return and speed in development are also among the concerns.

Table 2: *Challenges associated with innovation*

	Absolute Frequency	Percentage
Human resources	18	23,1
Time spent	13	16,7
Technical skills	12	15,4
Financial factors	8	10,3
Leadership	7	9
Game-information balance	6	7,7
Application Structure	5	6,4
Financial return	5	6,4
Speed in development	2	2,5
Organizational culture	2	2,5
Government regulations	0	0
TOTAL	78	100

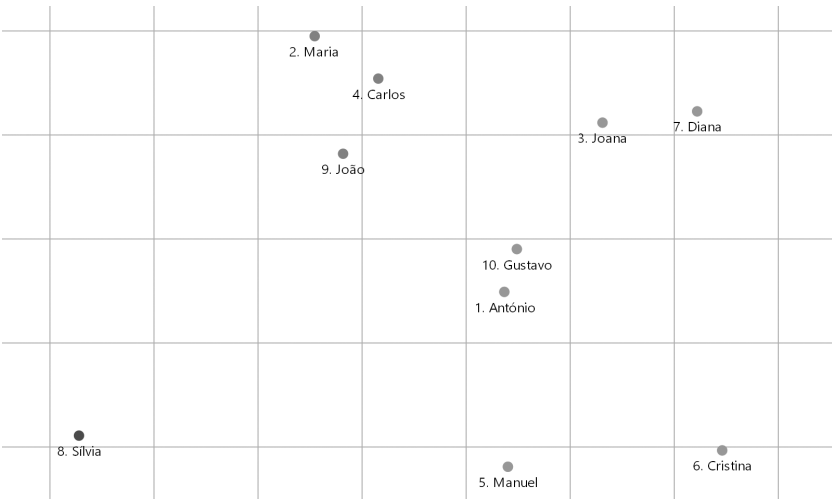
The code map (Figure 7) shows the relationships between the codes in a document for the dimensions “innovation challenges” and “journalists’ perceptions of innovation”. For a frequency of five or more, human resources, the main challenge of innovation, are often associated with technical skills, financial factors, attracting new audiences, and responding to changes affecting the business model. This means that, although innovation is seen by the participants as a way of responding to economic provocations and attracting new customers, newsrooms need new human resources whose skills involve challenging investments for a business model that has been in crisis for several decades. The organizational culture category is distant from the other clusters and has no established relationship, suggesting that it is not an obstacle in the Portuguese context. However, leadership is pointed out as a constraint.

Figure 7: Map of relationships between the codes for the dimensions “innovation challenges” and “journalists’ perceptions of innovation”



We used a document map (Figure 8) to see if there was any similarity between the documents in terms of the codes and variables assigned. This showed that there are two sets of documents with high similarity.

Figure 8: Document map



To recognize which variables may influence the views and positions of the newspaper industry in relation to innovation, two similarity matrices were applied to analyze the similarities between the documents of the participants with the same positions (Table 3). Maria, Carlos, and João, as well as Gustavo and António, have the same positions, the former as managers and the latter as technicians, suggesting that they tend to share the same opinions or value certain subjects more than others.

Table 3: *Similarity matrix of the documents according to the variable “management positions”*

Document name	2. Maria	3. Joana	4. Carlos	8. Sílvia	9. João
2. Maria	1,00	0,78	0,78	0,70	0,83
3. Joana	0,78	1,00	0,78	0,61	0,78
4. Carlos	0,78	0,78	1,00	0,65	0,74
8. Sílvia	0,70	0,61	0,65	1,00	0,74
9. João	0,83	0,78	0,74	0,74	1,00

The good receptivity to newsgames by the public was highlighted by 90% of the sample, and 10% didn’t know how to answer. There are essentially 3 ways to measure the return on investment, namely website metrics, mentioned 9 times, social media metrics, mentioned 4 times, and new customer acquisition, which was mentioned only 1 time.

The mechanisms commonly used on websites are creation of triggers in the code that detect certain actions on the special page, the length of the ‘scrolled’ reading, the times on the page, specific buttons, and various navigations. The return on social media is measured by the engagement that these projects generate in sharing and discussion.

From a journalist’s perspective, innovation comes in response to both business needs and new ways of reaching audiences. A perspective well reflected by one of Observador’s respondents.

“I believe that all innovation efforts, in a journalistic context, that allow bringing audiences closer and involving them with the news are extremely advantageous. On the one hand, they may have an impact on the viability of the business model by generating potential new readers. But also, because it is the obligation of journalism to adapt to the forms of consumption of news and reports, so that our work has the widest possible reach”. (Observador, 2023)

As shown in Table 4, for the newsroom professionals, innovation comes up as a response to changes affecting the business model with an absolute frequency of 32%. “New formats and innovation are essential in responding to change. Of course, the balance between time invested and immediate results is never balanced in the start-up phase. The time invested is almost always much greater than the immediate result”, as one of the participants from Expresso said.

Regarding the publics, the absolute frequency of responses is 28% for attracting new audiences and 20% for active involvement of the public. As the Renascença respondent says:

We believe that only by investing in new formulas can we reach new audiences and combat the complex challenges that the business faces. Looking for these formulas, experimenting, failing, and trying again is, in our view, crucial to finding more interesting paths for journalistic content and better results. Formulas cannot, however, distract from the essential: doing good journalism with criteria and balance.

Finally, about innovation, informational relevance appears with an absolute frequency of 12% in the answers; 4% point to threats to the work of newsrooms, and 4% to the minimization of the journalistic mission. No one mentions ethical issues.

Table 4: *Journalists' perceptions of innovation*

	Absolute Frequency	Percentage
Response to changes affecting the business model	8	32
Attracting new audiences	7	28
Active involvement of the public	5	20
Informational relevance	3	12
Threatens the work	1	4
Undermines the journalistic mission	1	4
Ethical issues	0	0
TOTAL	25	100

On the question about ongoing innovations, half of the participants do not know how to answer because of lack of knowledge. This suggests that boards and other management positions do not engage the entire organizational culture around change. As for the other half of the sample, 3 media are developing new multimedia formats, 1 is continuing its digital transformation, and 1 is exploring the potential of artificial intelligence for processes related to data development, analysis, and processing.

Conclusion

The news media have sought to enhance their presence in the digital environment. Not only as a means of survival, but also to assert themselves on an aesthetic, creative and informative level. In this multi-platform and multimedia context, there is a need to conduct longitudinal studies that follow the implementation and impact of innovation projects (García-Avilés et al., 2019).

Portuguese journalism strives to keep up with the new paradigm shift that characterizes Industry 4.0 to find possibilities to connect and build loyalty among different audiences on different platforms, while maintaining its credibility and relevance in the market. Still, the severe economic crisis it has faced in the last decade has reduced teams that have neither the time nor the technical skills to develop and test new forms of innovation.

The sporadic creation of newsgames projects is seen by companies as a new, more visual, interactive, and immersive way of telling real events. As an innovation, newsgames are not without risks, where a poorly conceived or executed narrative can create biases or prejudices on the part of the reader.

Besides the various human and technical resources, the main challenge associated with this form of innovation is the fact that the media have not yet managed to find a way to monetize it. As Pérez-Seijo and Vicente (2022) argue, this new form of storytelling and gamification is not a magic solution to all the problems of journalism, so different multimedia formats and business models should complement each other. Still, one cannot fail to highlight the importance of

newsgames to keep audiences informed and engaged, who can also be involved in the construction processes.

This complex and constantly evolving process requires a careful and collaborative approach to ensure that technologies are used effectively and responsibly. Collaboration and knowledge sharing between media and other industry players, such as universities and technology companies, are important in helping media overcome some of the challenges associated with innovation, such as lack of resources and sustainable business models.

While the study provides valuable information on how newsgames emerge in the media sector in the form of innovation, and some of the findings can be applied to other contexts, it is not without limitations. It should be taken into consideration that the study focused specifically on Portuguese media and on a convenience sample. All professionals involved in newsgames projects were considered and invited to participate.

The main limitation of the study is the choice of the data collection method (questionnaire surveys constituted only by open questions). Participants do not always give the necessary reflexivity to each topic. Future studies should seek to explore, mainly through semi-structured interviews, the extent to which cultural and institutional factors influence the diffusion of innovation in different regions or countries and explore the perspectives of other stakeholders, such as publics and advertisers, who can also play an important role in defining and promoting innovation in the media industry.

References

- Aoki, R. L. (2020). Aprendizagem baseada em jogos digitais aplicada ao ensino de redação jornalística. *Media and Jornalismo*, 20(36), 109–128. https://doi.org/10.14195/2183-5462_36_6
- Bardin, L. (2016). *Análise de conteúdo* (L. Reto & A. Pinheiro, Trans.). Edições 70. (Original work published 1977).
- Bogost, I., Ferrari, S., & Schweizer, B. (2012). *Newsgames: Journalism at play*. The MIT Press.
- Bonixe, L. (2020). Jornalismo radiofónico e inovação: Uma análise à cobertura de acontecimentos mediáticos. *Media and Jornalismo*, 20(36), 153–169. https://doi.org/10.14195/2183-5462_36_8
- Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.
- Cardoso, G., Baldi, V., Crespo, M., Pinto-Martinho, A., Pais, P. C., Paisana, M., & Couraceiro, P. (2019). *O que devem saber os jornalistas? Práticas e formação em Portugal*. OberCom. <http://www.obercom.pt/wp-content/uploads/2019/04/Relatorio-OberCom-Abril-2019.pdf>
- Cattani, D., & Amaro, M. (2018). Newsgame: Como os jogos podem salvar as notícias e fortalecer um novo negócio? *Estudos de Jornalismo*, 9, 61–74. https://www.revistaej.sopcom.pt/ficheiros/20190103-ej9_2018.pdf#page=60
- Coelho, P. M. (2016). *Rumo à Indústria 4.0* (Master's thesis). Universidade de Coimbra. <http://hdl.handle.net/10316/36992>
- Crespo, M., Foà, C., & Pinto-Martinho, A. (2018). Como o jornalismo lida com a inovação: Um estudo de caso das melhores práticas em Portugal. *Estudos de Jornalismo*, 9, 75–102. <http://hdl.handle.net/10071/17489>
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Sage.

de-Lima-Santos, M.-F., & Ceron, W. (2021). Artificial intelligence in news media: Current perceptions and future outlook. *Journalism and Media*, 3(1), 13–26. <https://doi.org/10.3390/journalmedia3010002>

Dogruel, L. (2014). What is so special about media innovations? A characterization of the field. *The Journal of Media Innovations*, 1(1), 52–69. <https://doi.org/10.5617/jmi.v1i1.665>

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115. <https://doi.org/10.1111/j.1365-2648.2007.04569.x>

Fante, A. (2018). Jornalismo aplicado: Arquiteturas da notícia. *Estudos de Jornalismo*, 9, 112–124. https://www.revistaej.sopcom.pt/ficheiros/20190103-ej9_2018.pdf#page=111

Ferrer-Conill, R., Foxman, M., Jones, J., Sihvonen, T., & Siitonen, M. (2020). Playful approaches to news engagement. *Convergence*, 26(3), 457–469. <https://doi.org/10.1177/1354856520923964>

Francis, D., & Bessant, J. (2005). Targeting innovation and implications for capability development. *Technovation*, 25(3), 171–183. <https://doi.org/10.1016/j.technovation.2004.03.004>

García-Avilés, J. A., Carvajal-Prieto, M., Arias, F., & De Lara-González, A. (2019). How journalists innovate in the newsroom: Proposing a model of the diffusion of innovations in media outlets. *The Journal of Media Innovations*, 5(1), 1–16. <https://doi.org/10.5617/jomi.v5i1.3968>

Gehlen, M. A., & Sousa, J. P. (2018). Jornalismo de dados em Portugal: Um estudo exploratório sobre práticas jornalísticas especializadas. *Estudos de Jornalismo*, 9, 125–139. <http://www.revistaej.sopcom.pt/edicao/234>

Gilchrist, A. (2016). *Industry 4.0: The industrial internet of things*. Apress.

Granado, A., Silva, D. S., & Vicente, P. N. (2020). Inovação nos media e nas indústrias criativas limítrofes—uma introdução. *Media and Jornalismo*, 20(36), 5–9. https://doi.org/10.14195/2183-5462_36_0

Hermann, M., Pentek, T., & Otto, B. (2016). Design principles for Industrie 4.0 scenarios. In *Proceedings of 49th Hawaii International Conference on System Sciences HICSS* (pp. 3928–3937). <https://doi.org/10.1109/HICSS.2016.488>

Iansiti, M., & Lakhani, K. R. (2020). *Competing in the age of AI: Strategy and leadership when algorithms and networks run the world*. Harvard Business Review Press.

Lupton, D. (2006). Sociology and risk. In G. Mythen & S. Walklate (Eds.), *Beyond the risk society: Critical reflections on risk and human security* (pp. 11–24). Open University Press.

Mayring, P. (2000). Qualitative content analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(2). <https://doi.org/10.17169/fqs-1.2.1089>

Meier, K., Schützeneder, J., García-Avilés, J. A., Valero-Pastor, J. M., Kaltenbrunner, A., Lugschitz, R., Porlezza, C., Ferri, G., Wyss, V., & Saner, M. (2022). Examining the most relevant journalism innovations: A comparative analysis of five European countries from 2010 to 2020. *Journalism and Media*, 3(4), 698–714. <https://doi.org/10.3390/journalmedia3040046>

Meyer, R. (2014, January 17). The New York Times' most popular story of 2013 was not an article. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2014/01/-em-the-new-york-times-em-most-popular-story-of-2013-was-not-an-article/283167/>

Miranda, J., Fidalgo, J., & Martins, P. (2021). Jornalistas em tempo de pandemia: Novas rotinas profissionais, novos desafios éticos. *Comunicação e Sociedade*, 39, 287–307. [https://doi.org/10.17231/comsoc.39\(2021\).3176](https://doi.org/10.17231/comsoc.39(2021).3176)

Pérez-Seijo, S., & Vicente, P. N. (2022). After the hype: How hi-tech is reshaping journalism. In J. Vázquez-Herrero, A. Silva-Rodríguez, M. C. Negreira-Rey, C. Toural-Bran, & X. López-García (Eds.), *Total journalism: Studies in big data* (pp. 41–52). Springer.

Silva, D. S., & Granado, A. (2021). Cobertura jornalística dos números da Covid-19: Casos de inovação em Portugal. *Obi.Media*. <https://doi.org/10.34619/x70x-posb>

Skilton, M., & Hovsepien, F. (2018). *The 4th industrial revolution: Responding to the impact of artificial intelligence on business*. Palgrave Macmillan.

Schwab, K. (2017). *The fourth industrial revolution*. Crown Publishing Group.

Takeuchi, H., & Nonaka, I. (2008). *Gestão do conhecimento* (A. Thorell, Trans.). Bookman. (Original work published 2004).

Vicente, P. N., & Pérez-Seijo, S. (2022). Spatial audio and immersive journalism: Production, narrative design, and sense of presence. *Profesional de la Información*, 31(5), Article e310519. <https://doi.org/10.3145/epi.2022.sep.19>

Annex 1 – Questionnaire survey

1. Can you explain to us what you mean by newsgames?
2. How did you come up with the idea of creating the first newsgames project where you work?
3. How many and which newsgames projects have you worked on?
4. What was your role in this project/s?
5. What are the stages associated with their development?
6. What journalistic criteria are used in the gamification of the news?
7. What were the main challenges?
8. What was the public's receptivity to the project/s?
9. How do you measure the return on this investment?
10. To what extent does innovation allow you to respond to the changes that impact the business model of the journalism industry?

11. Can you share what other i4.0 initiatives have been developed or are being planned at the media outlet where you work?
12. How would you define Portugal's state of the art in newsgames?
13. Do you think newsrooms should invest more in newsgames in Portugal?
14. In your opinion, what is the reason for not having so much investment?
15. What needs to change in the newsrooms for the gamification of news to become a reality?

Digital Technologies, Individualisation, and the Rise of Digital Populism

Dominic Spada

Introduction

Across the world, digital connectivity continues to rise. In the US alone, smartphone and tablet ownership increased by 42% from 2010 to 2016 (Langford et al., 2019), a penetration that reached over 80% in 2021 (Perrin, 2021). Social media penetration is also significant, for example in Hungary where around 71% of the population has a registered Facebook account (Statista, 2023). The data collected by these devices and platforms, combined with the reliance that we all have to them, create a world of their own that are controlled by the data cycle.

One of the social implications of this data collection and processing is its impacts on democratic process. In several of his writings, Alexis de Tocqueville notes the importance of community and communication in avoiding the effects of inherent individualisation in democracy. In Tocqueville's writings, the shift from aristocracy to democracy came

Dominic Spada, PhD Student, The American University of Paris, Paris,
ORCID: 0009-0000-8416-8043, email: dominic.spada@aup.edu

DOI: <https://doi.org/10.5281/zenodo.14203490>

with a breakdown of the social stability and obligation that had existed for centuries, leaving men to fend only for themselves and their immediate families. The risks of this, he explained, were the atomisation of society, and the opening that this presents for autocratic individuals to take advantage of the weakened social links.

This concept around individualism can be seen in many respects in 21st century political discussion, most prominently in recent memory within COVID-19 pandemic discussions where individualistic narcissism clashed with calls for collectivism. These prominent instances of individualism run deep in political contexts, where they have notably been linked to democratic backsliding, a trend which has gathered steam since the early 2000s. For the context of this discussion, backsliding will generally correspond to the decay of the institutions and values supporting democracy (Gora and de Wilde, 2022). Though a historic phenomenon, many have questioned the reasonings for this most recent backsliding, with rationales ranging from economic, migration, and climate crises, to religious and cultural clashes.

This article seeks to propose another possible cause for democratic backsliding in the 21st century, one linked to rising digital technology usage undercutting many democratic principles. Technology has disrupted the way we all communicate with one another and receive information, shifting political discussion from inherently public forums onto private bulletin board platforms. This shift sees the definition of “community discussion” also challenged, with the community no longer representing a consistent grouping of people with a common mindset. Instead, this community presents itself broken into individuals, each with their own interests in mind as they engage in political discussion. The result here is an atomised political society—one

lacking social ties that Tocqueville noted as being essential to maintaining democracy and avoiding despotism. This article does not seek to propose a new analysis of Tocqueville's writings per se, instead it uses Tocqueville to help us better understand contemporary backsliding under the contexts of digital despotism. The relevance of Tocqueville here comes in the comparison of the current democratic climate with that of Tocqueville. Where Tocqueville wrote on the democratic revolution, a sense of novelty in governance that had been emerging over several hundred years, we now write on democracy in the technological revolution. The vision that Tocqueville had in the outsets of the democratic emergence can hence give us unique insights in how to handle the emergence of a new structure in democratic governance.

This article is split into three sections. The first will engage a review of the literature on democratic backsliding, highlighting the absence of the individualising nature of technology as a cause. The following two sections will discuss Tocqueville's writings on individualisation and its risks, connecting them with digital technologies and modern backsliding with a brief case study on Hungary, followed by a conclusion.

Section 1: Literature Review

Definitions of liberal democracy extend from being simply composed of elections and political competition to more complex configurations of civil liberties; separations of power and rule of law (Gora and de Wilde, 2022). All of these aspects, to one extent or another, have

been challenged in recent years across established democracies, with many countries generally having considered to suffer some form of democratic backsliding. One of the most notable characteristics in studies on regime types since the 1970s has been the resiliency of authoritarian regimes despite successive waves of democratisation worldwide in the latter half of the 20th Century. Even with economic and social modernisation, autocratic governance in countries including China, Russia, and Hungary persevered or regained strength against strong democratic headwinds (Welzel, 2017). This trend is notably represented by the general increasing stagnation in democratic transition since the 1990s and an increase in democratic reversals since 2005 (Welzel, 2017). The stability provided by unitary authoritarian regimes provides to many an appealing form of governance, particularly to generalities in developing middle-income countries,¹ where economic development has become tied to further social developments and governance (Lee, 2018). More generally, commonly cited influences, albeit under varying circumstances, to the increased willingness to support autocratic leaders and increasing polarisation include declining political trust, economic conditions, racial issues, low social mobility, and other social factors (Kaufman and Haggard, 2019). A number of countries, prominently Hungary and Turkey, also saw the weakening of judicial independence crackdowns on media outlets, with Turkey going so far as to become one of the world's leaders in journalist imprisonments (Kaufman and Haggard, 2019). Further instability can further stem from "ethnic, racial, and religious cleavages" (Kaufman and Haggard, 2019, p. 418).

¹ Generalities, or a generality, throughout this paper refers to the active political society as utilised by Pierre Rosanvallon.

Notably, the democratic states facing the most significant cases of backsliding are those in which opposition strength was weakest,² while countries where opposition parties or coalitions remain strong have faced less severe backsliding despite autocratic leaders being elected (Kaufman and Haggard, 2019).³ Weakened oppositions came from two different directions, either from incapacity or unwillingness to form a united opposition,⁴ or from crackdowns on opposition from majority parties.⁵ In these cases, lack of strong constitutional protections and degradations in horizontal checks and balances allows the majority government to pass legislation limiting opposition access to political competition or,⁶ in the case of Poland, refuse to confirm the appointment of opposition officials to their roles (Nalepa, 2021). Two principal theoretical strands can help guide discussions on democratic regressions, the first on polarisation within the general electorate and regime dysfunction, while the second stresses polarisation among political elites and institutional failures (Kaufman and Haggard, 2019). When elite polarisation is high, tolerance

2 See Abdullah (2020); Pérez-Liñán, A., Schmidt, N., and Vairo, D (2019); Bermeo (2016); and Riaz (2021) for more detailed examples of the weakening of opposition parties and how this contributes to executive aggrandisement and related backsliding.

3 Such is the case in the United States under the Trump presidency, where strong opposition from the Democratic party in part balanced out the autocratic tendencies of Trump and his allies.

4 Such as the case of Turkey.

5 In these cases, lack of strong constitutional protections and degradations in horizontal checks and balances allows the majority government to pass legislations.

6 The examples referenced in footnote 13 offer a more in-depth the analysis of access limitations for opposition parties to political competition.

and self-restraint within political competitions weakens (Kaufman and Haggard, 2019), and in such cases voters will more willingly elect autocrats aligning more closely with their ideological preferences due to the high perceived cost of being ruled by an alternative ideology (Nalepa, 2021).

The countries with these weakened constitutional protections for oppositions are often those emerging from autocracies and whose democracies have yet to be consolidated. Transitional justice structures are often promoted in order to prevent backsliding during the transition from democracy to autocracy, a process including mechanisms for removing incumbent holdovers from autocratic regimes to limiting the influence of former autocrats (Bates et al., 2020). Where transitional justice runs into challenges, however, is where specific terms such as denazification or decommunisation, used to describe the both the structural and personnel transition to democracy, are used as a false-flag tactic to re-create an autocratic structure (Nalepa, 2021).

This time period of stagnation in democratic transition has also been an era in which economic liberalism has flourished, a concept which emphasises the individual as the key to economic growth and progression. Neoliberalism, in promoting mistrust between individuals through risk-reward calculations, creates tendencies which inhibit solidaristic behaviour (Lynch and Kalaitzake, 2020). The economic privatisation assists in the expansion of the protected personal sphere, delegitimising social programmes and movements that are core to distinct social powers (Brown, 2019). Benevolence and warmth in the public sector therefore become conditional, based on deservedness and inspiring a society of self-responsibilised individuals (Lynch and Kalaitzake, 2020). Democracy in this era becomes one of constrained

democracy, with its constructors dedicated to the protection of individual rights and a constraining of majoritarianism (Wilkinson, 2019). These intentions, though done in the name of preserving democracy and preventing backsliding, instead hollow out democratic institutions in order to protect the free market from constraint (Wilkinson, 2019), providing a further influence to backsliding in conjunction with previously mentioned challenges.

Though instances of backsliding have been consistently present since the outset of the democratic age, occurrences have begun to take on new forms in the 21st century. Historical democratic failures tended to be rife with violence, with governments succumbing to coups, civil wars, and seizures of power through physical showings of force (Bernhard, 2021). Many countries nowadays do not suffer the violent revolutions of past democratic failures, instead suffering as individuals and groups win power through executive and parliamentary powers and move to undermine democratic norms (Bernhard, 2021). To achieve these goals, legislative majorities have reduced or eliminated horizontal checks and balances allowing corruption and crony capitalism to flourish while allowing executive powers to expand through fundamental constitutional changes (Haggard and Kaufman, 2021). Weakening in the checks and balances have come as strong institutions are generally better at enforcing their commitments intertemporally (Ezcurra, 2021), a critical notion as temporalities shift between digital platforms and traditional mediums of debate. These capacities lead to the institutions being more stable over time (Ezcurra, 2021, p. 3).

These movements have presented themselves differently depending on the country. Certain instances see parliamentary power gains through executives riding on early-term popularity to gain popular

support, while others see fracturing opposition parties exploited to gain majorities during legislative elections (Haggard and Kaufman, 2021). Majorities can then be increased through voting disproportionalities and electoral quirks (Haggard and Kaufman, 2021).⁷ Electorates may choose closet autocrats, either individuals or parties, for subsequent terms as they view them as ideological incumbents rather than autocrats (Nalepa, 2021), allowing autocratic ideologies to emerge and power consolidation to happen through increased majorities.

Further constitutional changes have included reduced legislative independence, minimised judicial authority, and increased executive discretion, the last of these commonly termed executive aggrandization. Executive aggrandization develops as legislative majorities and supermajorities allow for delegation to the executive power, having the knock-on effect of constitutional changes limiting horizontal accountability, opposition, and political and civil freedoms (Kaufman and Haggard, 2019). Further tolerance for authoritarian power grabs expands after re-election, a condition that is highly dependent on polarisation throughout the political system (Nalepa, 2021).

These clandestine forms of backsliding, using incumbent governments and executive aggrandization, are based principally on media censorship, averting horizontal accountability, and election manipulation (Bauer and Becker, 2020). Censorship and information manipulation have historically been one of the primary tools of autocratic regimes, but they have taken new forms in the digital age. Coming now

⁷ For example, the disproportionality of the US Senate voting system overrepresenting rural populations and gerrymandering allowing House seats to be won. These allow for majorities in situations where the popular vote normally would not allow it.

in the form of platform manipulation and content prioritisation, these two actions which can be utilised by populist and autocratic leaning governments to their own benefit. All states engaged in backsliding undermine at least one of the three components backsliding, most prominently in the diminished fairness of elections and reduced civil and political rights (Bellamy and Kröger, 2021).⁸ All of these have been complicated by mass data collection and analysis to the point where candidates can be reasonably certain exactly which citizens voted for them in a given precinct (Bauer and Becker, 2020).⁹ The principal factor threading through the discussion on backsliding is the presence of polarisation. These discussions on present-day polarisation and resulting backsliding cannot be made without discussing the role of the multiplicity of digital technologies utilised by the generality and political elites alike.

Democratic Backsliding and Digital Technologies

Most studies focus on backsliding through endogenous factors, including political ideology; competition patterns; the capture of the state by elite coalitions; and the role of the EU (Holesch and Kyriazi,

-
- 8 The three core components of democracy are free and fair elections, the presence of civil and political rights, and an independent judiciary.
 - 9 This process was prominently first used in the United States by Barack Obama in his 2008 presidential campaign and expanded upon for his second candidacy in 2012. See Bimber (2014) and Issenberg (2013) for further discussion.

2022). Backsliding cannot be measured using just one indicator and may vary across states as to which components of their democracies have backslid, complicating the discussion when trying to compare and contrast backsliding (Gora and de Wilde, 2022). One of the main challenges to measuring backsliding is the possibility of overestimating backsliding when using indicators and the failure to distinguish movement on individual components of democracy (Gora and de Wilde, 2022). This is particularly relevant to the role of digital technologies on democratic backsliding, where technology may have varying and concurrent effects across a number of different democracy indicators.

Failures of mainstream parties to respond to disillusioned generalities and politically motivated manipulation of the mass media have also been noted as contributors to the rise of far-left and far-right political parties, as have “patronalism, party state capture, and exclusionary identity politics” (Holesch and Kyriazi, 2022, p. 3). While discussions around democratic backsliding are extensive and thorough, they globally fail to account for this rise in digital technology usage during this period of democratic backsliding. Those that do are principally tied to political polarisation. Numerous cases of democratic backsliding have been studied across the world, almost all of which possess histories or recent spikes of polarisation, and cases which did not see polarisation spikes were theorised to have already high levels of polarisation (Haggard and Kaufman, 2021). Polarisation, the state of vast social divisions and higher political tensions, reduce the generality’s capacity to oppose increasingly autocratic tendencies of political representatives (Svolik, 2019).

Democracies must have both a controversial and non-controversial setting, the former being the debate promoted by pluralism, and the latter having the established rules of the political game being fair, consisting of steadfast legal norms, and maintaining civility as a principle of human interaction. It is this latter that is challenged by present-day populists (Bauer and Becker, 2020), often presently by the use of digital media and other platforms. Both the controversial and non-controversial sides of democracies are further challenged by individualising digital platforms and are an aid to the polarisation influenced by populist leaders. This digitally-led polarisation has consequently led to trends of democratic backsliding.

This environment pits the two factors against one another, in an us-versus-them contest, with a short step from there to placing greater value on winning than maintaining democratic rights and order (Haggard and Kaufman, 2021). Even within the electorate there is this debate between preserving democratic rights and supporting representatives that represent their interests, where democratic principles become less important the more partisan that the environment becomes (Graham and Svolik, 2020). In creating this “us versus them” environment, a total societal categorisation develops making peaceful coexistence impossible (McCoy et al., 2018). The divisions and unwillingness to collaborate leads to inefficiencies in the democratic process, increased political stalemates, and swings between political extremes. These all in effect leads to increasing disillusionment and distrust of political institutions by the generality (Haggard and Kaufman, 2021). The distrust sowed by inefficiencies are only heightened as reduced civility and respect in democratic discourse allow for misgivings of own-party representatives to be more easily overlooked, including transgressions violating democratic norms (Tucker, et al., 2018).

The rise in polarisation critically lowers the communication between different groups, with a sort of snowball effect as the different groups continuously view each other's viewpoints more critically as ill-intentioned (McCoy et al., 2018). In the natural state of democracy, competing elites are not enemies to be destroyed, but rivals in a game to be defeated (Gora and de Wilde, 2022). In the developing polarised state, however, partisanship has arisen as a part of social identity, condensing opposition voices into highly competitive camps (Tucker et al., 2018). This polarised backsliding has accompanied rival elites often being targeted as traitors, betrayers of the people, or criminals by their opponents within the political sphere, an effect that leads to either mainstream parties being taken over or displaced by extremist groups from both sides of the political spectrum (Haggard and Kaufman, 2021). In doing so, a strongly partisan environment develops that makes the generality far less resilient to groups desiring to disrupt democratic process (Graham and Svolik, 2020).

This paper thus argues that increased digital technology usage by general society has been a contributor to the general backsliding seen in democracies across the world, with no, as of yet, way to fully understand which aspects it most directly affects nor how to effectively measure it. Because of the general usage of digital technologies and the widespread collection of personal data from such technologies, the impacts from their usage potentially affect many different factors in democratic backsliding. This may allow technology usage by both citizens and political organisations to be noted as a key factor in the general backsliding noted over the last 15 years. One potentially critical factor that has been overlooked is the influence that societal atomisation, as defined by Alexis de Tocqueville, caused by digital technologies has in democratic backsliding.

Section 2: Tocqueville, Individualisation, and Democratic Backsliding

Alexis de Tocqueville noted several threats to democracy throughout his works, the threat of corruption through poorly educated leaders, a condensing tyranny of the majority, and a satisfaction with equality leading to complacency being particularly notable examples. It is in fact this satisfaction with equality, an equality of condition, that results from a hyper-individualised society. Individualism is not wholly bad, with it resting as a “mature and calm feeling”, one that allows each member of society to “sever himself from the mass of his fellows” and form a “circle of his own” composed of friends and family (Tocqueville, 1998, p. 205). It is in this environment that the individual develops politically and is best suited to interact in democratic society. When this individualisation reaches a point of atomisation, where the links between individuals and groups are severed, is where democracy reaches a critical point. No longer are democratic societies able to unite effectively unite into a true community, with each individual member only focused on their own interests. With this very essence of the social contract collapsing, as individuals defer to the perceived opinion of the majority, the opportunity for despotism arises. One of the critical drivers behind this atomisation are various kinds of technologies, including modern digital technologies. With varying forms of connected devices in the hands of the vast majority of the population, democratic society has been rendered equal in condition at a level never seen before. This section will explore the relationship that Tocqueville creates between technologies of all kinds and individualisation, highlighting the threats that arise from the continued advancements of different devices.

Tocqueville does not himself explicitly mention the word “technology” in his writings, instead interpreting the world’s attitude towards technological development, analysing its effects on how humans see themselves and the world (Storey, 2013).¹⁰ It is in *Journey to England* where Tocqueville aligns technology with political freedom, questioning whether societies’ laws “give men the courage to seek prosperity, freedom to follow it up, the sense and habits to find it, and the assurance of reaping the benefit” (Tocqueville, 1958, p. 116). The health of a democratic society is thus inextricably attached to their abilities to advance economically by virtue of technological progression. This technological romanticism, where technology is a tool to human perfection (Storey, 2013), does not come without downsides, however. Severe individualisation through specialisation arises from workers dedicating themselves to the production of single objects, (Tocqueville, 1998), degradations which allow Tocqueville’s equality of condition to develop while isolating individuals from the community around them.

That individualism frequently occurs in democratic societies is not foreign, in fact Tocqueville regards it as unavoidable in democratic republics as a consequence of the equality of condition (Villa, 2005).¹¹ Democracy itself is inherently individualising, driving citizens to pursue their self-interests within a material world (Tocqueville, 1998).

10 In a sense, the psychic appeal of technology which convinces democratic society to embrace its transformations in such drastic ways.

11 A general social and historical equality, not one of equal rights or of equal financial standing

Because all individuals are on equal ground, this being Tocqueville's equality of condition, material interests are the only fashion in which to elevate oneself over another, a freedom of movement that can only come from a society free of stable power elites (Tocqueville, 1998, p. 206). The permanence of aristocratic society has "the effect of closely binding every man to several of his citizens" (Tocqueville, 1998, p. 205), a state brought about by the duties and obligations that each individual has to maintaining his family's position within the system, even sacrificing personal gratification to do so. This form of intense attachment is lost in democracies, where "the bond of human affection is extended, but it is relaxed" (Tocqueville, 1998, p. 206), owing significantly to the dissolution of corporate ties that maintain aristocracies (Villa, 2005). The severing of the links between society, Tocqueville argues, comes from the increasing lack of distinction between classes as each individual may rise or fall as they so desire (Tocqueville, 1998, p. 206). This lack of distinction further produces is a misapprehension of total independence by the individual, that they owe nothing to nobody else and control their own destinies (Villa, 2005).

Individualisation and Democratic Backsliding

The risk with individualism is that it spawns selfishness (Tocqueville, 1998, p. 205). The search for self-understanding is the key to intelligent democratic government, as society needs to "teach democracy to know itself" in order for the system "to guide... and discipline itself" (Tocqueville, 2013). The digital age has seen this exact description play out to an even more extreme form than described by de Tocqueville.

Today, connections between individuals are more possible than ever before, as digital platforms connect hundreds, thousands, and even millions of people. The quality of connection between those people, however, is questionable at best (Parkinson, 2009).¹²

Knowledge of the self has transpired through mass data collection, processing, and agglomeration, except the self does not (generally) have access to the collections of data that help us understand ourselves. Instead, it is presented as an extension of the through digital assistants such as web platforms or home devices providing recommendations. This is the threat that is presented by the digital individualisation of the 21st century.

The lack of connection described by Tocqueville plays out in two distinct ways in the developing despotic digital society. In democracies, the exploitation and public virtue he describes is achieved by individual need for ease and efficiency, achieved by data-collecting and aggregating digital technologies. The influence comes from the exploitation of self-interest for profit purposes with no concern for consequences; this creating the space for polarised ideas to take hold and allowing despots the opportunity to emerge. Authoritarian governments use technologies to promote self-censorship, and through developments such as China's Social Credit System, detach communities through social ranking. By promoting a society where violations of the norms are at an individual level, the rulers exert their full control through indifference, as each individual is more concerned about

¹² The 'public space' has long been discussed as necessary and has undergone significant shifts in identity during the digital transition. A neglect of the physical public sphere can have significant effects on the health of democratic society.

their individual ranking without any consideration to the power exerted over them, complicated further by Tocqueville's assertion that man is individually powerless (Tocqueville, 1998, pp. 196–197).

The increasingly individualised society is not a new phenomenon, but one that has been progressing steadily since the mid 1900s, a significant modern factor of which is an increasingly polarised political environment. Polarisation reduces the electorate's ability to punish politicians when their policies undermine democratic norms, since voters put partisanship over values (Graham and Svobik, 2020). This combines with this inherently individualising nature of technology to further reduce the ability for individuals to utilise popular checks. The manners in which the generality now participate in democratic civil society, principally through digital means, have increased polarisation and decreased freedom of choice.

These conditions have aided in the democratic backsliding that has been pervasive across the world since the turn of the century. This has happened alongside a shift from group-based prediction making to individualised prediction and the identification of new groups and subgroups based on novel data compilations as opposed to traditional data grouping strategies (Harcourt, 2015). Such individualised prediction creates power over the individual, which is much easier to control than a full group of individuals (Harcourt, 2015), an environment which gives governments or businesses more control over a population. The reduction in influence of individuals is magnified by the selective and individualising nature of algorithmic filter bubbles, preying upon natural human tendency to seek out what they are most comfortable with and reducing them to only see content that lies, in Tocqueville's terms, within man's own heart.

This extreme individualisation caused by the digital technological revolution, and the effects that come with it, can be seen manifesting in the digital age to a wider group of people. No longer is it restricted to the traditional industrial worker, as all individuals within digital society are “working” for the data collectors; this meaning that every single connected individual is akin to the industrial worker in the 1830s. The constant distraction and simplifications that connected devices bring to society can also call back to Tocqueville’s observations of the societal degradations in the 1830s (Storey, 2013). Tocqueville later stated that laws can no longer protect this individual, “a theory... more powerful than customs and laws has bound him to a craft, and... a place in society that he cannot leave”, leaving him “stationary” (Tocqueville, 1998, p. 254).

The Importance of the Community

The community to Tocqueville is one of the fundamental gateways to democracy, a meaningful way of combatting the effects of individualism. Community in the eyes of Tocqueville specifically refers to a political community, often one formed from engaging in public affairs. In the act of committing to public affairs, to which Tocqueville explicitly states must be done “in public”, individuals begin to notice that they are “not so independent of his fellow man” because “in order to obtain their support, he must often lend them his hand” (Tocqueville, 1998, p. 210). Individualistic behaviours and societies are not necessarily the struggle, but atomisation and the resulting polarisation and alienation that is; individuals must be able to think for

themselves but must not only think of themselves in order for democratic society to avoid despotism.

The community as described by Tocqueville, however, refers not solely to the people within a community, but also to communal knowledge (Bjereld et al., 2009).¹³ Humans aspire to greatness because they are aware of their capacities and their smallness within the world, creating the individual environment for the desire for self-knowledge (Storey, 2013). Stemming from the heightened importance of individual achievement and standing individuals becomes incentivised to seek the reasons of things exclusively by themselves and wholly within themselves (Hebert, 2007). The individual, when ensconced in this circumstance, makes no effort to seek explanation or information from external sources, motivated by a false assumption of extraordinary judgement and capability of the individual (Hebert, 2007).¹⁴

This pervasive individualism has helped foster and ingrain intellectual conformism through the diminishment of meaningful intellectual liberty (Hebert, 2007). Collectivist societies are more cohesive, maintain better social networks, and are better able to overcome problems affecting the society as a whole (Ezcurra, 2021). This contrasts with governments in individualised societies, which Tocqueville noted as being dangerously overcommitted and feebly incapable as the “truly isolated” become increasingly numerous (Storey, 2013). This combination of

¹³ This despite knowledge and education, according to Tocqueville, be-
longing exclusively to an individual

¹⁴ The amount of effort to effectively do this would be overburdensome
on humans, necessitating the existence and acceptance of the moral
authority of religion and philosophy.

overcommitment and incompetence consequently creates a breeding ground for instability and autocracy.

When a society is composed of individuals who are fully independent from each other, “no one is seen to exert a great or still less a lasting power over the community”, depending then solely on voluntary actions (Tocqueville, 1998, p. 194). This community, “devoid of influence, sees individuals passively defer to public opinion in isolation, a public opinion determined by no self in particular” (Tocqueville, 1998, p. 195). The reduction of influence by any individual brought about by democratic equality of condition reduces the reliance that an individual has on those around him, owing “nothing to any man” and expecting “nothing from any man”, leaving them alone to manage their own destinies and confining each individual “entirely within the solitude of his own heart” (Tocqueville, 1998, p. 206).

Local communities have been difficult to define as societies have become increasingly individualised (Nelson, Lewis, and Lei, 2017). Constant and continued digital surveillance and exposure furthers the creation of a non-physical commons, where we interact and create connections not with each other, but with digital constructions of each other. The surveillance of our exposures threatens to create a second surreal-punish society without democratic oversight or beyond legal jurisdictions (Harcourt, 2015).¹⁵ Physical communities have seen their importance diminished in the political sense (Nelson et al., 2017). What has developed in digital participation is a participatory community based on loose individual ties, with the interest of individuals founded principally in the movement, not the individuals in the

¹⁵ This form of surveil-punish society is in reference to Michel Foucault's *Discipline and punish: The birth of the prison*. See Foucault, (2012).

movement. This removes the direct connected community from the civic participation, with people only participating if it serves their personal agenda or feed, creating an engagement where the generality is entirely alone in the physical world while participating in democratic political activity (Nelson et al., 2017).

Younger generations are more active politically online than previous generations, at least visibly active, this despite their reticence to trust national politicians, and showed a majority planned to partake in presidential elections (Statista, 2021). They also form connections through online communities allowing them to follow the political and civic issues that are most important to them (Nelson et al., 2017). Despite this, a scepticism remains due to the nature by which they receive their news. Simple political participation by voting does not inherently indicate a functional democratic society (Schmitter and Karl, 1991).¹⁶ The most impactful of these, as described by Tocqueville, are effective information sharing and contest among competing groups. As these groups may be restricted by their increased reception of the news via bulletin board platforms, the challenges of individualised content streams emerge (Barberá, 2020).

21st Century Threats to Democracy

The threat of equality, and thus by association, individualism, is the opportunity it presents for despots, who exploit the predisposed

¹⁶ Other necessary parts of a functional democracy include a public realm, inclusive citizenship for suffrage, political competition, regular elections, and majority rule.

social detachment and turn the same equality into a public virtue and “raising barriers to keep them asunder” (Tocqueville, 1998, p. 210). Individualism thus creates at one time a dangerously overcommitted and feebly incapable government, as the truly isolated become increasingly numerous (Storey, 2013). The solution to the threat of despotism is not to reduce power down to the individual, however, to prevent central government from abusing its power. This is best done, according to Tocqueville, by drawing as many individuals into the public realm as possible (Tocqueville, 1998, p. 211).

Several interconnected threats to democracy have emerged in the 21st century from the advancement of the digital society, composed of fake news, hate speech, and surveillance (Keuhn and Salter, 2020). Contemporary digital advertising sways public opinion through targeted political advertising while being elevated by microcelebrities who link shocking content and provide visibility and financial value to extreme ideologies. This increased visibility reduces the willingness of different groups from circulating potentially controversial stories or ideas, reinforced by surveillance practices which produce chilling effects (Keuhn and Salter, 2020).¹⁷ Challenges to minority groups are of particular importance owing to the fact that they are often used as scapegoats or otherwise targeted in the “us vs them” nature of polarised populism. Their self-censorship in face of online attacks reduces their willingness or ability to adequately defend themselves (Keuhn and Salter, 2020).

At the heart of these threats lie filter bubble and echo chamber mechanisms, which exacerbate human tendencies to seek out information

¹⁷ Self-censorship

and environments in which they are familiar and comfortable. Filter bubbles occur as people seek information and the algorithm-based browser or platform learns what they consume as well as their tendencies (Dahlgren, 2021; Pariser, 2011). The resulting personalised online ecosystems narrow scopes of information, opinions, and resources for users, Twitter as an example has algorithms that prioritise popular and more recent posts, creating “toxic techno cultures” which subordinate dissenting opinion (Dahlgren, 2021). That the internet can create homogenous online societies is not a new argument, with balkanised digital communities being discussed for as long as the internet has been around (Barberá, 2020). While real-life interactions force diverse connections, filtering technologies and their automated screening processes accelerate the homogenisation of feeds, allowing the creation of like-minded communities that isolate individuals from challenging information (Putnam, 2001).

The continuous process evolving to increasingly personalise information stream makes it difficult for individuals to find information that challenges their beliefs (Dahlgren, 2021; Pariser, 2011). Complicating this are people’s habits of clicking on continuously more polarising material. This makes the individual or, more likely, the algorithm determining the information flows the gatekeepers of information through the “Daily Me” concept (Barberá, 2020).¹⁸ If this feedback loop created by the filter bubble consistently presents this information to a user, this can present to them as the mainstream political belief. In terms of Tocqueville, this can be viewed as a forced deference to the

¹⁸ The “Daily Me” concept was first popularized by Nicholas Negroponte of MIT, and points to the idea that the news shown on digital feeds are tailored to show an individual only the news that they care about.

general ideology. This general ideology that is fabricated develops not out of the sharing of opinion, but of the manipulation of human tendencies.

Filter bubbles have several outcomes with ramifications on democratic society, including overconfidence, large confirmation bias, decreased curiosity, and a skewed vision of the world (Dahlgren, 2021; Pariser, 2011). In the context of democratic society within filter bubbles, pluralism is significantly reduced as one viewpoint becomes the predominant vision to an individual, thus reducing the capacity of pluralism to limit the group's influence. The importance of pluralism of opinion is that it leads to disagreements and conflicts of interest which limit any single individual's or group's ability to influence the overall society (Bellamy and Kröger, 2021). Instead, humans are now deprived of an intimate and private space, areas where you can be free to develop yourself and control that development, as opposed to having your development controlled for you, instead these environments controlled by digitally curated content streams.

Without space come detrimental effects on civic discourse, democracy through the facilitation of intellectual and ideological isolation, preventing shared understanding of contemporary ideas and issues. The reduction of plurality induced by the state of extreme societal atomisation is the essence of individualisation's threat to democracy, as the common will of the people can only materialise in compromises of interest promoted by pluralist settings (Bauer and Becker, 2020). If the individual is deferring more to popular opinion, as expressed on social media through bulletin board posting sites and media presence, the narrower a world viewpoint that they internalise. Personalisation on social media lock an individual into a content bubble, most

specifically as it is the environment in which the individual is interacting creating the bubble. A direct contrast develops here with the environment Tocqueville says is necessary for successful democracy, where the personal interests and self-development are necessarily done in a diverse and pluralist environment. In a system free from algorithmic influence, individuals may have the opportunity to engage with other sources or interact with a group that exists outside the sphere of influence, but when there is no alternative, the remaining opinions become the popular opinion; this can only be reinforced by algorithmic control.

Studies have indicated that deliberation is noted to have seen the most significant deterioration.¹⁹ Technologies have disrupted traditional forms of democratic deliberation and digital deliberation falls short of the modes of deliberation that Tocqueville noted as necessary for effective democracy. Tocqueville proclaimed that “the best-informed inhabitants of a district use their information to discover new truths that may augment the general prosperity”. This is inherently challenged by the nature of technologies, which prey on innate human tendencies to seek out information with which they are already comfortable to trap them within information bubbles (Pariser, 2011; Putnam, 2001; Sunstein, 2018), preventing them from seeking out and finding the information needed to augment their well-being in the manner predicted by Tocqueville. That echo-chambers are identity-securing in nature, tending to seek out information that corresponds to what users already know, can be exploited by political content-reinforcement of information bias, polarising effects (Keuhn and

¹⁹ See for example, Gora and de Wilde (2022), Orhan (2022), Curato (2014), and Coppedge (2017).

Salter, 2020). Majority rule through election is a legitimate and effective mechanism for conflict resolution, but only when done through processes that are via anonymous, neutral, and free from bias (Bellamy and Kröger, 2021), conditions that are innately challenged by algorithmic manipulations.

The good in humans is not inherently private but is contingent upon man's interactions within a structure of others. By failing to exist within this, rationally individualised humans will struggle with self-understanding (Hebert, 2007). Failing at self-understanding makes developing political opinion that much more difficult and allows the individual the opportunity to passively defer to the masses. What we see in present-day democratic society is individuals whose self-understanding is in many ways determined for them and a curated generality with whom to defer opinion to handed on a platter.

The rise of the influence of political representatives over the individual, importantly done through private channels of democratic political engagement, weaken the political structures that maintain horizontal checks and balances of democratic governments. Majority rule through election is a legitimate and effective mechanism for conflict resolution, but only when done through processes that are via anonymous, neutral, and free from bias (Bellamy and Kröger, 2021). Beyond the general issues of filter bubbles raised previously, the black box nature of digital algorithms frequently faces criticisms for incorporating biases, particularly from their training datasets (Caliskan et al., 2017).

Political Instability, Digital Technologies, and Democratic Backsliding

Some studies have associated individualism to political instability or highlight the importance of maintaining a society rife with social links (Annett, 2000; Ezcurra, 2021; Gibson, 2001). The issue is that individualism is merely discussed in terms of political instability and violence more generally, not specifically narrowed down within the contexts of democratic backsliding. This analysis suggests that one of the root causes of recent democratic backsliding may be political instability instigated by increased usage of digital technologies (Ezcurra, 2021). Political instability is a result, whether directly or indirectly, of confrontations between groups where power is unequally distributed (Ezcurra, 2021), unequal distributions that have arisen in the present day due to the power of digital data usage.

Liberalism more generally, but in particular neoliberalism, pushes to make clear distinctions between the public and the private (Harcourt, 2015), a distinction that is inherently difficult in democratic society because of governance by the people. Technology has in a sense flipped that concept, the private is now public, by dissolving the fragmented and illusioned divides between the public and the private and creating a digital agora turning private forums into arenas of public discourse (Ranchordás, 2017). Because of the nature of bulletin board platforms where these agorae exist, these are no longer group discussions, instead they are collections of people discussing issues on an individual level. This results not solely from a market competition on social media, that is to make the most money and competing for audiences,

but also from the general day-to-day competition of individuals to make their lives look as “perfect” as they can because others’ lives seem to be more perfect than their own. Some digital markets, such as subscription and peer-to-peer transaction-based markets, see their most successful workers be the ones who disseminate populist ideologies (Böcskei and Németh, 2022).

Activism plays a critical role in democracies, and digital engagement has certainly played an immense role in the spread of activist campaigns and opened avenues for activist engagement. However, this engagement has gone both ways, from the examples mentioned above, to the disinformation campaigns that have become pervasive and influential in elections since 2015 (Nelson et al., 2017). Politics once meant individuals working together to solve problems through a vibrant public sphere designed for reasoned and justified debate (Habermas, 1991), a space that has become increasingly intangible and out of the reach of ordinary citizens (Nelson et al., 2017). This mould has slowly shifted as civic participation has begun to be redefined (Ranchordás, 2017). Previously, democratic participation would have been explained by family, local community associations, and neighbourhoods.

Today the internet offers new modes of mobilisation, with hashtag campaigns on Twitter playing influential roles in elections and individuals their jobs and highlighting racial injustices within the policing system (Nelson et al., 2017). Digital forms of democratic engagement empower citizens in several different ways, providing wider public attention, more easily evading censorship, and facilitating the coordination and logistics of public movements (Tufekci, 2014). The result of this is a convergence of the digital and the social, with the digital

platforms influencing or even producing the social structures that exist today (Poell and van Dijck, 2015; Sorice and Blasio, 2019). Digital medias have shifted the temporal expectations of 21st Century social structures, with platforms of all kinds continuously updating and sharing information for users (Coleman, 2020). This instantaneity within digital activism is in fact a crucial component, allowing both activist groups and individuals to share photos and resources or in disseminating communications with other engaged parties (Melki and Mallat, 2014).

Though the instantaneousness of digital democracy can help developed more responsive and devolved forms of democracy (Kuo, 2019), the need for immediate results pulls into tensions with even the most modern conceptions of democratic legitimacy (Rosanvallon, 2011). Online activist campaigns are popular because their results can be immediately seen (Nelson et al., 2017), directly contrasting with traditional democratic structures, where temporalities are much slower and spread out, with results needing months or even years to play out. Other factions of democratic government, such as judiciary branches of government, are not designed to be responsive to public opinion as compared to the legislative and executive branches of government (Kuo, 2019), while all three face similar movements towards open and transparent governance. Digital political engagement is further challenged as governments learn to respond to new mediums of engagement, done often through restricting information flows or making the space challenging to interact with through the use of paid users (Tu-fecki, 2014), these coming prominently as populist and anti-democratic political actors make use of the digital platforms.

Ever since EU enlargement in 2004, the EU has faced growing threats of democratic backsliding, beginning when Fidesz questioned the legitimacy of the 2006 elections (Bakke and Sitter, 2022).²⁰ Three aspects of democratic backsliding are generally at play in EU discourse, complications with the rule of law from judges and EU institutions and deterioration of deliberation and citizen disengagement from different groups of scholars (Gora and de Wilde, 2022). The EU's fundamental failures also play a role, as the bloc lacks the adequate measures to enforce compliance. Even where EU treaties do contain compliance mechanisms, triggering them has prompted defiance from governments to push back and further backsliding (Holesch and Kyriazi, 2022).²¹ Finalising this requires unanimity of all member states except the accused, which became a challenge as Hungary and Poland both defended each other during debates in the European Parliament (Holesch and Kyriazi, 2022). Legitimation of backsliding came importantly from having a key ally in each other, with Orban and sequential Polish leaders routinely serving as the other's partner in European affairs (Holesch and Kyriazi, 2022).

²⁰ Fidesz lost the 2006 parliamentary elections and immediately challenged their legitimacy.

²¹ Backsliding states within the EU can offer each other mutual protection via vetoes against the initiation of Article 7 TEU procedures, which may be undertaken should there be a consistent breach in values named in Article 2.

One particularly interesting case of backsliding is that of Hungary, as it has endured relatively peaceful conditions of backsliding since 2010 against the backdrop of their position within the democratic liberal–normative European Union. This form of progressive backsliding can be analysed as a case of electoral authoritarianism, with leading parties creating uneven electoral competition (Bernhard, 2021). Hungary was at one point a leader in 21st Century democratisation, but has seen continuous centralisation of powers, electoral field manipulation, media takeovers, and assaults on civil society groups by successive Viktor Orban governments (Rupnik, 2023).

Hungary's Fidesz party, led by Viktor Orban, took power in 2010 in the midst of an economic recession and a left coalition mired in scandal, and secured continued supermajorities through 2020. This near decade and a half of sequential Fidesz governments has allowed legislation that undermines democratic quality (Holesch and Kyriazi, 2022). Party rhetoric often takes an anti-West stance, nationalist, and xenophobic messaging, which only increased due to conflicts with the European Commission over border controls during the 2015 immigration crisis (Bernhard, 2021). Major shocks to the system surrounding the beginning of the last decade, most prominently the Great Recession and subsequent Euro Crisis and then the European Refugee Crisis of 2015 re-ignited nationalist agendas and arguments against the liberal global market, allowing Fidesz to gain strength (Bernhard, 2021). The nationalist rhetoric was particularly critical to boosting Orban in 2016, where a referendum on tightening immigration was influential in reversing a mid-crisis drop in polls (Rupnik, 2023).

The steady and continuous backsliding in Hungary since 2010 has been influenced by executive aggrandizement (Holesch and Kyriazi,

2022). Fidesz continued condensing powers in the Parliament, Constitutional Court, and several other state institutions while redrawing constituency boundaries in an attempt to diminish social accountability (Bernhard, 2021). Orban has committed to creating an “illiberal democracy”, the construction of which creates an uncertain role for politically neutral institutions (Rupnik, 2023). Along with this have come provisions curtailing “consultation, debate, and scrutiny” of new legislation within the parliament (Bakke and Sitter, 2022). As the Hungarian system more easily converts pluralities into constitutional majorities, the use of which has been used to reduce constraints via constitutional reforms, Fidesz faces fewer horizontal constraints on its power (Bernhard, 2021). The combination of the decreased importance of impartial institutions combined with the inherent weak points of the Hungarian system have presented the milieu for Fidesz to maintain its stranglehold on power.

Hungary and Fidesz’ use of digital media

Alt-right and extremist organisations and views, including political parties, are able to cut out the media middlemen by using unregulated, money-driven platform providers to bypass censorship and marginalisation from traditional media (Böcskei and Német, 2022). This has allowed digital bulletin board platforms to become influential in helping several populist political groups to gain ground, including the English Defence League, the Five Star Movement in Italy, and the Fidesz party in Hungary (Grabbe, 2019). The media, even traditional forms, in Hungary have faced continued attacks and struggled to

maintain their independence since the fall of communism. Though this pressure has continuously come from both sides of the political divide, Fidesz has taken steps since 2010 to oversee, control, and regulate the media out of line with democratic norms (Wilkin, Dencik and Bognár, 2015); these illiberal steps have since been extended to more modern technologies.

These platforms instead become information hubs, with Hungarian parties using platforms as part of a top-down communication strategy for party initiatives and positions (Bene, Farkas, and Burai, 2021). Taking advantage of this, Hungary's Jobbik party, a far-right party often in opposition to Fidesz, quickly emerged as a leader in the Hungarian political scene on social media usage (Böcskei and Német, 2022). Other opposition parties also took advantage of digital platforms as they were pushed out of traditional media platforms (Centre for Media Pluralism and Media Freedom, 2022). Fidesz, in reaction, has developed a vast government media campaign with government agencies making large investments into social media (Centre for Media Pluralism and Media Freedom, 2022). Digital platforms, prominently Facebook, have become central hubs for political information in Hungary, incentivising political parties to pay increase their attention to their presence on these platforms (Bene et al., 2021). The influence of Fidesz within these digital platforms is clear, with the party receiving significantly more 'likes' and 'shares' than the next closest party (Bene et al., 2021).²²

The influence of Fidesz within the social media process begins early on, with influencers having the ability to be trained by the Fidesz

²² According to this study, Fidesz received a median of 870 shares per post, compared to the Democratic Coalition's 174 as the next closest, and 4052.5 likes per post, compared to DK's 515.

linked Megafon organisation (Balogh, 2021; Ozsvath, 2021).²³ Pro-government influencers, sponsored by the Fidesz party, continue to amplify conservative voices through social media platforms already used by Hungarians, in addition to party dominance over the traditional media (Böcskei and Németh, 2022). Ahead of the 2022 elections alone, around \$2.5 million was spent by Megafon on these influencers pushing Fidesz government opinions and ideologies (Rutai, Bhat, Shah, and Mambondiyani, 2022). It is interesting to note this special effort made towards promoting messages using social media. The trend is that people are engaging more and more with modern methods of receiving and processing information, especially in Hungary where 60% of Hungarians getting their news from Facebook (Rutai et al., 2022). Where Fidesz differentiates themselves from their competition is their use of these platforms for political messaging, devoting more than half of their posts to policy issues as other parties devote less of this digital space to this kind of messaging (Bene et al., 2021).²⁴

Despots find that it is in their “interest to render... subjects equal and to leave them ignorant, in order more easily to keep them as slaves” (Tocqueville, 1998, p. 190). A society left believing one constant story, such as the one told through Hungary’s state media policy, leaves it more easily manipulable and open to despotism. Within this equality of knowledge, citizens see their aptitudes for cultural and scientific thinking, which Tocqueville argues are crucial for individuals to escape their present situations and engage with diverse information types.

²³ This is based on reports from both NGOs and news sources.

²⁴ Findings in this study showed opposition parties focusing heavily on traditional offline mobilisation efforts than Fidesz, this perhaps in an attempt to find some way of mobilising support as they continue to struggle with digital engagement due to issues accessing these spaces.

Platforms of “public” debate create level playing field for debate and discourse, allowing for individuals to directly interact with each other outside of the traditional spectrums of political discourse. This further breakdown of conditional hierarchy continues the trend of individualisation that Tocqueville noted as an inherent part of democracy, leaving the individuals even more manipulable than before. The complication of public action online is that deliberation is most often done on platforms that are privately owned, a situation particularly complicated in Hungary’s situation. Bulletin board platforms such as Facebook and Twitter, among others encourage the development of political communities, but inherently do so within the private spectrum and complicate the necessity to perform public duties in the public. This draws significant tensions when public bodies influence and utilise these platforms as their main forums for communication.

Conclusion

While significant research has taken place on the disruptive nature of technologies in our societies, we still have only a partial sense of how these technologies tie into the theoretical and practical imaginations of democratic society. The threats that Tocqueville presents are no less prevalent today than they were when he was writing *Democracy in America* in the 1830s, with the alarm bells every day being sounded from every corner of the Earth. To many, our societies today are in many respects more disconnected and disjointed than ever before, threatening the democracies that many take for granted.

Digital technologies have from the start been presented as a way of creating more functional democratic systems as all individuals in a society would be able to interact more freely with information and with their representatives. While digital platforms have aided in groups envisioning new forms of democratic governance, their potential is roadblocked by how information is distributed on social media. In Western digital spaces, private platforms control the information flow, holding free-flowing news and accuracy hostage for financial gains. Governments and their institutions can then take advantage of this by playing the algorithm game to their benefit, as seen in the case of Hungary. There are two possible reasons that governments are focusing on social media marketing, either that people are consuming more and more media on social media platforms over traditional forms of media (i.e., newspaper, television, radio, etc.), or that government officials have identified social media as the easiest to manipulate in order to have their message be the dominant voice.

The filter bubble trap keeps users, even those that are politically aware and active, stuck in a cycle of information that prevents them from seeing the full scope of their reality. Instead of a full truth, individuals are presented with only the truth the algorithm deems profitable, preying on human tendency to seek out information they are most comfortable with. Despite the research presented and synthesised regarding filter bubbles, their true effects are still subject to significant debate and this trap requires further significant research.

Evidence suggests that the potential for a relationship between increasing technology usage and decreasing voting habits in addition to the increasing distrust of the media as independent from government sources (Keuhn and Salter, 2020). This possibility, in

combination with the thoughts presented in this paper, lead to the notion that the role that technology plays in legitimating and intensifying factors leading to backsliding across the world. This research must be furthered by discussions on the true impacts of filter bubbles. Additional discussions on how to quantify the effects of digital technologies on democratic process are also, as traditional indicators fail to explicitly account for this. Subsequent research may also explore the impacts of silent national interests in data gathering platforms, concerns raised by the recent revelations during the TikTok hearings in the United States.

Without truly identifying the how, no significant proposals to counteract further societal atomisation can be made. One thing is for sure, that the limitation of the impact of algorithms on content in digital spaces is essential for maintaining true democratic process as it has been imagined. Failures to ensure this will only lead further to the rise of despotism, whether in the traditional sense as remarked by Tocqueville, or in new ways not yet imagined.

I thank Susan Perry, Claudia Roda, and Stephen Sawyer for their valuable comments and insight on this paper.

References

- Abdullah, W. J. (2020). "New normal" no more: Democratic backsliding in Singapore after 2015. *Democratization*, 27(7), 1123–1141. <https://doi.org/10.1080/013510347.2020.1764940>
- Annett, A. (2000). IMF working paper: Social fractionalization, political instability, and the size of government. *International Monetary Fund*. <https://www.imf.org/external/pubs/ft/wp/2000/wp0082.pdf>
- Bakke, E., & Sitter, N. (2022). The EU's enfants terribles: Democratic backsliding in Central Europe since 2010. *Perspectives on Politics*, 20(1), 22–37. <https://doi.org/10.1017/S1537592720001292>
- Balogh, E. S. (2021, October 20). The Fidesz propaganda machine gears up against Marki-Zay. *Hungarian Spectrum*. <https://hungarianspectrum.org/2021/10/20/the-fidesz-propoganda-machine-gears-up-against-marki-zay/>
- Barberá, P. (2020). Social media, echo chambers, and political polarization. In N. Persily & J. Tucker (Eds.), *Social media and democracy: The state of the field, prospects for reform* (pp. 34–55). Cambridge University Press.
- Bates, G., Cinar, I., & Nalepa, M. (2020). Accountability by numbers: A new global transitional justice dataset (1946–2016). *Perspectives on Politics*, 18(1), 161–184. <https://doi.org/10.1017/S1537592719000756>
- Bauer, M. W., & Becker, S. (2020). Democratic backsliding, populism, and public administration. *Perspectives on Public Management and Governance*, 3(1), 19–31. <https://doi.org/10.1093/ppmgov/gvz026>
- Bellamy, R., & Kröger, S. (2021). Countering democratic backsliding by EU member states: Constitutional pluralism and 'value' differentiated integration. *Swiss Political Science Review*, 27(3), 619–636. <https://doi.org/10.1111/spsr.12448>
- Bene, M., Farkas, X., & Burai, K. (2021). Same strategy, but different content: Hungarian parties' Facebook campaign during the 2019 EP election. In J. Haßler, M. Magin, U. Russmann, & V. Fenoll (Eds.), *Campaigning on Facebook*

in the 2019 European Parliament election (pp. 119–134). Springer International Publishing. https://doi.org/10.1007/978-3-030-73851-8_8

Bermeo, N. (2016). On democratic backsliding. *Journal of Democracy*, 27(1), 5–19. <https://doi.org/10.1353/jod.2016.0012>

Bernhard, M. (2021). Democratic backsliding in Poland and Hungary. *Slavic Review*, 80, 585–607. <https://doi.org/10.1017/slr.2021.145>

Bimber, B. (2014). Digital media in the Obama campaigns of 2008 and 2012: Adaptation to the personalized political communication environment. *Journal of Information Technology & Politics*, 11, 130–150. <https://doi.org/10.1080/19331681.2014.895691>

Bjereld, U., Ekengren, A.-M., & Schierenbeck, I. (2009). Power, identity, modernity: Individualisation and destabilisation in a globalised world. *International Review of Sociology—Revue Internationale de Sociologie*, 19, 263–272. <https://doi.org/10.1080/03906700902833577>

Böcskei, B., & Németh, S. (2022). From unknown subculture to political mainstream: The American alt-right and the Hungarian right. *The Interdisciplinary Journal of Populism*, 3, Article 3. <http://real.mtak.hu/155830/>

Brown, W. (2019). The personal, protected sphere must be extended. In *In the ruins of neoliberalism: The rise of antidemocratic politics in the West* (pp. 89–122). Columbia University Press. <http://www.jstor.org/stable/10.7312/brow19384.7>

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183–186. <https://doi.org/10.1126/science.aal4230>

Bátorfy, A., Bleyer-Simon, K., Szabó, K., & Galambosi, E. (2022). Monitoring media pluralism in the digital era: Application of the Media Pluralism Monitor in the European Union, Albania, Montenegro, The Republic of North Macedonia, Serbia & Turkey in 2021. *Centre for Media Pluralism and Media Freedom, Robert Schuman Centre*. <https://cadmus.eui.eu/bitstream/handle/1814/74692/MPM2022-Hungary-EN.pdf>

- Coleman, R. (2020). Making, managing and experiencing 'the now': Digital media and the compression and pacing of 'real-time'. *New Media & Society*, 22(9), 1680–1698. <https://doi.org/10.1177/1461444820914871>
- Coppedge, M. (2017). *Eroding regimes: What, where, and when?* V-Dem Working Paper 2017:57. <http://dx.doi.org/10.2139/ssrn.3066677>
- Curato, N. (2014). Participation without deliberation. *Democratic Theory*, 1(2), 113–121. <https://doi.org/10.3167/dt.2014.010212>
- Dahlgren, P. M. (2021). A critical review of filter bubbles and a comparison with selective exposure. *Nordicom Review*, 42(1), 15–33. <https://doi.org/10.2478/nor-2021-0002>
- Ezcurra, R. (2021). Individualism and political instability. *European Journal of Political Economy*, 66, Article 101959. <https://doi.org/10.1016/j.ejpoleco.2020.101959>
- Gibson, J. L. (2001). Social networks, civil society, and the prospects for consolidating Russia's democratic transition. *American Journal of Political Science*, 45(1), 51–68. <https://doi.org/10.2307/2669359>
- Gora, A., & de Wilde, P. (2022). The essence of democratic backsliding in the European Union: Deliberation and rule of law. *Journal of European Public Policy*, 29(3), 342–362. <https://doi.org/10.1080/13501763.2020.1855465>
- Grabbe, H. (2019). Populism: What does it mean for Europe, our societies, and our minds? *Faith in Europe*. <https://faithineurope.org.uk/wp-content/uploads/2020/08/grabbe.pdf>
- Graham, M. H., & Svolik, M. W. (2020). Democracy in America? Partisanship, polarization, and the robustness of support for democracy in the United States. *American Political Science Review*, 114(2), 392–409. <https://doi.org/10.1017/S0003055420000052>
- Habermas, J. (1991). *The structural transformation of the public sphere*. MIT Press.

Haggard, S., & Kaufman, R. (2021). The anatomy of democratic backsliding. *Journal of Democracy*, 32(4), 27–41. <https://doi.org/10.1353/jod.2021.0050>

Harcourt, B. (2015). *Exposed: Desire and disobedience in the digital age*. Harvard University Press. <https://scholarship.law.columbia.edu/books/108>

Hebert, L. J. (2007). Individualism and intellectual liberty in Tocqueville and Descartes. *The Journal of Politics*, 69(2), 525–537. <https://doi.org/10.1111/j.1468-2508.2007.00548.x>

Holesch, A., & Kyriazi, A. (2022). Democratic backsliding in the European Union: The role of the Hungarian-Polish coalition. *East European Politics*, 38(1), 1–20. <https://doi.org/10.1080/21599165.2020.1865319>

Issenberg, S. (2013). How President Obama's campaign used big data to rally individual voters. *Technology Review*, 116, 38–49.

Kaufman, R. R., & Haggard, S. (2019). Democratic decline in the United States: What can we learn from middle-income backsliding? *Perspectives on Politics*, 17(2), 417–432. <https://doi.org/10.1017/S1537592718003377>

Kuehn, K., & Salter, L. (2020). Assessing digital threats to democracy, and workable solutions: A review of the recent literature. *International Journal of Communication*, 14, 2589–2610.

Kuo, M.-S. (2019). Against instantaneous democracy. *International Journal of Constitutional Law*, 17(2), 554–575. <https://doi.org/10.1093/icon/moz029>

Langford, A. T., Solid, C. A., Scott, E., Lad, M., Maayan, E., Williams, S. K., & Seixas, A. A. (2019). Mobile phone ownership, health apps, and tablet use in US adults with a self-reported history of hypertension: Cross-sectional study. *JMIR mHealth and uHealth*, 7(1), Article e12228. <https://doi.org/10.2196/12228>

Lee, J. (2018). Understanding authoritarian resilience and countering autocracy promotion in Asia. *Asia Policy*, 13(4), 99–122. <https://www.jstor.org/stable/26533132>

- Lynch, K., & Kalaitzake, M. (2020). Affective and calculative solidarity: The impact of individualism and neoliberal capitalism. *European Journal of Social Theory*, 23(2), 238–257. <https://doi.org/10.1177/1368431018786379>
- McCoy, J., Rahman, T., & Somer, M. (2018). Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities. *American Behavioral Scientist*, 62(1), 16–42. <https://doi.org/10.1177/0002764218759576>
- Melki, J., & Mallat, S. (2014). Digital activism: Efficacies and burdens of social media for civic activism. *Arab Media & Society*, 19, 1–19.
- Nalepa, M. (2021). Transitional justice and authoritarian backsliding. *Constitutional Political Economy*, 32(3), 278–300. <https://doi.org/10.1007/s10602-020-09315-5>
- Nelson, J. L., Lewis, D. A., & Lei, R. (2017). Digital democracy in America: A look at civic engagement in an internet age. *Journalism & Mass Communication Quarterly*, 94(1), 318–334. <https://doi.org/10.1177/1077699016681969>
- Orhan, Y. E. (2022). The relationship between affective polarization and democratic backsliding: Comparative evidence. *Democratization*, 29(4), 714–735. <https://doi.org/10.1080/13510347.2021.2008912>
- Ozsvath, S. (2021, June 2). Hungary's Fidesz party seeks to conquer social media. *Deutsche Welle*. <https://www.dw.com/en/hungarys-fidesz-party-seeks-to-conquer-social-media/a-56479198>
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Press.
- Parkinson, J. R. (2009). Does democracy require physical public space? In R. Geenens & R. Tinnevelt (Eds.), *Does truth matter?* (pp. 131–148). Springer. https://doi.org/10.1007/978-1-4020-8849-0_8
- Pérez-Liñán, A., Schmidt, N., & Vairo, D. (2019). Presidential hegemony and democratic backsliding in Latin America, 1925–2016. *Democratization*, 26(4), 606–625. <https://doi.org/10.1080/13510347.2019.1566321>

- Perrin, A. (2021, June 3). Mobile technology and home broadband 2021. *Pew Research Center: Internet, Science & Tech*. <https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/>
- Poell, T., & van Dijck, J. (2015). Social media and activist communication. In C. Atton (Ed.), *The Routledge companion to alternative and community media* (pp. 527–537). Routledge. <https://doi.org/10.2139/ssrn.2624911>
- Putnam, R. D. (2001). *Bowling alone: The collapse and revival of American community*. Simon & Schuster.
- Ranchordás, S. (2017). Digital agoras: Democratic legitimacy, online participation, and the case of Uber-petitions. *The Theory and Practice of Legislation*, 5(1), 31–54. <https://doi.org/10.1080/20508840.2017.1279431>
- Riaz, A. (2021). The pathway of democratic backsliding in Bangladesh. *Democratization*, 28(1), 179–197. <https://doi.org/10.1080/13510347.2020.1818069>
- Rosanvallon, P. (2011). *Democratic legitimacy: Impartiality, reflexivity, proximity*. Princeton University Press.
- Rupnik, J. (2023). Illiberal democracy and hybrid regimes in East-Central Europe. In “*Illiberal democracies*” in *Europe: An authoritarian response to the crisis of liberalism* (pp. 9–16). <https://hal-sciencespo.archives-ouvertes.fr/hal-03980745>
- Rutai, L., Bhat, M., Shah, M., & Mambondiyani, A. (2022). Radical timelines. *Index on Censorship*, 51(4), 42–45. <https://doi.org/10.1177/03064220221144893>
- Schmitter, P. C., & Karl, T. L. (1991). What democracy is... and is not. *Journal of Democracy*, 2(3), 75–88. <https://doi.org/10.1353/jod.1991.0033>
- Sorice, M., & De Blasio, E. (2019). Platform politics in Europe: E-democracy and digital activism. *International Journal of Communication*, 13, 5715–5733. <https://ijoc.org/index.php/ijoc/article/view/10807>
- Statista. (2021, March). Youth voter turnout in presidential elections in the United States from 1972 to 2022. <https://www.statista.com/statistics/984745/youth-voter-turnout-presidential-elections-us/>

Statista. (2023, April). Number of Facebook users in Hungary from September 2018 to March 2023. <https://www.statista.com/statistics/1029770/facebook-users-hungary>

Storey, B. (2013). Tocqueville on technology. *The New Atlantis*, 40, 48–71. <https://www.jstor.org/stable/43152771>

Sunstein, C. R. (2018). *#Republic: Divided democracy in the age of social media*. Princeton University Press. <https://doi.org/10.2307/j.ctv8xnhtd>

Svolik, M. W. (2019). Polarization versus democracy. *Journal of Democracy*, 30(3), 20–32. <https://doi.org/10.1353/jod.2019.0039>

Tocqueville, A. de. (1958). *Journeys to England and Ireland* (K. P. Mayer, Ed.; G. Lawrence & K. P. Mayer, Trans.). Yale University Press. (Original work published 1833 & 1835).

Tocqueville, A. de. (1998). *Democracy in America* (P. Renshaw, Ed.; H. Reeve, Trans.). Wordsworth Editions Limited. (Original work published 1835 & 1840).

Tocqueville, A. de. (2013). Appendix 3: Letter from Alexis de Tocqueville to Silvestre de Sacy. In *Tocqueville: The aristocratic sources of liberty* (pp. 335–336). Princeton University Press. <https://doi.org/10.1515/9781400846726-022>

Tucker, J. A., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature (SSRN Scholarly Paper No. 3144139). <https://doi.org/10.2139/ssrn.3144139>

Tufekci, Z. (2014). Social movements and governments in the digital age: Evaluating a complex landscape. *Journal of International Affairs*, 68(1), 1–18. <https://www.jstor.org/stable/24461703>

Villa, D. (2005). Hegel, Tocqueville, and ‘individualism’. *The Review of Politics*, 67(4), 659–686. <https://www.jstor.org/stable/25046471>

Kruse, S., Brunkert, L., & Welzel, C. (2019). A tale of culture-bound regime evolution: The centennial democratic trend and its recent reversal. *Democratization*, 26(3), 422–443. <https://doi.org/10.1080/13510347.2018.1551192>

Wilkin, P., Dencik, L., & Bognár, É. (2015). Digital activism and Hungarian media reform: The case of Milla. *European Journal of Communication*, 30(6), 682–697. <https://doi.org/10.1177/0267323115595528>

Wilkinson, M. A. (2019). Authoritarian liberalism in Europe: A common critique of neoliberalism and ordoliberalism. *Critical Sociology*, 45(7–8), 1023–1034. <https://doi.org/10.1177/0896920519837325>

World Bank. (n.d.). Hungary. <https://data.worldbank.org/country/HU>

Emotional Dynamics in Online Videos: Integrating Machine Learning and Content Analysis

Sara Major
Aleksandar
Tomašević

Introduction

The introduction of television as a medium for mass communication in the 20th century represented a massive shift in the way political information is produced, disseminated, and evaluated. Television was the first channel of communication that primarily relied on the audio-visual format rather than text (Grabe and Bucy, 2009) which, in return, lent a great deal of importance to the image of the leader and emphasized their role at the expense of the political party. This trend in personality-based political communication was only exacerbated in recent years by the rise of the Internet, especially Web 2.0 technologies,

Sara Major, Teaching Assistant, Department of Sociology, Faculty of Philosophy, University of Novi Sad, Novi Sad, ORCID: 0009-0008-7901-5489, email: sara.mejdzor@gmail.com

Aleksandar Tomašević, Assistant Professor, Department of Sociology, Faculty of Philosophy, University of Novi Sad, Novi Sad ORCID: 0000-0003-4863-6051, email: atomashevic@ff.uns.ac.rs

DOI: <https://doi.org/10.5281/zenodo.14203655>

as social media now allows public figures to bypass political parties as intermediaries altogether and interact directly with the electorate (Garzia, 2017). Politicians in this way become primary signifiers of a particular movement or ideology (Mendonça and Caetano, 2021) and can proliferate their political messages freely without the constraints of established structures such as the mainstream media (Lalancette and Raynauld, 2019; Moffitt, 2022). With this personalization of political discourse comes an increase in emotionalization (Mazzoleni, 2014; Udris and Lucht, 2014), too, especially with the rise of contentious populist movements which employ a more affective and divisive political style (Bonansinga, 2020; Norris and Inglehart, 2019; Salmela and von Scheve, 2018).

Because political content online is now more prevalent than ever, it can have very tangible effects on voter behavior and election outcomes (Fujiwara et al., 2021; Zhuravskaya et al., 2020), and it is thus an incredibly valuable area of research. Understanding the characteristics of affective and personalized communication between political leaders and the electorate—be it face-to-face and mediated by a screen—can, by extension, help researchers understand the ways voters are mobilized into action and how political support for a particular leader is consolidated (Sullivan and Masters, 1988).

Most studies of this phenomenon today are divided along the verbal-nonverbal line, i.e., they are most often either text-based (Ekström et al., 2018; Haselmayer and Jenny, 2017; Widmann, 2021) or visual (Mendonça and Caetano, 2021; Salojärvi, 2019; Stewart and Ford Dowe, 2013), rarely both, although there are some notable exceptions (e.g. Bucy et al., 2020). To be able to fully conceptualize the performance of politics in media, it is necessary to consider all elements of

communication, including the verbal, tonal and non-verbal elements of political discourse. Still, this is easier said than done, as methods for analyzing these different elements vary widely in their effectiveness and accessibility. Methods and tools for text analysis are, for example, rather advanced, while visual methods are not nearly as well developed. Traditional methods for the analysis of non-verbal behavior in video, particularly facial expressions of emotion (e.g. Ekman and Friesen, 1976; Masters et al., 1986), are generally time-consuming and only allow for analysis of a limited amount of data as they require manual frame-by-frame coding done by highly trained researchers. These conditions make it significantly harder to conduct exhaustive research on one subject, often leading to the fragmentation of research into smaller, more comprehensible sections (i.e., only verbal, or only non-verbal components).

The field of machine learning, however, may remedy the limitations linked to visual research and bring it up to speed with its text-based counterpart, as it offers researchers the possibility to automatically and efficiently process large quantities of video footage (Walker and Boyer, 2018), which makes the prospect of conducting extensive far more likely and accessible. The present study attempts to illustrate one such method. We apply a machine learning model for emotion detection to a single YouTube video of a political figure's public address and demonstrate how it can be utilized to detect change points in the affective dynamics of speech. The model returns a time-series of seven basic facial expression proportions (anger, disgust, fear, happiness, sadness, surprise, and neutrality) for the length of the entire video, after which various change point detection methods are applied to pinpoint key moments when non-verbal emotional shifts in the address occur. These key moments are then located in a transcription of

the video and contextualized with the verbal content of the speech, which shows that the change points are not arbitrary moments in time but rather reflect rhetorical switches in the address. We showcase how this model can provide an outline for more robust quantitative and qualitative analysis in the future, and thus may benefit researchers looking to conduct broad-scope research on the non-verbal and verbal elements of political communication.

Data

Data used in our analysis comes from a single YouTube video showing a public address of the president of Serbia, Aleksandar Vučić, posted on his official YouTube channel (Вучић: Три Пута Размислите До 30. Септембра, 2022). The video is titled “Think again before September 30th” (in Serbian: “Tri puta razmislite do 30. septembra”) and the length of the video is 25 minutes and 19 seconds.

The video captures a press conference that took place after the president’s meeting with the EU Special Representative for the Belgrade–Pristina Dialogue, Miroslav Lajčák, in June of 2022, which the latter chose not to be a part of. We have selected this video due to its delicate subject matter (that is, Kosovo–Serbia relations) and the familiar format of the president’s mode of public communication, where a lengthy monologue is followed by questions both from independent and pro-government journalists. Due to the length of the address and its emotionally charged topic, we expect the time-series analysis to reflect affective changes in the president’s non-verbal emotional expressions following tonal shifts in the speech itself.

After an emotion detection procedure is applied to the video, we filter only the results related to President Vučić's facial expression. The resulting processed data set consists of 7 time series, one for each score of basic emotion (anger, disgust, fear, happiness, sadness and surprise) and one for scores of neutral expression. Each score is between 0 and 1 and in each frame these 7 scores sum up to 1. These time series are then used as input for change point detection methods.

The emotion detection part of the analysis was performed in Python and the code is available at: <https://github.com/atomashevic/face-of-populism>. The change point detection part of the analysis was performed in R programming language and the code notebook, as well as the processed data are available at: <https://github.com/atomashevic/emotions-change-point>.

Emotion Detection

The emotion detection methodology applied here is based on our previous work (Major and Tomašević, 2023). We select 300 uniformly distributed frames of a YouTube video and process them using machine learning methods. The first task of the analysis is face detection, which is based on a convolutional neural network (CNN) approach (Zhang et al., 2016). Once this algorithm detects one or more faces in a selected frame of the video, a second machine learning procedure is applied with an emotion detection task for each facial expression.

Our emotion detection approach relies on an existing solution—a pre-trained CNN based on the mini-Xception architecture (Arriaga et al.,

2017), which performs a task aimed at detecting six basic facial expressions of emotion as well as a neutral expression. This architecture is suitable for the analysis of still frames extracted from video material because it was originally developed for real-time emotion detection from video input for robotic applications. It has a 66% accuracy rate on the FER2013 dataset (Goodfellow et al., 2013) which is a standard benchmark for emotion detection.

The output of this analysis on a single video frame is an array of 7 emotion scores, each representing the score of one expressed emotion. Each score is matched with one of 7 labels: “angry”, “disgust”, “fear”, “happy”, “sad”, “surprise” and “neutral”. Figure 1 shows examples of processed frames with labels and scores for each emotion for every facial expression.



Figure 1. Examples of processed frames. The left frame shows a highly neutral facial expression, while the right frame shows a high aggregate score of negative emotions, with fear being the dominant one.

When results from all frames are taken together for each of the 7 scores we have a time series of scores. All time series are indexed by

the ordinal number of the processed frame, starting from 1 (first selected frame). For the sake of clarity, in Figures 2 and 3 the indexing of time series and the x-axis are converted to minutes of the video, based on the fact that the video shows 25 frames per second and that we have uniformly selected total of 300 frames, where the subject's face was detected in a total of 268 frames.

Change Point Detection

Upon extracting the time series data, we proceeded with the quantitative analysis based on change point detection methods. This collection of methods is used to detect frames of the video in which a significant change in emotional expression occurs. We were interested in selecting the frames that we can say, with some degree of certainty, represent points in time where a change in some statistical property of the time series happens. That can refer to a change in the mean of the time series. For example, if we have a video where the average sum score for negative emotions is 0.2 in the first 20 frames, and in the next 20 frames we have the total average (of 40 frames) rise to 0.75, then we can assume a significant moment with a sharp increase in negative emotional expression occurred somewhere between the 20th and 40th frame.

The same principle holds if the variance of the time series changes, in which case we could observe, for example, the change from stable emotional expression—where we have a neutral expression score varying from 0.3 to 0.6 for the first 20 frames—to an unstable series of emotional expressions—with higher range and variance of the neutral

expression, ranging from 0.9 to 0.1 in the next 20 frames. Again, we expect that there are one or more points between the 20th and 40th frame where the pattern of emotional expression meaningfully changes.

A slightly more complex case of change point detection occurs when analyzing multivariate time series data. In the abovementioned examples, either the sum score of negative emotions or the time series of neutral expressions is processed for a single time series. We can thus process all seven time series (six basic emotions and a neutral expression) at the same time and detect changes in covariance of individual series, indicating the points in time where the relationship between different emotions changes. This type of change point detection method gives us detailed insight into the affective dynamics of a video, and time points may signify the moments in a public speech where switches between dominant emotions happen. For example, it allows us to pinpoint moments where the subject (e.g. a politician) shifts from a section of speech characterized by expressional switches from anger to fear, to a section which is more characterized by switches from sadness to a neutral expression.

Existing methods for change point detection can be classified into three groups. The first group consists of the simplest methods, capable only of univariate time series analysis, which return change points based on changes in the mean and variance of a single series of data points. The second group is exclusively focused on multivariate analysis and the analysis of time series covariance. The third group is the most flexible and allows change point analysis on both univariate and multivariate data.

As with other methods of statistical analysis, there are differences between frequentist and Bayesian methods. Frequentist change point

detection methods return a list of time points where statistically significant changes in some property of the time series has occurred. The notion of statistical significance is usually based on statistical tests or alternative approaches such as regularization. On the other hand, Bayesian methods return a probability distribution defined over a set of all time points which stores the probability that any point is a change point. Results of the change point detection are inferred from such a probability distribution.

Because change point detection methods haven't been applied to the type of emotion time series we are using in this paper, we apply a mix of different methods from all categories. Our analytical strategy is to go with the broadest set of change points detected with different methods and then to proceed with the qualitative analysis. Three different methods are included in three different R packages: Bayesian change point detection methods implemented in *bcp* package (Erdman and Emerson, 2008), high-dimensional change point detection implemented in *hdcd* package (Londschien et al., 2020), and Bayesian change point methods with spike and slab prior implemented in *solocp* package (Cappello et al., 2023).

The first Bayesian method (Erdman and Emerson, 2008) is flexible and can be applied to both univariate and multivariate cases. Because it is one of the oldest methods we utilize and is essentially a Markov chain Monte Carlo (MCMC) implementation of an older statistical approach (Barry and Hartigan, 1993), it is computationally intensive. It returns an informative summary of a probability distribution defined over all time points, which reflects the degree of uncertainty of the detection procedure. Following previous work, from such a posterior probability distribution we select time points where the probability of them

being change points is greater than 0.8. This returns a robust result with time points denoting striking changes in the statistical properties of emotion time series. We apply it to the univariate case of the sum score of negative emotions and the score of neutral expression, as well as to the multivariate case of all scores collectively.

The second method is a frequentist approach developed for high-dimensional data settings (Londschien et al., 2020). High-dimensional settings are those where the number of variables is approximately the same or greater than the number of observations ($p > n$). Even though our setting is not high-dimensional, this approach is useful as it is an advanced frequentist method exclusively developed to process multivariate time-series with different lengths and with missing values, which makes this approach useful for the multivariate analysis of noisy time series like the ones we are dealing with. This method returns a list of detected change points based on multivariate input and we use it for the analysis of a time series for the six emotions.

Finally, the third method is a new Bayesian approach for change point detection based on the spike and slab prior (Cappello et al., 2023). The key difference here is that at the beginning of the analysis there is a pre-selection method of likely candidates for change points based on the simple estimation of differences in statistical properties of the time series. The analysis then proceeds by estimating the exact probability distribution over the candidate change points, and we subsequently select those whose posterior probability of being a change point is high. The advantage this approach has over Erdman and Emerson (2008) is that it is not based on MCMC and is thus much faster, however the disadvantage is that it only works in the univariate case.

Overall, we have three methods from all three categories regarding the type of time series data they can be applied to, two of which are Bayesian methods and one of which is frequentist. These approaches represent the current state-of-the-art in their respective categories and the aggregated results will provide us with a solid starting point for the qualitative analysis.

Results

The applied change point detection algorithms produced six different change points related to three different time series: (1) a time series of negative emotion scores (sum scores of anger, fear, disgust and sadness), (2) a time series of neutral expression scores, and (3) multivariate time series of all six basic emotions (Figure 2).

Figure 2 shows the 12-frame rolling average time series for all emotions. This rolling average means that for the n -th point of the time series we take into account the average of the previous 12 ($n-11$ to n) frames, i.e. the average score for each emotion in the previous 12 frames leading up to a particular point in time. Given that the average time distance between frames is around 5 seconds, altogether the moving average for one minute of the video is considered. This approach reduces the impact of unavoidable misclassification errors of the emotion detection procedure and reduces the volatility of the time series. Figure 2 shows that we have changes in the most dominant emotion expressed during the speech, with substantial switches from “happy” to “sad” and “fear” in the first five minutes of the video, followed by the dominance of “fear” in the consecutive five minutes,

and then switches between “surprise”, “fear” and “sad” in the second half of the video. Other emotions either have a constantly low value (“disgust” and “happy”) or are changing over time, but their scores are consistently lower than the most dominant emotions (“anger”). Visual inspection of Figure 2 gives us insight into the dynamics of dominant emotions but fails to provide more precise information about key moments in the speech where these dynamics change.

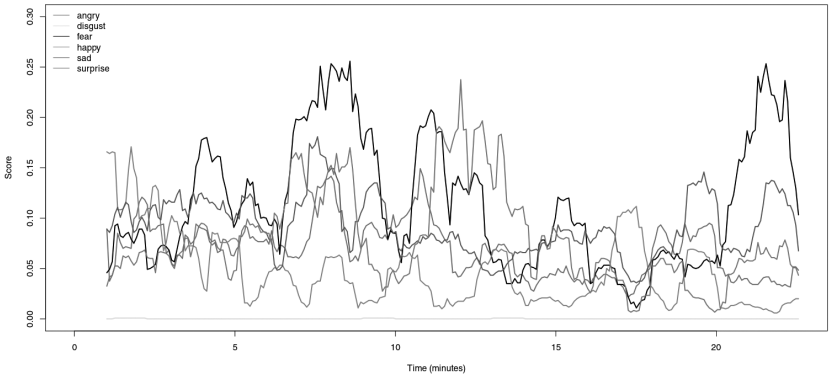


Figure 2 – Rolling average time series for six emotions

Detailed insights into affective dynamics are provided by change point detection methods. The Bayesian change point detection algorithm detected no change points for neutral expressions, meaning that there are no points in time for which the posterior probability of them being change points is greater than 0.8. For negative emotions, three different change points were detected in frames 91, 225, and 227. Since frames 225 and 227 are seconds apart in the video, we treat them as a single change point in a time series of negative emotions. The same approach applied to multivariate time series found change points in frames 84 and 94. Once again, since change points in frames

84, 91, and 94 are close to one another (all three occur within one minute), we treat them as a single change point.

The second method, the high-dimensional change point detection algorithm (Lonschien et al., 2020), was applied only in a multivariate time series of all six emotions. It returned two change points in frames 157 and 213.

Finally, results of the Bayesian change point detection approach based on spike and slab priors (Cappello et al., 2023) returned frames 228 and 268 for time series of neutral expressions and frames 180 and 264 for time series of negative emotions. These frames were selected with a high posterior probability of them being the true change points of these series. We treat frame 228 as the same change point as frames 225 and 227, and frames 264 and 268 as another single change point.

This gives us a total of six change points which are shown in Table 1. We combine change points which are less than a minute apart into a single point, consistently choosing the one that came last in the video, given that it reflects the rolling average of the previous minute of the speech. Timestamps are calculated to be 3 seconds before the detected change point in order to provide more context into the part of the speech in which a specific emotional change point was detected.

Description of changepoints is given in terms of dynamics of negative emotions and neutral expression because it is less volatile and more amenable to interpretation. Figure 3 displays each of the six change points (A-F) on a time series of negative emotions and neutral expression.

Table 1: Detected change points and their description

Change points	Frame number	Timestamp	Change description
A	94	7m51s	Decline of negative emotions
B	157	13m9s	Rise of neutral expression
C	180	15m6s	Continual rise of neutral expression to its maximum
D	213	17m52s	Burst of negative emotions
E	228	19m3s	Continual rise of negative emotions
F	268	22m10s	Rise of neutral expression

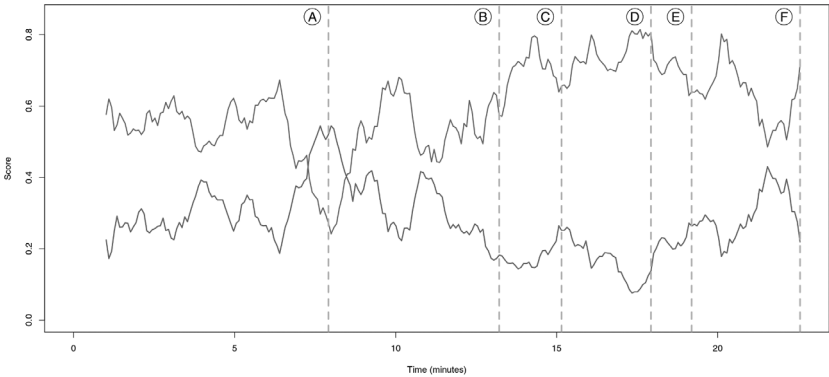


Figure 3 – Time series of negative emotions scores and the neutral expression score. Letters A-F denote change points described in Table 1.

Finally, we locate the given change points within a transcription of the video, taking the verbal context into consideration and comparing the emotion scores to what was said in the minute surrounding a specific change point. It should be noted that the original video transcription is in Serbian and that the following excerpts are translations made for illustrative purposes, which are subject to translator interpretation.

Change point A occurs at the end of the 8th minute of the video. The section of the president's speech directly preceding the change point involves his assessment of international institutions as hypocritical and their treatment of Serbia as unjust when compared to their treatment of Ukraine. A decline in negative emotions ensues as he turns to citing an excerpt from a United Nations Resolution.

[07:51] Vučić: But when you say the word 'Serbia' together with 'Charter of the United Nations' and 'Resolution of the United Nations'—suddenly they're all deep in thought, thoughts so deep one would think they're all Einstein or Hegel, *they look off into the distance as if they don't understand a word you're saying... but in Resolution 1244 it says: 'Reaffirming the commitment of all Member States to the sovereignty and territorial integrity of the Federal Republic of Yugoslavia and the other States of the region', but nowhere does it mention that Kosovo is one of the 'other states' in the region...*

The next change point (B) occurs about 13 minutes into the video, after Vučić continues his criticism of the West and addresses the speculation that he would order an invasion of neighboring countries following the Russian invasion of Ukraine. A rise in neutral expression appears as he switches to speaking about the status and treatment of himself and Serbia in international relations.

[13:09] Vučić: The point is that the West allows itself to do anything it wants, and everything they say is 'based on the law' even when *it has nothing to do with the law—and even less with*

justice—and anyone that disagrees is told: ‘Hey, you’re going against democracy’... ‘Democratic countries’, and so on. Do you think I’m worried about what you’re going to do and how you’re going to react? About how ‘Little Putin is provoking everyone’, and whatever else they say about me? Shame on you for everything that you do, you cause crisis after crisis here...

Similarly, change point C transpires at a point two minutes later where the president once again shifts his narrative from a negative appraisal of Western powers to the status of Serbia and his own efforts to reach a compromise regarding the normalization of Kosovo–Serbia relations. When speaking of Serbia and its purported victimization, there is a continual rise of neutrality in his facial expression.

[15:06] Vučić: All I ask is that they... just cut back a little on their audacity and arrogance... just cut back a little... and respect the rights... of those they’ve already killed before... so I hope it doesn’t occur to them to do it again... *and truth be told, we have nowhere to go from here, and I say this as the president of Serbia... we’ve fought for a compromise for ten years, for ten years none of them have wanted a compromise, they’ve just wanted to humiliate Serbia... they just want us to agree to what they’ve already decided...*

Contrastingly, after a section of the president’s speech in which he talks about his own veracity at the end of the 18th minute, a burst of negative emotions (change point D) is detected at the moment he

returns to an evaluation of how international actors have jeopardized Serbia's status, and he subsequently opens up the floor for questions.

[17:52] Vučić: I think this is the roughest press conference I've ever held, but I held it—with a smile on my face, believe me—and I've said everything I think and everything that is true, and every word I've said has been absolutely truthful. As far as my own fate will have it... with this conference I've gone down in history because at least I've said everything to their face, everything that they're doing, every way they're deceiving us, every way they're destroying us. *And I reiterate: the problem isn't that they think we're stupid, the problem is that they think they're so smart...* Go ahead, do you have any questions?

The final two change points take place during answers Vučić gives to reporters' questions around the 19th and 22nd minute mark. Change point E occurs in response to a question about future diplomatic dialogue posed by RTS (Serbian public broadcasting service) reporter Tamara Stojisavljević, and shows a marked, continual rise of negative emotion as he dismisses the very notion that dialogue with Kosovo and EU authorities is possible.

[19:03] Vučić: What dialogue? To what end? After these decisions—a dialogue? About what? About the Association of Serb Municipalities where they can't even put a banner up? It's like you're not even listening to me, Tamara. It's like you're not even listening. About what—about how they arrest people for having a tiny Serbian flag in their car? They arrested Bogdan

Laban because he had a Serbian flag in his trunk. *What, is that a criminal offense? I told Miroslav Lajčák—imagine getting arrested for walking with a Slovak flag and being held in some jail in Pristina for three hours. How is that acceptable, huh? How is that acceptable? Did anyone condemn that? Did anyone condemn the fact that they arrest people for putting up banners of the Association of Serb Municipalities—signed in Brussels, mind you...*

Lastly, change point F—an increase in neutral expression—arises when the president answers PINK (pro-government media outlet, see Castaldo, 2020) reporter Gordana Uzelac’s question about what will happen if Serbia’s requests are not met by September 30th and whether it is possible conflict will arise in northern Kosovo.

[22:10] Vučić: I don’t think we have any other... weapon, other than the truth. We’ll make our requests and I believe *someone will be smart enough to fulfill them. That’s all I can tell you.* But allowing another Operation Storm, another Operation Flash—as I said eight years ago and many times since—that is certainly not going to happen.

Discussion

In this study, we deployed a machine learning approach for detecting facial expressions of emotions in a YouTube video of President Vučić. This approach produced time series data describing the sequence of

emotion scores for six basic emotions and a neutral facial expression. Several change point detection methods were applied to these time series and to time series of the sum scores of negative emotions. As a result, we obtained a list of six timestamps in the video where significant changes occur in its temporal affective dynamics. We extracted the transcript of the video around these six change points, and qualitative interpretation revealed that the retrieved change points are not mere statistical artifacts, and that they correspond to important discursive shifts in the speech itself. While it is difficult to generalize based on limited data, there seems to be a shift towards a more neutral facial expression when the subject of the president's speech is the status of Serbia and himself in international relations, especially when they are framed as unjustly victimized by the West (change points B, C, and F). Similarly, there is a decline in negative emotions when he reads from an existing text and is thus not 'instinctively' speaking (change point A). The two change points relating to a rise in negative emotion (D and E) occur when the president answers a question from the Serbian public broadcasting service and turns his attention to the actions (or lack thereof) of the authorities mediating Kosovo-Serbia relations. Finally, it was interesting to note the decline of negative emotions (change point F) when answering questions from pro-government media outlet. Overall, while these insights do not constitute an in-depth qualitative analysis of the verbal elements of the address, they do lend support to the idea that the detected change points represent significant affective shifts in the speech, and point to a viable line of inquiry for future research on the matter.

While this study focused on a single video of President Vučić's public address, the methodology employed can be applied to analyze emotional dynamics in different types of online political communication.

The analysis of a larger dataset of videos from different contexts and speakers was presented elsewhere (Major and Tomašević, 2023) and results based on quantitative statistical analysis revealed significant differences in emotional expressions between populist and non-populist leaders. In future work, a further exploration of the mixed-methods approach outlined here can be extended to cover different talks by populists and non-populists around the globe and unveil the qualitative, discursive differences which follow different patterns of their affective dynamics. This could further enhance our knowledge of how emotional dynamics influence audience engagement, persuasion, and the effectiveness of political communication.

This study has significant limitations. Firstly, the analysis is based on a pre-trained machine learning model for emotion detection, which may have inherent biases and limitations. Training a model on a dataset of frames extracted from a diverse sample of political speeches “in-the-wild” would potentially improve the accuracy of emotion detection in this particular setting. Secondly, the change point detection methods used in this study were adapted to the analysis of emotional time series, but their performance in this specific context has not been extensively evaluated. Even though a combination of different methods was used, there remains a probability that retrieved change points do not represent the true dynamics of the time series analyzed. Finally, qualitative interpretation was performed without an established coding scheme and without cross-validation by multiple coders and the results are therefore subjective in nature and prone to researcher bias.

In conclusion, this study demonstrates the prospect of integrating machine learning based workflows for facial expression emotion

detection and content analysis to analyze emotional dynamics in on-line videos. The findings shed light on the emotional expressions and rhetorical modes employed in a YouTube video of President Vučić's public address. By utilizing change point detection methods, the study identifies significant shifts in emotional expressions, providing insights into the affective dynamics of the speech. We found that these change points correspond to shifts in framing the subject of speech as either the victim or perpetrator of some form of injustice (e.g., Serbia and 'the West', respectively). While the case study presented here is limited in scope, this approach can serve as a foundation for future research on emotional dynamics in various digital communication contexts, ultimately furthering our understanding of the role of emotions in political communication.

References

- Arriaga, O., Valdenegro-Toro, M., & Plöger, P. (2017). Real-time convolutional neural networks for emotion and gender classification (arXiv:1710.07557). *arXiv*. <https://doi.org/10.48550/arXiv.1710.07557>
- Barry, D., & Hartigan, J. A. (1993). A Bayesian analysis for change point problems. *Journal of the American Statistical Association*, 88(421), 309–319. <https://doi.org/10.1080/01621459.1993.10594323>
- Bonansinga, D. (2020). Who thinks, feels: The relationship between emotions, politics, and populism. *PArtecipazione e CONflitto*, 13(1), 83–106. <https://doi.org/10.1285/120356609V13I1P83>
- Bucy, E. P., Foley, J. M., Lukito, J., Doroshenko, L., Shah, D. V., Pevehouse, J. C., & Wells, C. (2020). Performing populism: Trump's transgressive debate style and the dynamics of Twitter response. *New Media & Society*, 22(4), 634–658. <https://doi.org/10.1177/1461444819893984>
- Cappello, L., Madrid Padilla, O. H., & Palacios, J. A. (2023). Bayesian change point detection with spike-and-slab priors. *Journal of Computational and Graphical Statistics*, 32(4), 1488–1500. <https://doi.org/10.1080/10618600.2023.2182312>
- Castaldo, A. (2020). Back to competitive authoritarianism? Democratic backsliding in Vučić's Serbia. *Europe-Asia Studies*, 72(10), 1617–1638. <https://doi.org/10.1080/09668136.2020.1817860>
- Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, 1(1), 56–75. <https://doi.org/10.1007/BF01115465>
- Ekström, M., Patrona, M., & Thornborrow, J. (2018). Right-wing populism and the dynamics of style: A discourse-analytic perspective on mediated political performances. *Palgrave Communications*, 4(1), 83. <https://doi.org/10.1057/s41599-018-0132-6>

Erdman, C., & Emerson, J. W. (2008). bcp: An R package for performing a Bayesian analysis of change point problems. *Journal of Statistical Software*, 23, 1–13. <https://doi.org/10.18637/jss.v023.i03>

Fujiwara, T., Müller, K., & Schwarz, C. (2021). The effect of social media on elections: Evidence from the United States. *NBER Working Paper 28849*. <http://www.nber.org/papers/w28849>

Garzia, D. (2017). Personalization of politics between television and the internet: Leader effects in the 2013 Italian parliamentary election. *Journal of Information Technology & Politics*, 14(4), 403–416. <https://doi.org/10.1080/19331681.2017.1365265>

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., ... Bengio, Y. (2013). Challenges in representation learning: A report on three machine learning contests (arXiv:1307.0414). *arXiv*. <https://doi.org/10.48550/arXiv.1307.0414>

Grabe, M. E., & Bucy, E. P. (2009). *Image bite politics: News and the visual framing of elections*. Oxford University Press.

Haselmayer, M., & Jenny, M. (2017). Sentiment analysis of political communication: Combining a dictionary approach with crowdcoding. *Quality & Quantity*, 51(6), 2623–2646. <https://doi.org/10.1007/s11135-016-0412-4>

Lalancette, M., & Raynauld, V. (2019). The power of political image: Justin Trudeau, Instagram, and celebrity politics. *American Behavioral Scientist*, 63(7), 888–924. <https://doi.org/10.1177/0002764217744838>

Londschien, M., Kovács, S., & Bühlmann, P. (2020). Change point detection for graphical models in the presence of missing values (arXiv:1907.05409). *arXiv*. <http://arxiv.org/abs/1907.05409>

Major, S., & Tomašević, A. (2023). The face of populism: Examining differences in facial emotional expressions of political leaders using machine learning (arXiv:2304.09914). *arXiv*. <http://arxiv.org/abs/2304.09914>

- Masters, R., Sullivan, D., Lanzetta, J., Mchugo, G., & Englis, B. (1986). The facial displays of leaders: Toward an ethology of human politics. *Journal of Social and Biological Systems*, 9(4), 319–343. [https://doi.org/10.1016/S0140-1750\(86\)90190-9](https://doi.org/10.1016/S0140-1750(86)90190-9)
- Mazzoleni, G. (2014). Mediatization and political populism. In F. Esser & J. Strömbäck (Eds.), *Mediatization of politics* (pp. 42–58). Palgrave Macmillan.
- Mendonça, R. F., & Caetano, R. D. (2021). Populism as parody: The visual self-presentation of Jair Bolsonaro on Instagram. *The International Journal of Press/Politics*, 26(1), 210–235. <https://doi.org/10.1177/1940161220970118>
- Moffitt, B. (2022). How do populists visually represent ‘the people’? A systematic comparative visual content analysis of Donald Trump and Bernie Sanders’ Instagram accounts. *The International Journal of Press/Politics*. <https://doi.org/10.1177/19401612221100418>
- Norris, P., & Inglehart, R. (2019). *Cultural backlash: Trump, Brexit, and authoritarian populism*. Cambridge University Press. <https://doi.org/10.1017/9781108595841>
- Salmela, M., & von Scheve, C. (2018). Emotional dynamics of right- and left-wing political populism. *Humanity & Society*, 42(4), 434–454. <https://doi.org/10.1177/0160597618802521>
- Salojärvi, V. (2019). Populism in journalistic photographs: Political leaders in Venezuelan newspaper images. *Iberoamericana—Nordic Journal of Latin American and Caribbean Studies*, 48(1), 28–39. <https://doi.org/10.16993/iberoamericana.439>
- Stewart, P. A., & Ford Dowe, P. K. (2013). Interpreting President Barack Obama’s facial displays of emotion: Revisiting the Dartmouth Group. *Political Psychology*, 34(3), 369–385. <https://doi.org/10.1111/pops.12004>
- Sullivan, D. G., & Masters, R. D. (1988). “Happy warriors”: Leaders’ facial displays, viewers’ emotions, and political support. *American Journal of Political Science*, 32(2), 345–368. <https://doi.org/10.2307/2111127>

Udris, L., & Lucht, J. (2014). Mediatization at the structural level: Independence from politics, dependence on the market. In F. Esser & J. Strömbäck (Eds.), *Mediatization of politics*. Palgrave Macmillan.

Walker, E. B., & Boyer, D. M. (2018). Research as storytelling: The use of video for mixed methods research. *Video Journal of Education and Pedagogy*, 3(1), Article 8. <https://doi.org/10.1186/s40990-018-0020-4>

Widmann, T. (2021). How emotional are populists really? Factors explaining emotional appeals in the communication of political parties. *Political Psychology*, 42(1), 163–181. <https://doi.org/10.1111/pops.12693>

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10), 1499–1503. <https://doi.org/10.1109/LSP.2016.2603342>

Zhuravskaya, E., Petrova, M., & Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, 12(1), 415–438. <https://doi.org/10.1146/annurev-economics-081919-050239>

Александар Вучић. (2022, June 29). Вучић: Три пута размислите до 30. септембра. [Video]. YouTube. <https://www.youtube.com/watch?v=knhZJ-clSbWQ>

The Power of Images: Visual Framing of the Ukraine Conflict on Instagram

Gheorghe
Anghel

Introduction

The current research topic holds great importance in the global context, especially with the ongoing war between Russia and Ukraine, a conflict that has international repercussions and reverberations. Although the military conflict started last year, the history of misunderstandings between the two states is much older. A significant moment in recent history occurred in 2014 when Russia annexed Crimea, generating considerable international focus and debate. The action orchestrated by the Kremlin leader “was a triumph of political manipulation over national interests and common sense” (Ragozin, 2019) and involved a carefully crafted strategic disinformation campaign (Zeitsoff, 2017: 2). Since then, tensions have increased, reaching a peak when Vladimir Putin ordered the attack on Ukraine, claiming it was an

Gheorghe Anghel, Assistant professor, Department of Journalism, Faculty of Journalism and Communications Sciences, University of Bucharest, Bucharest
ORCID: 0000-0001-5054-2041, email: gheorghe.anghel@fjsc.ro

DOI: <https://doi.org/10.5281/zenodo.14203819>

“act of self-defense” because, he said, “Russia did not want to occupy Ukraine but would demilitarize and de-Nazify the country” (Ukraine Conflict: What We Know about the Invasion, 2022).

Understanding the profound influence of the digital sphere, particularly social media, is crucial for comprehending the dissemination of information and its impact on public opinion during the ongoing hostilities between Russia and Ukraine. In today’s interconnected society, where we inhabit both physical and virtual worlds, the digital sphere plays a significant role in shaping public opinion and our understanding of events. This influence extends to instances of armed conflict, including Russia’s military aggression toward Ukraine. There are several reasons why this phenomenon occurs. One of the main factors is the growing number of people becoming increasingly active in the digital world (Kemp, 2023). Access to the internet has significantly expanded, and there is a widespread trend toward the digitalization of information consumption. As individuals engage more in the virtual world, it affects them cognitively. The information and stimuli they receive from the digital space significantly influence their behavior and decision-making.

Additionally, social media plays a vital role in modern society, effectively serving as a parallel world for many of us. Currently, there are about 4.76 billion users worldwide, with 77.8% being over 18 years of age. On average, users spend roughly 2 hours and 31 minutes per day on social media platforms (Kemp, 2023). Social media has also become one of the primary gateways to news consumption; more than 90% of the world’s population owns a smartphone, and trends indicate a growing tendency for news to be accessed through social networks. This phenomenon is particularly common among people aged 18 to 24, who “have an even weaker connection with websites and apps, preferring to access news via side-door routes such as social

media, search, and mobile aggregators” (Newman, 2022). With the rise of social media, communication has shifted from being one-way to two-way, resulting in a more interactive dialogue between journalists and their audience (Adornato, 2021: 5).

This paradigm shift has also influenced the visual framing of conflicts, as the production and distribution of images occur at an astonishing speed compared to the era before the emergence of social media. Social networks are now “used to shape public opinion, mobilize supporters, coordinate military activities, and collect information for targeting purposes. It has increasingly become the ‘weapon of choice’ for both state and non-state actors” (Reynolds, 2016: 13).

Therefore, it is vital to comprehend how information is displayed on social media platforms and how these tools are used in communication. This understanding can aid in critically assessing content, grasping framing strategies, manipulating public opinion, and preventing the dissemination of misinformation. A study on the visual framing of the Ukraine conflict may help shed light on how social media users are exposed to selective and filtered information, potentially changing our approach to global conflicts and events. Ultimately, the study’s importance lies in its capacity to provide valuable insights into how mass media depicts and approaches the Ukrainian conflict.

Literature Review

Building upon the introductory remarks that emphasize the importance of social media in mass communication, specifically in relation to armed conflicts, the literature review emphasizes the powerful

influence of media organizations in selectively presenting visual frames that shape public perception and comprehension of such events.

“The most utilized mass communication theory of the present era” (Bryant and Miron, 2004: 695), framing “is widely used in the social and behavioral sciences” (Lecheler and De Vreese, 2018: 7). According to Goffman, frames are mental tools or “schemata of interpretation” (1986, p. 21) that individuals use to understand and give meaning to social situations and the world around them. Frames serve as a means of storytelling that combines visual and textual elements to provide context to events and guide their interpretation in a particular direction (see Gamson and Modigliani, 1987). Given that “framing is inherent in all storytelling, it plays a major role in journalism” (Schwalbe and Dougherty, 2015: 145) and contributes to the formation of public opinion because “the way information is structured affects cognitive processing, and ultimately, the meaning people derive from that information” (Boaz, 2005: 350).

Journalists have the power to shape the audience’s perception of an event by selectively choosing which part of the information to present. This can be done either intentionally or unintentionally, resulting in a particular portrayal of the event. Media frames serve the purpose of organizing the world “both for journalists who report it and, to some important degree, for us who rely on their reports” (Gitlin, 1980: 7). According to Gitlin, media frames refer to “persistent patterns of cognition, interpretation, and presentation, of selection, emphasis, and exclusion, by which symbol-handlers routinely organize discourse, whether verbal or visual” (ibid.: 7).

Therefore, frames are defined both by the writing, by the approach angle of the article, also by images that “influences frame construction”

(Schwalbe, 2013a: 243), because “photographs command more attention than words alone... and elicit a spontaneous emotional response” (Schwalbe and Dougherty, 2015: 142).

Images have the power to present fragments of reality in a way that text cannot. Be it still or moving, images help build and give meaning to the surrounding reality, enabling viewers to better comprehend the events presented, be it a marathon, a plane crash, or an armed conflict. As in the case of news frames where reality is constructed (mainly) through the generated/selected text, journalists/photojournalists can define reality through the images they choose to document an event. Practically, depending on the composition of those images, specific frames will be generated (Fahmy et al., 2022: 32) and influence the public that will see the images. The same applies to wars as well, because “images play a key role in modern mediatized conflicts, promoting particular ways of understanding those conflicts, what they are about, and who drives them” (Ojala et al., 2017: 474). Especially in the case of armed conflicts, “visuals become an effective tool for creating persuasion and gaining public support for government, national security, and military actions” (Fahmy and Wanta, 2007: 20).

In the case of the Iraq conflict, the US mass media played a significant role in shaping the public’s perception of the war by promoting certain frames while neglecting others. Thus, the American perspective involved emphasizing heroism in favor of failure, freedom in favor of destruction, but also omitting images of injured or deceased people (Schwalbe, 2006: 269). Thus, as Sloan and Stuart (1996) wrote, “public opinion is a prime factor in war” (as cited in Fahmy and Wanta, 2007: 17).

Griffin and Lee undertook a visual content analysis of 1104 Gulf War (1990–1991) related photographs showcased in Time, Newsweek, and

US News and World Report magazines (Griffin 1995: 813). Their analysis revealed that the pictorial coverage primarily revolved, in all the publications, around a limited range of images, particularly focusing on documenting military weaponry and technology (Griffin and Lee, 1995: 813). The study also demonstrated that the photographs exhibited a clear bias towards the United States, with a deliberate avoidance of depicting images related to death, injury, and funerals (Griffin and Lee, 1995: 813). A few years later, Griffin conducted a comparative examination of the photographic coverage in American news magazines during three military interventions by the United States in southwest Asia: Gulf War, Afghanistan, and Iraq (ibid.: 382). The research findings uncovered a pattern: news magazines consistently sided with the official narrative put forth by Washington. The American president frequently appeared in war photographs, visually portrayed as a “strong and confident leader” (ibid.: 399). At the same time, the greatness of the American nation was implied through numerous frames featuring “US troops, weapons, and military hardware” (ibid.: 399). It is worth noting that two media institutions can visually frame the same conflict differently based on their political views, approach to war, national perspectives etc. For instance, the first Gulf War was framed differently in the Arab and American media. In the Arab world, to obtain support for the Iraqi people, “graphic images of human suffering were shown to arouse national sentiments in support for the oppressed people in Iraq”, and in the American media “portrayed the first Gulf war as a bloodless affair” (Fahmy and Wanta, 2007: 18).

Fahmy and Kim (2008) conducted an analysis comparing the visual portrayal of the 2003 Iraq War in two prominent newspapers: The New York Times representing the *US press* and *The Guardian* representing the British press (ibid.: 444). The authors concluded that both

newspapers depicted the event in distinct ways, primarily influenced by the extent of military engagement and public sentiment within their respective countries. This depiction was independent of the political inclination of the news organizations and the perceived significance of the matter being reported (ibid.: 458).

In 2010, Fahmy investigated “the contrasting visual narratives in English and Arabic-language transnational press in covering the 9/11 attack and the Afghan War” (Fahmy, 2010: 696). The investigation revealed that the Arabic-language *Al-Hayat* newspaper framed the story of the terrorist attack by downplaying visual elements that focused on the victims, instead emphasizing images depicting material destruction and the impact of the planes. *Al-Hayat* published fewer photographs showcasing mourning individuals, no images of 9/11 funerals, and victims of the tragic event (ibid.: 711). The other newspaper, the IHT English-language, underlined “the emotion of guilt in 9/11” by “humanizing 9/11 victims” with the help of “images of citizens mourning, mug shots of the deceased and funerals, and focused less on depictions of material loss” (ibid.: 711).

Parry (2011) initiated a content and visual framing analysis focused on British press photography during the period of the 2003 Iraq invasion. The objective was to examine the interconnected concepts of “humanitarian warfare and photographic humanism” (ibid.: 1186). The conclusion of the research showed that the humanitarian frame was “largely pictured in supportive terms, at least in the high-circulation, pro-war press” (ibid.: 1187).

Schwalbe (2013) conducted an analysis of the visual framing of the Iraq invasion in prominent news magazines, namely *Time*, *Newsweek*, and *U.S. News and World Report*. The study revealed that these

influential publications predominantly presented the conflict from an American-centered viewpoint, emphasizing aspects such as conflict itself, politicians, and human-interest stories. Also, they “neglected alternative viewpoints, such as antiwar protests, destruction, Iraqi military leaders and troops, and the human toll” (ibid.: 239).

Also, the Sri Lankan Civil War was analysed by Neumann and Fahmy (2012) in order to see the extent to which the visual coverage of the conflict relied on war and peace frames. The authors examined news photographs in three leading western news wires: *Associated Press*, *Agence France Press*, *Reuters*. Findings concluded that “war frames continue to prevail in Western newswires (when taken as a whole), even though only a slim majority of the visuals in this study (52.7%) outweighed photographs of peace negotiations and anti-war demonstrations” (ibid.: 194)

Greenwood and Jenkins conducted a study focusing on the visual framing of the Syrian conflict that took place between 2011 and 2012. The researchers analyzed a total of 193 photographs “published in two news magazines and nine public affairs magazines available to American readers to determine whether visual framing differs among magazines with differing purposes and differing political/editorial orientations” (Greenwood and Jenkins, 2015: 207). The findings revealed that viewers were exposed to a balanced representation of both civilian victims and active combatants in the images of the conflict. The photographs diverge from potentially glorifying combatants as valiant fighters for freedom or defenders of the government. Instead, they unequivocally highlight the price paid by non-combatants in the conflict (ibid.: 222).

Also, in 2015, a research was conducted to see how *Time*, *Newsweek*, and *U.S. News and World Report* visually covered the 2006 Lebanon War. Through a quantitative content analysis, Schwalbe and Dougherty (2015) “revealed that the military conflict and human interest frames dominated visual coverage of the seven-week war” (ibid.: 141).

Makhortykh and Sydorova (2017) “investigated the use of social media for visual framing of the conflict in Eastern Ukraine” from 2014. They analysed how the war “was represented and interpreted in pro-Ukrainian and pro-Russian online communities” (ibid.: 359). The research results showed two different approaches to the same conflict. The pro-Ukrainian forces framed the conflict from a patriotic perspective as a good war “through the predominant use of such thematic categories as military machines, combatants and trophies”. On the other side, the Russians, “presented the conflict as humanitarian crisis, emphasizing the destruction and suffering that the war brought through the predominant use of civilians and ruins categories” (ibid.: 376–377).

The same conflict was analysed from a European perspective. Ojala et al (2017) examined “the visual coverage of the Ukraine conflict in *The Guardian*, *Die Welt*, *Dagens Nyheter*, and *Helsingin Sanomat*” (ibid.: 474). Findings indicated “in the almost total absence of direct depictions of death, the papers often employed images of civilians alongside images of material destruction or symbols such as candles, flowers, crosses, or coffins to illustrate the human cost of the conflict” (ibid.: 480).

Anghel (2023) conducted an analysis to examine the visual representation of Russia’s aggression towards Ukraine in 2022 on Instagram. The

study focused on the video content uploaded by three media outlets: BBC News, Ria Novosti, and Ukrinform. The analysis specifically examined video clips published during the period of February 24 to April 24, 2022. A total of 169 video clips were analyzed, with BBC publishing 87 materials and Ria Novosti publishing 80. It should be noted that Ukrinform had only 2 video clips, primarily featuring still images. Consequently, the analysis considered materials solely from the first two media institutions. The author utilized the research grid presented by Makhortykh and Sydorova (2017) to conduct the analysis. This grid was applied to both the first frame and the complete video material(s). According to the research findings, the first frame of the videos by both agencies primarily featured civilians as the most frequent category. Next, the BBC depicted the devastation of war through visuals of combatants and ruins, whereas RIA emphasized officials before transitioning to combatants and ruins. As anticipated, RIA—a Russian state-controlled agency, prominently showcased President Vladimir Putin's image in the category of officials, while BBC News opted for President Zelensky's image. The results of the Pearson test showed that „there is a tendency for videos that include in the first frame the image of a civilian man or combat images to receive more likes from visitors” (Anghel, 2023: 179) Instagram users also tend „to like or comment less where the video clip opens with an image of a group of civilians” (Anghel, 2023: 180). The results do not change much after analyzing the materials in their entirety, and not just the first frame.

Based on Makhortykh and Sydorova (2017) theoretical approach, the study aims to answer three research questions:

RQ1 What are the main categories encountered?

RQ2 What differences in framing the conflict can be detected between the 3 agencies?

RQ3 How do visitors interact with the posts, and which thematic categories in the frames attract the most attention?

Corpus and Methodology

This corpus comprises images shared on Instagram by three news agencies—BBC News, RIA Novosti, and Ukrinform—during the period of February 24 to April 24, 2022. The analysis encompassed a total of 717 images extracted from 219 posts that were directly associated with the conflict in Ukraine. Among the three agencies, BBC News had the largest number of posts, with 118 in total, featuring a total of 304 images. RIA Novosti had 17 posts with 35 images, while Ukrinform had 84 posts with 378 images.

To achieve the best outcomes, I reviewed all image posts from the three Instagram accounts and selected the ones related to the Ukraine conflict. Regarding the materials published by RIA and Ukrinform, a caption translation software was used to overcome the language barrier, especially for posts that did not make a clear reference to the conflict in Ukraine. Following this stage, the posts that were not related to the research subject, but also those that included video clips, were removed from the corpus.

To determine the frequency of specific thematic categories in the selected images, an analysis grid was utilized, based on the model developed by Makhortykh and Sydorova (2017). The two authors visually

framed the conflict in Ukraine, from 2014, identifying several themes present in the analyzed static images. Among those highlighted by the two authors, for the current research the following categories were included in the analysis grid: (1) „combat images”; (2) „animals”; (3) „civilians” (with the following subsections: woman, man, child, groups), (4) „soldiers”; (5) „Russian/Ukrainian/international officials”, (6) „military equipment”, (7) „landscape”, (8) „ruins” (9) „symbols” (national/religious), (10) „casualties” and (11) „injured”. In the analysis I also collected the number of comments and the number of likes for each post.

During the third phase, a unique ID code was assigned to each post and recorded in a database. If a post contained multiple images in a carousel, the analysis grid was applied to the first image and to all the images in the carousel separately. Special attention was given to the first image in each post to determine if it had any recurring themes or categories and to identify any potential correlations between these themes and the followers’ interactions with the multimedia content. Also, the first fifty materials were coded separately by two coders. Discussions were held where the options were different, and criteria were established to standardize the analysis.

Results

According to the analysis, in terms of weight, regarding the number of published images, Ukrinform leads with a percentage of 52.7%, followed by the BBC with 42.3% and Ria Novosti with 4.8%. However, when considering the number of posts, the hierarchy changes, the top

being led by BBC News, with 53.8%, followed by Ukrinform, with 38.3% and Ria Novosti, with 7.7%.

The information will be organized into different sections in the upcoming part of the article. Initially, I will emphasize the outcomes achieved by applying the analysis grid to the first image of the posts. Then, I will showcase the results for all the images in each post, particularly for carousel-style posts. The categories were marked only once in carousel posts, even if they had a higher recurrence. In addition, both combined and individual data for each media outlet will be provided for all three agencies.

Most Used Categories:

First Image of Posts.

Overall results.

Regarding the first image in the carousel, civilians are the most frequent category, appearing 97 times in the photos shared by all three agencies on their Instagram accounts. Regarding the subcategories, the statistics indicate a more significant recurrence of the one dedicated to civilian groups—34, followed by female civilian—22, child—23 and male civilian—18 (more info in Figure 3).

The category of symbols (including religious and national symbols) was the second most frequently used, appearing 46 times. In the posts of BBC News, the symbols were encountered 11 times: 2 times religious, 8 times national symbols of Ukraine and only once national symbols of Russia. In their Instagram posts, Ukrinform used images

with symbols to prominently feature Ukraine’s national symbols 24 times, while symbols representing Russia were only included once. Additionally, religious symbols were present twice. Ria, as expected, mainly used the national symbols of Russia in seven posts and only once those of Ukraine; no religious symbols were found. In total, at the level of the three agencies, the national symbols of Ukraine were encountered 33 times, of which 32 times they were promoted by the BBC and Ukrinform (Figure 1).

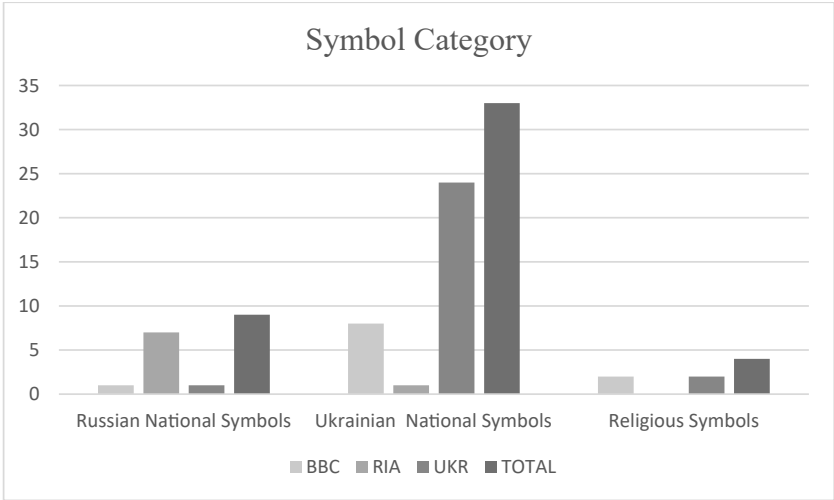


Fig. 1: *Usage of the symbol category in the first image of the posts*

The next recurring category is that of soldiers, being encountered 40 times in the posts of the three agencies: BBC—23, Ukrinform—16, Ria—1. It should be noted that Ukrinform did not publish any pictures of the officials from either country, Russia or Ukraine. There was an expectation that President Zelensky, who is often regarded as a prominent figure in the Ukrainian conflict, would be depicted in the pictures. However, contrary to expectations, no such images were published

by the Ukrainian news agency. Images of both Vladimir Putin and the President of Ukraine were posted by the BBC and Ria (see Figure 2). Categories such as combat images and injured are found with a shallow frequency, almost to none.

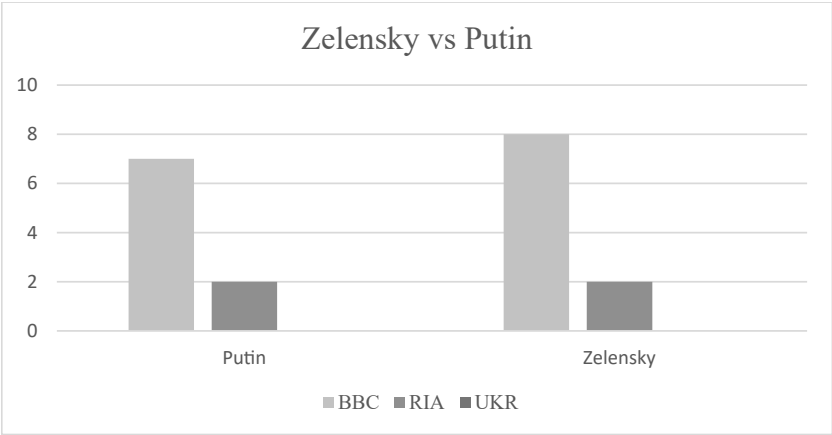


Fig. 2: Usage of the officials category in the first image of the posts

*Most Used Categories:
First Image of Posts.
Results Specific to
Each News Agency*

At an individual level, the BBC used the civilian category in 51 instances, while Ukrinform used it 41 times, indicating their primary focus on this category. The BBC used several images that brought to the fore the subcategories of civilian women—15 and children—14. Instead, Ukrinform relies on the subcategories of civilian groups—21, children—8, women—6 and men—6.

During the analyzed period, RIA featured the category of civilians only 5 times in the first images of their Instagram posts, indicating it was not their most frequently used theme. In contrast, the most frequently used category in RIA's posts was symbols, which appeared eight times. About the next most recurring category, the BBC focused on soldiers—23, Ukrinform on religious and national symbols—27, and Ria on civilians—5 and officials—4.

In reference to the third most used category at particular level, the BBC displays military equipment—18, while Ukrinform highlights soldiers—16. Ria, on the other hand, features military equipment and soldiers with only one recurrence.

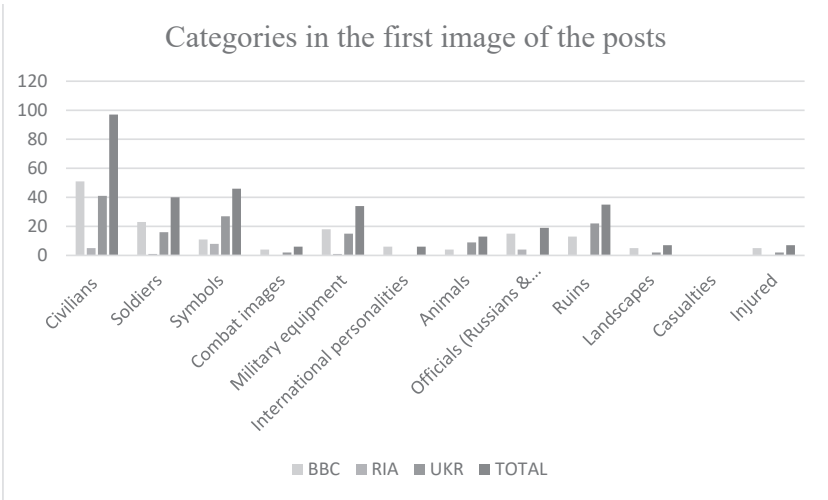


Fig. 3 Usage of the categories in the first image of the posts

Correlations: Categories in First Post Images and Number of Interactions

I utilized the Pearson correlation test to examine a potential connection between the thematic categories identified in the first image of the posts and the number of comments (Table I) as well as the number of likes (Table II). The Pearson test indicates whether the variables (thematic category, respectively the number of comments) tend to increase simultaneously (positive correlation) or in different directions (negative correlation). Table I summarizes the results of the correlation coefficients. The closer the correlation coefficient value is to 1, in the case of positive correlation, respectively -1, in the case of negative correlation, the stronger the correlation. In this particular context, it is worth mentioning that the majority of thematic categories, excluding officials, civilian men, and symbols, do not exhibit statistically significant correlations (since the p-value, which indicates the statistical significance of the test, is higher than 0.05) nor do strong correlations between variables appear (even without statistical power).

A weak positive correlation ($R = 0.1322$) was observed between the civilian man category and the number of comments, but it did not reach statistical significance (p-value = .05076). This indicates that using the civilian man theme may potentially boost the number of comments, but it cannot be assured as the test lacks statistical significance (p-value > 0.05). Similarly, it was discovered that there is a statistically significant correlation (p-value = 0.01448) between the symbol category and the number of comments. However, the correlation is small and negative ($R = -0.165$), meaning that the use of national or religious symbols in images leads to a decrease in the number of comments.

In the analysis, a strong statistical power (p -value = 0.0007782) was found in the correlation between officials and the number of comments. Although the correlation is not particularly strong ($R = 0.2254$), the test showed a positive correlation between these two parameters. The significant statistical power suggests a tendency for an increase in the number of likes when the thematic category focuses on officials.

Table 1: *Pearson correlations between category and number of comments*

Parameter 1	Parameter 2	Correlation coefficient, statistical significance <i>R (DF) = r, p-value; DF = 219</i>
Civil woman	Number of comments	$R = 0.02369$, p -value = .7274
Children	Number of comments	$R = -0.03139$, p -value = 0.6441
Officials	Number of comments	$R = 0.2254$, p -value = 0.0007782
Soldiers	Number of comments	$R = -0.03286$, p -value = 0.6286
Symbols	Number of comments	$R = -0.165$, p -value = 0.01448
Civil man	Number of comments	$R = 0.1322$, p -value = .05076
Combat images	Number of comments	$R = -0.005807$, p -value = 0.9319
Military Equipment	Number of comments	$R = -0.1086$, p -value = 0.1088
Civilian group	Number of comments	$R = -0.1253$, p -value = 0.06409
Animal	Number of comments	$R = -0.1271$, p -value = 0.06051
Ruins	Number of comments	$R = -0.05259$, p -value = 0.4387
Landscapes	Number of comments	$R = 0.05776$, p -value = 0.395
Injured	Number of comments	$R = 0.0774$, p -value = 0.2541
The result is significant at $p < .05$		

Also, the Pearson test was applied to determine any potential correlations between the different thematic categories in the initial image and the number of likes (Table 2). Upon comparing the current

findings with the previously obtained results (from the analysis of the thematic category and the number of comments, as shown in Table I), several similarities can be observed with a few exceptions. Specifically, it is worth noting that there are no statistically significant correlations between the number of likes and most categories in the first image of the posts. However, when examining the thematic category officials, a stronger correlation with the number of likes is evident ($R = 0.2852$). This correlation has a highly significant statistical power (p-value is 0.00001825), especially when compared to the correlation with the number of comments ($R = 0.2254$, p-value = 0.0007782, as shown in Table I).

It is evident that the correlation between the symbol category and the number of likes is strengthened. While a weak negative correlation was found between symbol and the number of comments ($R = -0.165$, p-value = 0.01448, table I), the correlation between symbol and the number of likes was considerably stronger (p-value = 0.0009446), yet still negative. This implies that when symbol appears in the first image of a carousel, both the number of comments and likes decrease, with the latter being more affected.

On the other hand, a weak but statistically significant correlation was observed between the category civil group and the number of likes ($R = -0.1464$, p-value = 0.03032). This indicates that the appearance of civil group results in a decrease in likes. However, no significant results were obtained in the correlation with the number of comments.

Regarding the thematic category landscapes, a weak positive correlation ($R = 0.1648$) was found, but it was statistically significant p-value = 0.01462). This implies that the appearance of landscapes in the first image tends to increase the number of likes on the post.

Table 2: Pearson correlation between the category in the first image of the carousel and the number of likes

Parameter 1	Parameter 2	Correlation coefficient, statistical significance R (DF) = r, p-value; DF = 219
Civil woman	Nr. of likes	R = 0.07935, p-value = 0.2423
Children	Nr. of likes	R = 0.00976, p-value = 0.8858
Officials	Nr. of likes	R = 0.2852, p-value = 0.00001825
Soldiers	Nr. of likes	R = 0.0389, p-value = 0.5669
Symbols	Nr. of likes	R = -0.2219, p-value = 0.0009446
Civil man	Nr. of likes	R = 0.04644, p-value = 0.4942
Combat images	Nr. of likes	R = -0.003821, p-value = 0.9552
Military equipment	Nr. of likes	R = -0.07776, p-value = 0.2518
Civilian group	Nr. of likes	R = -0.1464, p-value = 0.03032
Animal	Nr. of likes	R = -0.1133, p-value = 0.09442
Ruins	Nr. of likes	R = -0.09779, p-value = 0.1492
Landscapes	Nr. of likes	R = 0.1648, p-value = 0.01462
Injured	Nr. of likes	R = 0.08521, p-value = 0.2091
The result is significant at $p < .05$		

The Pearson test was also applied to determine if there are differences regarding the number of comments/likes when the image of the presidents of the two states—Zelensky or Putin, is used in the first photo of the posts. President Zelensky was assigned the code 1, President Putin was assigned the code 2. In the case of the number of comments and the number of likes, no statistically significant differences were obtained in comparing the two presidents. Thus, for the number of comments, a very weak, negative and statistically insignificant correlation resulted ($R = -0.1965$, $p\text{-value} = 0.4201$), indicating a weak tendency to increase the number of comments when Zelensky appears in the post's first image. The same tendency to increase the

number of likes when Zelensky appears is also noticeable in the case of the correlation with the number of likes, $R = -0.2461$ and $p\text{-value} = 0.3099$. However, both correlations are weak and do not have statistical significance.

In the next phase, President Zelensky was assigned code 1, President Putin was assigned code 2, and their absence was coded as a 3. After applying the test, weak but statistically significant negative correlations were obtained. Thus, for the correlation with the number of comments, a correlation coefficient $R = -0.2127$ and $p\text{-value} = 0.001544$ was obtained, indicating a tendency to increase the number of comments when the image of the two presidents is present. The result confirms the previously obtained correlation (Table I), where a positive and statistically significant correlation was obtained, indicating an increase in the number of comments when officials appear in the main frame. Similarly, a correlation coefficient of $R = -0.2731$ and $p\text{-value} = 0.00004186$ was obtained in the case of the same analysis in relation to the number of likes. The result indicates a tendency to increase the number of likes when, in particular, Zelensky and Putin appear, compared to their absence, and confirms the correlation obtained in Table II (which indicates a positive, statistically significant correlation between the presence of officials and the number of likes-hate). The fact that the results are consistent, although the coding system was different, validates the analysis performed and strengthens the conclusions obtained.

Most Used Categories:
All Images of Posts.
Overall Results.

The analysis grid was also applied to all the images in a post (if carousel). Each thematic category was marked in the grid only once for optimal results, although it could appear multiple times in the image carousel.

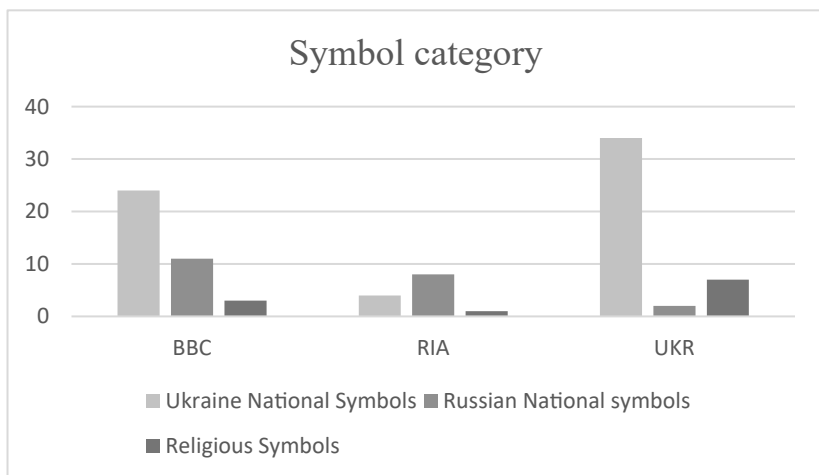


Fig. 4: Usage of the symbol category in the first image of the posts

The findings showed that civilians were the most commonly used category, with a recurrence of 184 times, including subcategories such as: civilian group—64, civilian woman—45, civilian man—41 and children—34. Other high recurring categories included: symbols (national and religious)—94 soldiers—56, military equipment—53, and ruins—47, officials—26, animals—25, while some categories were used

less frequently, such as landscapes – 15, combat images—13, international personalities—6, wounded—9 and casualties—2. As for the symbols, the category is divided into: Ukrainian national symbols—62, Russian national symbols—21 and religious symbols—11.

Most Used Categories:

All Images of Posts.

Results Specific to

Each News Agency

BBC and Ukrinform frequently used the category of civilians, Ukrinform—90 and BBC—88, while Ria Novosti only used it 6 times. Both BBC and Ukrinform focused on groups of civilians (BBC—27, Ukrinform—35), women (BBC—22, Ukrinform—21) and children (BBC—20, Ukrinform—13). The second most frequently used category by BBC News was that of symbols—38 appearances, with an emphasis on the symbols of Ukraine, which were present 24 times; Russian symbols were used 11 times and religious symbols 3 times. Also, the BBC insisted on the following categories: soldiers—32, military equipment—28, ruins—22 and officials—18.

After civilians, the second most used category by the Press Agency of Ukraine was that of symbols, with a recurrence of 43 times: 34 national symbols of Ukraine, 7 religious, and 2 national symbols of Russia. The top was completed by military equipment and ruins—24, followed by soldiers—23, and animals—18. The categories with the least recurrence were: landscapes—5, combat images—3, officials—3, injured—3.

The symbol category was the most common in Ria Novosti's posts. The Russian state-controlled news agency used it 13 times: Ukrainian national symbols—4, Russian national symbols—8, and religious symbols—1. The following recurring categories were: civilians—6, of-
ficials—5 and military equipment—1.

Interestingly, the use of images of the two presidents involved in the conflict, Putin and Zelensky, was not heavily emphasized by any of the news agencies. The BBC used both presidents' images equally, while Ukrainians only used Putin's image twice and Zelensky's image once. Ria Novosti also did not emphasize the category of officials, with Pu-
tin's image appearing 3 times and Zelensky's image appearing twice.

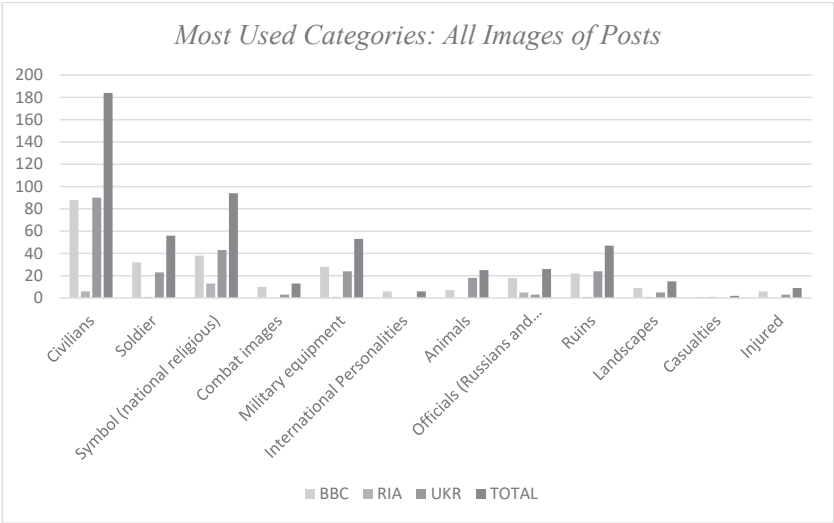


Fig. 5: Usage of the symbol category in the first image of the posts

Most Commented¹ Posts

A list of the posts with the most comments was created after analyzing all the materials. As expected, most interactions were recorded on the posts of BBC News due to their popularity and large number of Instagram followers compared to Ukrinform and Ria Novosti. The post with the highest number of comments, 28.500 (registered at the beginning of the research—May, 2022), dates from March 7 and features a Russian athlete wearing a letter Z on his chest, a symbol associated with the Russian invasion of Ukraine. In the context of the conflict, the symbol was linked to the national symbols of Russia due to its use in Russian propaganda. The athlete's gesture of wearing the symbol expressed support for the military action against Ukraine, which likely increased interaction on the post.



Figure 6: Most commented image. Source: BBC News

1 I want to clarify that the number of comments and likes mentioned in the article was recorded between May and June 2022 (the moment when the research was started). While some posts may have received more or fewer comments and likes since then, I have maintained the original data collected at that time.

The second image² with a higher number of comments was posted on February 24 and included the ruins category; it announced the start of Russia's invasion of Ukraine and registered 10.600 comments, likely due to the correlation between the ruins category and the significance of the event.

The next two posts with high interactions featured images of President Putin calling for volunteers to participate in the conflict in Ukraine (posted³ on March 11, with 9.262 comments) and threatening to use nuclear bombs in the conflict (posted⁴ on February 25, with 9.089 comments). The last image⁵ in the top five contains the landscape category and announced the killing of Ukrainian soldiers defending Snake Island (posted on February 27, with 7.006 comments).

Aside from the general top, I also selected the first three posts for each agency based on the number of comments. The first three posts from BBC News were the same as those in the main chart, so I reviewed only the materials from Ukrinform and Ria Novosti.

2 Post available at the web address: <https://www.instagram.com/p/CaX-ONuLNLIv/>

3 Post available at the web address: https://www.instagram.com/p/Ca9yAF7o_y0/

4 Post available at the web address: <https://www.instagram.com/p/CafIT8QIDiA/>

5 Post available at the web address: <https://www.instagram.com/p/CaacP2sKkV8/>

Ukrinform: Most Commented Posts

Ukrinform's posts received the lowest number of interactions among the three agencies. Thus, the most commented image included the ruins category and registered 17 comments⁶, which refers to the reprehensible gesture of the Russian army to vandalize by shooting a monument of the poet Taras Shevchenko in the town of Borodianka in the Kyiv region. The second post showed an image of an injured child due to the raids of the Russian army and collected 13 comments⁷, while the third post⁸ presented the marriage between a Ukrainian soldier and his girlfriend and garnered 11 comments.

Ria Novosti: Most Commented Posts

The image with the most comments from Ria Novosti includes the national symbols category and shows the Independence Monument of Ukraine. The February 24 post⁹, which has garnered 6.876 comments,

6 Post available at the web address: <https://www.instagram.com/p/CcAy-b1VqOVL/>

7 Post available at the web address: <https://www.instagram.com/p/CbQCNcYINxp/>

8 Post available at the web address: <https://www.instagram.com/p/Cb5c-mXZIYUs/>

9 Post available at the web address: <https://www.instagram.com/p/CaWv7CJMYQh/>

announced a military operation in Donbas. The following post in the top three includes the category of officials and features the image of President Putin. The post¹⁰, which has garnered 1.313 comments, dates from February 25 and announces that Russia is ready to send its delegation to Minsk for negotiations with Ukraine. The third image¹¹, which recorded 1003 comments, was posted on March 8 and included the category of officials, presenting the image of President Vladimir Zelensky. The post announces various steps to negotiate a compromise to end Ukraine's conflict.

Likes

Regarding the number of likes, the top five materials include posts exclusively from the BBC. The post with the most likes contains the category of officials and presents President Vladimir Zelensky. The material was published on Instagram on February 28, shortly after the outbreak of the conflict, and collected 572,958 likes.

The second post¹² of the top was also published on February 28 and contains the landscape category. The image brings Snake Island to the fore and announces that the Ukrainian soldiers who defended the small territory did not lose their lives. The post has garnered 418,474

10 Post available at the web address: <https://www.instagram.com/p/CaZ-rLPFMuU0/>

11 Post available at the web address: https://www.instagram.com/p/Ca1in79s-9_/

12 Post available at the web address: <https://www.instagram.com/p/Cah-7GX-on-B/>



Figure 7: Most liked post. Source: BBC News

likes. Another post¹³ that gathered an impressive number of likes, 404.205, dates from February 26, also shortly after the outbreak of the war, and contains the categories soldier and male civilian. The post announces the destruction of some apartments near Kyiv following the bombings organized by the Russian army. The top is closed by the posts of February 24 and 25, which announce the beginning of Russia's aggression against Ukraine (includes the ruins and soldiers categories¹⁴ and collected 390850 likes), respectively, the killing of the Ukrainian soldiers who defended Snake Island (includes the landscape category¹⁵ and collected 376146 likes).

13 Post available at the web address: <https://www.instagram.com/p/Cac-NAfKqg5a/>

14 Post available at the web address: <https://www.instagram.com/p/CaX-ONuLNliV/>

15 Post available at the web address: <https://www.instagram.com/p/CaacP2sKkV8/>

In addition to the overall top posts, I also specifically selected the first three posts with the highest number of interactions from each news agency. The posts from the BBC coincide with those in the main chart, so I reviewed only the materials from Ukrinform and Ria Novosti.

Ukrinform: Most Liked Posts

The most appreciated materials posted by Ukrinform recorded extremely few likes compared to the posts from the other two agencies under analysis. Thus, the most appreciated image¹⁶ dates from March 3 and has collected 335 likes. The post includes the soldier category and presents the marriage between a man and a woman who are part of the army of Ukraine. The second-most liked post captures a civilian setting fire to a straw representation of President Vladimir Putin. The material¹⁷ from March 6, 2022, collected 330 likes. The last post brings to the fore a happy event, namely the marriage of a soldier and his girlfriend. The post¹⁸, which includes the soldier category, collected 235 likes and was published on April 3, 2022.

16 Post available at the web address: <https://www.instagram.com/p/Capnd89oVdT/>

17 Post available at the web address: <https://www.instagram.com/p/Caxs1Lmookx/>

18 Post available at the web address: <https://www.instagram.com/p/Cb5cmXZIYUs/>

The material leading the top of the most appreciated posts is the one from February 25th. The post¹⁹ includes the category of officials and presents the image of President Putin; it has collected 13.935 likes. The next most liked post²⁰ is from February 24th. The image does not contain any of the categories included in the analysis grid but shows a map of Russia highlighting airports closed as a result of the conflict in Ukraine. The post has collected 7.605 likes. The post²¹ that ends the top is the one that recorded the most comments among the materials published by RIA. Dating from February 24, the image shows the Independence Monument of Ukraine (national symbols category) and has collected 6.876 likes.

Conclusions

The study examined how the conflict in Ukraine was visually framed on Instagram by BBC News, Ukrinform, and Ria Novosti. The main thematic categories used by the three agencies and the differences in visual framing of the conflict were identified. The analysis grid was independently applied to the first image of the post in order to test the

19 Post available at the web address: <https://www.instagram.com/p/CaZ-rLPFMuU0/>

20 Post available at the web address: <https://www.instagram.com/p/CaXG-FKHs9YL/>

21 Post available at the web address: <https://www.instagram.com/p/CaWv7CJMYQh/>

correlation between the categories present in the first image and the number of interactions: likes and comments, as well as the remaining images in the post (carousel type). Each category was marked in the grid only once, although it could appear multiple times in the image carousel.

The results indicate that the most frequently encountered category across the posts of the three agencies was that of civilians, suggesting a focus on the impact of war on the civilian population and the humanitarian perspective of the conflict. Regarding civilians, there is more emphasis on the group, practically implying a depersonalization of the war victims. Furthermore, the subcategory of women is used more frequently than men, while the subcategory of children is used significantly less, in line with a pattern already observed in the visual framing of conflicts, as “previous research suggested that the depiction of children during war has been a matter of great sensitivity and controversy” (Neumann and Fahmy, 2012: 190). Wells (2007) notes that many publications covering a conflict visually do not use images of injured or killed children, as they might “offend the sensibilities of readers” (as cited in Schwalbe, 2013: 253).

Unlike in other conflicts where „the photographs make it clear that non-combatants pay a price” (Greenwood and Jenkins, 2015: 222), most often the price being their life, in this case, the price depicted in the photographs refers more to the status of being a refugee—an existing pattern in war photography (see Parry, 2011: 195). As for civilians, the results are similar to those of the study conducted by Ojala et al. (2017), which analyzed the visual framing of the Ukraine conflict in various European newspapers during the period of 2014–2015. Images depicting deceased individuals were not shown, while civilians were

presented in images that featured symbols of destruction—ruins, coffins, flowers, candlelight—precisely to highlight “the human cost of the conflict” (Ojala et al., 2017: 480). Thus, in the present study as well, the emphasis was placed on generating empathy towards the situation of civilians. BBC and Ukrinform chose to present more Ukrainian civilians in empathic situations, such as being displaced from their homes or witnessing the destruction of their properties or expressing the pain of losing loved ones. On the other hand, Ria Novosti focused more on the material difficulties faced by Russian citizens as a result of the conflict.

Additionally, national and religious symbols represent the second significant category, highlighting the importance of identity and cultural significance in the context of war. On the other hand, there is noticeably less usage of categories directly associated with violence and destruction, such as combatants, military equipment, and ruins. This indicates a deliberate avoidance of bloody representations and morbid aspects of war. Although there are documented cases of injuries and casualties resulting from the conflict, they represent exceptions and have limited presence in the selected images. Thus, the main focus is on the representation of civilians and symbols, while categories associated with the horrors of war are neglected or underrepresented, a fact confirmed by the existing literature, which shows that dramatic images are also avoided in other wars (Griffin, 2004; Fahmy and Kim, 2008; Griffin, 2010).

Regarding the visual framing of the conflict from the perspective of each agency, it can be observed that each media institution adopts a specific approach but generally avoids excessive focus on violence and destruction, preferring to convey humanitarian, political, and

symbolic aspects of the conflict. There are also several similarities between the approaches of the British and Ukrainian news agencies. BBC and Ukrinform aim to highlight the impact of the conflict on the Ukrainian population, particularly showcasing groups of civilians seeking refuge from the war. In contrast, RIA emphasizes a more patriotic tone, focusing on national symbols and officials.

The following categories with increased recurrence in BBC News posts are symbols (national and religious), followed by soldiers, military equipment, ruins, and Russian and Ukrainian officials. It is interesting to note that in BBC's posts, soldiers are not presented on the battlefield, but rather "on backstage", a pattern found in other cases of visual framing of conflicts (Griffin, 2010: 30). Typically, their life stories are visually highlighted, or they are portrayed as heroic figures protecting civilians. Additionally, a crucial aspect is that the images primarily feature Ukrainian soldiers and not Russian soldiers, which once again emphasizes the media support BBC provides to Ukraine. A similar aspect was highlighted during the Gulf War when publications such as Time, US News and World Report, and Newsweek focused on American troops (Griffin and Lee, 1995: 818), contributing to influencing public opinion.

Regarding officials, BBC remains impartial and uses images of both presidents equally. However, it is surprising that Ukrinform does not utilize the image of President Zelensky, a true international icon of the conflict. The Ukrainian news agency only uses the image of its own president once, while the image of President Putin appears twice on Ukrinform's Instagram account.

Furthermore, according to Ojala et al. (2017), by using the category of international personalities, a war can be visually framed as a

„geopolitical conflict” (ibid.: 490). Thus, „by visually highlighting the presence of international leaders,” those images reinforce „a sense of the power and responsibility of other states in the Ukraine conflict and its resolution” (Ojala et al., 2017: 490). In this regard, BBC News is the only media institution among the three that visually frames the conflict as a geopolitical one since it uses images of international personalities, albeit with lower recurrence compared to other categories. Their posts include images of VIPs like Arnold Schwarzenegger, Joe Biden, Boris Johnson, Michael David Rosenberg (Passenger), personalities who publicly condemn the conflict in Ukraine, thus influencing public opinion.

The Ukrainian news agency also focuses on the category of civilians, with greater attention given to groups, men, and women. In their posts, a significant recurrence of symbols, military equipment, ruins, and soldiers can be observed. Categories with lower recurrence include animals, combat images, officials, and the injured. Interestingly, in the images posted by Ukrinform, state officials are present in a limited number, while President Zelensky’s image appears only once, in contrast to the two appearances of President Putin. Additionally, the news agency avoids showing images of victims who died or were injured as a result of the war.

On the other hand, Ria Novosti places special emphasis on national and religious symbols, as well as the categories of civilians and officials. The agency avoids using combat images and other categories associated with militarization and destruction. Through their approach, Ria Novosti aims to provide a perspective on the conflict that emphasizes symbolic, humanitarian, and political aspects, while avoiding a detailed presentation of violence and its consequences. By

using national symbols, both the Russian and Ukrainian news agencies visually frame the conflict as a national power struggle (see Ojala et al., 2017), in the hope of influencing public opinion and mobilizing support for the war. Each of the two media institutions promotes its own symbols, as expected, while BBC News predominantly uses Ukrainian symbols, visually supporting Ukraine in framing the conflict as a national power struggle.

It is interesting to note that the research findings initiated by Anghel (2023) are similar. The category of civilians is the most frequently encountered throughout the video content, followed by the categories of symbols, ruins, and combatants (Anghel, 2023: 175). The same categories are also used specifically by BBC and RIA, with both media institutions most commonly using the categories of civilians, symbols, ruins, and combatants (Anghel, 2023: 175). Regarding officials, BBC News featured President Zelensky in frames 12 times, while President Putin appeared 3 times. As expected, RIA used the subcategory of President Putin without ever using an image of President Zelensky (Anghel, 2023: 175).

Based on the analysis, we can say that the conflict in Ukraine is visually framed as a „clean war” (Fahmy and Kim, 2008: 435) or „sanitized war” (Griffin, 2010: 27) with an emphasis on civilians, but without primarily presenting images of killed or injured individuals. The existing literature indicates similarities with the visual framing of the first Gulf War, where very few images of deceased or injured individuals appeared in the media (Newton as cited in Fahmy and Kim, 2008, p. 435; Griffin and Lee, 1995: 813), with a focus placed on civilians, configuring more of a „human interest frame” (see Schwalbe, 2013: 244; Schwalbe and Dougherty, 2015).

Following the results obtained by applying the Pearson test, relevant conclusions can be drawn regarding the influence of visual categories on user interaction on social media. The analysis indicates that the presence of images featuring civilian men in the first image of posts can impact the number of comments generated by users. However, it should be noted that this increase is not statistically significant and cannot be considered guaranteed in generating interactions. The study findings indicate that posts featuring national or religious symbols tend to receive fewer comments and likes. Conversely, posts containing images of political officials, particularly Presidents Zelensky and Putin, generate more user engagement. This suggests that people are more inclined to respond positively to posts that highlight political figures within the context of the ongoing conflict, with a slight preference for Zelensky. Additionally, posts showcasing a group of civilians as the primary image tend to receive fewer likes. On the other hand, the presence of landscape or natural scenery images has a positive impact on the number of likes.

These findings are consistent with the results obtained in the previous video content analysis conducted by Anghel (2023), which highlighted the same decreasing trend in interactions when the first frame of the video material featured a group of civilians. Moreover, a positive correlation has been observed between the presence of the category „civilian male” and the number of likes, and in the case of video materials, the „combat images” category showed a positive correlation with both the number of likes and the number of comments. These findings underscore the importance of visual categories in influencing user interactions and suggest that certain types of images can stimulate or inhibit user engagement and participation in a specific context.

It is interesting that the data regarding the increase or decrease in interactions based on the depicted category differs, in certain situations, from those recorded by Makhortykh and Sydorova (2017). Their research indicated that users tend to appreciate „those images that articulated light-hearted aspects of the conflict, such as pets or symbols... as well as combatants” (ibid.: 376), which did not occur in the first two months of the current conflict in Ukraine. However, there are also some similarities, as images featuring „political or military leaders” (Makhortykh and Sydorova, 2017: 376) receive more likes than others, as is also the case in the present study. It is possible that interactions are influenced by existing trends in social media, such as the focus on political and military personalities responsible for managing the conflict, figures known internationally.

The present study makes a significant contribution in understanding how the mass media in general, and news agencies in particular, influence the public’s perception of armed conflicts and their victims. The research findings can be useful in gaining a better understanding of how the media selects and presents images related to conflicts, enabling us as the audience to approach news consumption and source evaluation with a more critical and informed mindset. Moreover, this study highlights the potential gaps in visual reporting during conflicts, underscoring the significance of seeking information from multiple sources to attain a more balanced perspective. The study’s findings indicate that specific types of visuals, such as images depicting civilians affected by the conflict, national symbols, and key figures involved, have a substantial influence on user engagement in terms of likes and comments. This suggests that these visual categories effectively capture the audience’s attention, resulting in active responses and interactions, or conversely, diminished engagement. These

findings emphasize the important role of visual categories in online communication, serving as powerful tools for news agencies, organizations, and individuals who aim to convey messages and encourage active participation on issues related to armed conflicts.

The research also examines how the three news agencies construct the narrative about the Ukrainian conflict with the help of social media. The prominent categories, including civilians and national symbols, reveal the particular focus on showcasing the direct impact on the lives of individuals affected by the conflict and the emotional connection that these images can foster among the audience. By emphasizing the human aspects of the conflict, a profound emotional impact can be generated, compelling the audience to become more engaged and concerned about the situation in Ukraine.

However, even the mere existence of this type of construction is a gain for the audience, as it requires them to engage with a broader field of understanding when encountering images of different conflicts. This understanding of the mechanisms through which a visual narrative can be constructed can help overcome the filter bubbles that can hold each of us captive.

The study has several limitations. Firstly, it focused solely on three news agencies: BBC News, Ukrinform, and Ria Novosti. Furthermore, the analysis period was limited to two months from the onset of the conflict. The short analysis period was due to the research being initiated at the end of April and the beginning of May when no additional data was available at that time. Moreover, the analysis focused solely on the visual aspect and did not incorporate a multimedia approach that would have included the analysis of the post captions. The data would have been more conclusive in understanding the correlation

between the category found in the post and the number of interactions if the textual component accompanying the image was taken into account. For example, posts with the highest number of interactions, likes, and comments, could have been better understood by integrating the caption into the analysis. One image that generated a significant number of interactions included the categories of civilian and national symbol. However, the high number of interactions could have been better understood by considering the caption, which explained that it was a Russian athlete participating in an international sports competition wearing the letter „Z,” a symbol associated with the Russian military in the Ukrainian conflict.

These limitations suggest avenues for future research. For instance, the text of each post could be analyzed to identify specific news frames that can be correlated with the visual research findings, providing a more holistic view of the analyzed materials. Additionally, considering that the current article focused exclusively on static images, and there is a separate research that focused on the same agencies, corpus, and period, but exclusively on the video aspect (see Anghel, 2023), a comparative analysis of both datasets along with a textual analysis of the captions accompanying the images in the posts would be appropriate.

References

- Adornato, A. (2021). *Mobile and social media journalism: A practical guide for multimedia journalism* (2nd ed.). Routledge.
- Anghel, G. (2023). Social media and visual framing of the conflict in Ukraine: BBC News, Ria Novosti, and Ukrinform. In R. Surugiu & G. Stepanov (Eds.), *Conjuncturi digitale* (pp. 165–181). Tritonic.
- Boaz, C. (2005). War and foreign policy framing in international media. *Peace Review*, 17(4), 349–356. <https://doi.org/10.1080/10402650500374603>
- Bryant, J., & Miron, D. (2004). Theory and research in mass communication. *Journal of Communication*, 54(4), 662–704. <https://doi.org/10.1111/j.1460-2466.2004.tb02650.x>
- Fahmy, S. (2010). Contrasting visual frames of our times: A framing analysis of English- and Arabic-language press coverage of war and terrorism. *International Communication Gazette*, 72(8), 695–717. <https://doi.org/10.1177/1748048510380801>
- Fahmy, S., & Kim, D. (2008). Picturing the Iraq War: Constructing the image of war in the British and US press. *International Communication Gazette*, 70(6), 443–462. <https://doi.org/10.1177/1748048508096142>
- Fahmy, S. S., Taha, B. M., & Karademir, H. (2022). Journalistic practices on Twitter: A comparative visual study on the personalization of conflict reporting on social media. *Online Media and Global Communication*, 1(1), 23–59. <https://doi.org/10.1515/omgc-2022-0008>
- Fahmy, S., & Wanta, W. (2007). What visual journalists think others think: The perceived impact of news photographs on public opinion formation. *Visual Communication Quarterly*, 14(1), 16–31. <https://doi.org/10.1080/15551390701361632>
- Gamson, W. A., & Modigliani, A. (1987). The changing culture of affirmative action. In R. Braungart (Ed.), *Research in political sociology* (pp. 137–177). Jai Press.

Gitlin, T. (1980). *The whole world is watching: Mass media in the making and unmaking of the New Left*. University of California Press.

Greenwood, K., & Jenkins, J. (2015). Visual framing of the Syrian conflict in news and public affairs magazines. *Journalism Studies*, 16(2), 207–227. <https://doi.org/10.1080/1461670X.2013.865969>

Griffin, M. (2004). Picturing America's "War on Terrorism" in Afghanistan and Iraq: Photographic motifs as news frames. *Journalism*, 5(4), 381–402. <https://doi.org/10.1177/1464884904044201>

Griffin, M. (2010). Media images of war. *Media, War & Conflict*, 3(1), 7–41. <https://doi.org/10.1177/1750635210356813>

Griffin, M., & Lee, J. (1995). Picturing the Gulf War: Constructing an image of war in *Time*, *Newsweek*, and *U.S. News and World Report*. *Journalism & Mass Communication Quarterly*, 72(4), 813–825. <https://doi.org/10.1177/107769909507200405>

Kemp, S. (2023, January 26). *Digital 2023: Global overview report*. Datareportal. <https://datareportal.com/reports/digital-2023-global-overview-report>

Lecheler, S., & De Vreese, C. H. (2018). *News framing effects* (1st ed.). Routledge. <https://doi.org/10.4324/9781315208077>

Makhortykh, M., & Sydorova, M. (2017). Social media and visual framing of the conflict in Eastern Ukraine. *Media, War & Conflict*, 10(3), 359–381. <https://doi.org/10.1177/1750635217702539>

Neumann, R., & Fahmy, S. (2012). Analyzing the spell of war: A war/peace framing analysis of the 2009 visual coverage of the Sri Lankan civil war in Western newswires. *Mass Communication and Society*, 15(2), 169–200. <https://doi.org/10.1080/15205436.2011.583192>

Newman, N. (2022, June 15). Overview and key findings of the 2022 digital news report. *Reuters Institute*. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022/dnr-executive-summary>

- Ojala, M., Pantti, M., & Kangas, J. (2017). Whose war, whose fault? Visual framing of the Ukraine conflict in Western European newspapers. *International Journal of Communication*, 11(25), 474–498. <https://ijoc.org/index.php/ijoc/article/view/5917>
- Parry, K. (2011). Images of liberation? Visual framing, humanitarianism, and British press photography during the 2003 Iraq invasion. *Media, Culture & Society*, 33(8), 1185–1201. <https://doi.org/10.1177/0163443711418274>
- Ragozin, L. (2019, March 16). Annexation of Crimea: A masterclass in political manipulation. *Al Jazeera*. <https://www.aljazeera.com/opinions/2019/3/16/annexation-of-crimea-a-masterclass-in-political-manipulation>
- Reynolds, A. (Ed.). (2016). *Social media as a tool of hybrid warfare*. NATO Strategic Communications Centre of Excellence.
- Schwalbe, C. B. (2006). Remembering our shared past: Visually framing the Iraq War on U.S. news websites. *Journal of Computer-Mediated Communication*, 12(1), 264–289. <https://doi.org/10.1111/j.1083-6101.2006.00325.x>
- Schwalbe, C. B. (2013a). Visually framing the invasion and occupation of Iraq in TIME, Newsweek, and U.S. News and World Report. *International Journal of Communication*, 7, 239–262. <https://ijoc.org/index.php/ijoc/article/view/924>
- Schwalbe, C. B. (2013b). Visually framing the invasion and occupation of Iraq in TIME, Newsweek, and U.S. News and World Report. *International Journal of Communication*, 7, 239–262. <https://ijoc.org/index.php/ijoc/article/view/924>
- Schwalbe, C. B., & Dougherty, S. M. (2015). Visual coverage of the 2006 Lebanon War: Framing conflict in three US news magazines. *Media, War & Conflict*, 8(1), 141–162. <https://doi.org/10.1177/1750635215571204>
- Ukraine conflict: What we know about the invasion*. (2022, February 24). *BBC News*. <https://www.bbc.com/news/world-europe-60504334>
- Zeitsoff, T. (2017). How social media is changing conflict. *Journal of Conflict Resolution*, 61(9), 1970–1991. <https://doi.org/10.1177/0022002717721392>

Part II

The Digital Transformation of Society

True Investment: Post-Authentic Honesty on Social Media

Hans-Georg
Moeller

Introduction

Methodologically, this paper differs from most contributions to this volume, as the author is not a social scientist but an academic philosopher. In accordance with a common approach in this discipline, this paper is largely written in an essayistic style and aims to develop a theoretical framework that can be applied to a broader analysis of social and psychological phenomena and existential questions. It does not rely on new empirical research, the collection of empirical data, or a comprehensive overview of scientific studies. Instead, it critically reflects on what the author perceives to be a conceptual weakness in both academic and non-academic discourses on social media. Specifically, it suggests a reevaluation of the often-prevalent semantics of “authenticity” in those discourses and advocates for their eventual replacement with a semantics of “proficiency.”

Hans-Georg Moeller, Full Professor, Department of Philosophy and Religious Studies,
University of Macau, Macau ORCID: 0000-0001-6185-3224, email:
hmoeller@um.edu.mo

DOI: <https://doi.org/10.5281/zenodo.14203888>

A starting point for critiquing the semantics of authenticity and its use in connection with mass and social media is the paradoxical ascription of “authenticity” to Donald Trump, the former (and perhaps future) American president—or, more accurately, to his political profile. The influential political scientist Francis Fukuyama once defined Trump as “the perfect practitioner of the ethics of authenticity that defines our age,” specifying that “he may be mendacious, malicious, bigoted, and unpresidential, but at least he says what he thinks” (Fukuyama 2018: 119). Fukuyama’s definition is self-contradictory: how can someone be perfectly authentic and at the same time mendacious? This contradiction has been explicitly addressed in an article published in *Newsweek* titled “Donald Trump is ‘Perhaps the Most Authentic’ President Ever, Despite Constantly Saying Things That Aren’t True, Former White House Aide Cliff Sims Says” (Hutzler 2019). Referencing the famous CNN anchor Anderson Cooper, the *Newsweek* article highlights the surprising fact that Trump is often described as authentic not despite, but because of, his frequently inaccurate or exaggerated statements, which it calls the “great paradox of the Trump presidency.” However, the present essay argues that this paradox of authenticity arises not as much from the illustrious persona, or profile, of Donald Trump, but rather from a misuse of the notion of authenticity. The (social) media profiles of Trump, like those of many other individuals, should not be measured by the outdated and therefore inappropriate standards of an “ethics of authenticity,” but rather by a currently emerging “ethics of prolificity.”

Once Donald Trump’s (social) media profiles are understood as operating within a post-authentic framework of prolificity, the paradox of authenticity vanishes. While Trump certainly does not display his “authentic self” in his media performances, it can be reasonably assumed

that he is indeed “truly invested” in the profiles he curates. Such true investment represents a post-authentic form of honesty that, from the perspective of authenticity, makes Trump—and by no means only him—appear as a paradoxical character who is “weirdly honest about his lying” (Lopez 2016).

Identity Technologies and Different Kinds of Honesty

As outlined in more detail elsewhere (Moeller and D’Ambrosio, 2021), the formation of both individual and collective identity can be achieved through the existential and social practice of different “identity technologies” named “sincerity,” “authenticity,” and “proficiency.”

As an identity technology, sincerity operates with an orientation to social roles. In traditional societies, and in many social contexts today, people shape a sense of selfhood by aligning with the rather fixed norms and expectations of roles they are either born into—such as gender roles, roles in the family (e.g. “brother” or “daughter”), or roles derived from social status (e.g. “aristocrat”)—or roles they adopt, for instance, through marriage (e.g. “wife” or “husband”) or as a profession (e.g. “farmer”). Other such roles with fixed expectations may include religious roles or roles in sports or the military. Typically, such roles are performed and validated in the context of co-present peers, e.g., family members, members of a religious or military group, or a sports team. Under such conditions, the successful formation of identity, both from the perspective of the individual and of the group, depends

on *sincere commitment* to one's role. Such sincere commitment consists of behavioral and psychological components. One must, for instance, not only act as is expected from a "wife", but also feel and think in line with this role in order to sincerely identify as "wife." While, under conditions of sincerity, people will often commit to a multiplicity of roles, these roles are nevertheless typically perceived as somehow congruent with one another, for instance by association with a specific gender: Religious, military, and professional roles, for example, tended to be aligned with certain gender roles.

In modern society, the orientation to fixed social roles became increasingly impractical. Accelerated social dynamics fueled by amplified mobility, rapid technological innovations, a capitalist economy, and growing social complexity challenged the stability of traditional roles. Rather than sincerely committing to lifelong externally provided roles, people oriented themselves inward in their search for identity. Individualism was on the rise and resulted in what the philosopher Charles Taylor labeled the "age of authenticity." (Taylor, 2018) From the perspective of the age of authenticity, sincere commitment to fixed social roles appears as mere conformity and as an obstacle to true authenticity. One's authentic self, it is now assumed, is unique and internal, and does not correspond to prefabricated social expectations. Accordingly, authenticity consists in the pursuit of originality. This originality can be associated with an underlying individual self that one has "always been" but that is in need of *discovery* because it has been obscured by social conventions. Alternatively, it can be associated with a self that has yet to be individually *created* through one's own efforts, often in defiance of social expectations. Innumerable modern novels, including the whole genre of *Bildungsroman*, described such efforts of individuals to "find" or "make" themselves by

shedding the “masks” that society imposed on them. Given its emphasis on originality and uniqueness, authenticity cannot be validated only by social approval. Instead, one’s authenticity needs to be *recognized*—in Hegel’s sense of the term (Hegel, 2017)—by other authentic individuals. Expressed in philosophical terminology, authenticity is “intersubjective.”

Authenticity, however, became increasingly fragile when society became postmodern. One way of defining postmodernity is to conceive of it according to the philosopher Guy Debord, as a “society of the spectacle” (Debord, 1994). According to Debord, the traditional economy, which centered on the commodification of material goods, has been replaced by an economy centered on the commercial exchange of images—which is what Debord calls the “spectacle”. The spectacle is not only exemplified by the rise of a gigantic show business in the form of a media industry, which produces and trades images rather than “real” objects, but also by the supplementation of material or “real” goods with *brands*. In the contemporary economy, the “logo” (Klein, 2010) determines the value of a commodity more than its material or use value. According to Moeller and D’Ambrosio (2021), the branding of goods in the economy has been extended to the branding of individual and collective identities in the form of profiles. Profiles are curated self-images that are projected to an often-anonymous public. Like the traditional orientation to social roles, the orientation to profiles is external, and not directed inward to an “original” self. Instead, profiles are validated in social-validation feedback loops. On social media, for example, profiles are validated in the form of views and likes; the profile of an AirBnB host is validated by their “Super-host” status; academic profiles are validated by publications in highly ranked venues or by citation numbers; political profiles are validated

by elections and media attention, etc. In comparison with traditional social roles, however, social expectations regarding profiles are highly volatile, unstable, and difficult, if not impossible, to predict. Importantly, one cannot be born into a profile but must actively curate one. Moreover, profiles are mostly not validated by co-present peers, but by a “general peer,” simulated, for instance, in the academic “double-blind” peer-review process or manifested in a non-specific multitude of followers on social media, or voters in the spectacle of a political contest. Unlike traditional roles in the mode of sincerity, profiles are often unrelated to one another and not closely aligned: There may be little coherence, for instance, between an academic’s professional profile and her profile on Tinder.

Different identity technologies are aligned with different moral regimes, or different ethical frameworks. Sincerity is typically tied to a morality of honor. Honor can be defined as the moral acclaim with which sincere commitment to roles is rewarded. A woman sincerely committing to the roles of “faithful wife” and “dedicated daughter”, for instance, may be regarded as an honorable role model for other women. Similarly, devoted monks, courageous soldiers, or honest shopkeepers can be regarded as moral exemplars. Accordingly, under conditions of sincerity, “honor” or “truthfulness” is found in reliable role behavior and deep emotional and intellectual role identification.

In authenticity, moral acclaim is due to those who are regarded as “true to themselves.” Since authenticity is found in the pursuit of originality, originality becomes an eminent moral value. Authenticity values “making a difference,” or being different. It praises innovation and condemns any form of “plagiarism.” What is more, a set of “human rights” singles out individualistic rights to “self-determination,”

“autonomy”, or “free speech” as especially valuable. Under conditions of authenticity, honesty lies in expressing one’s personal thoughts and feelings freely and accurately, and in representing one’s “true self” in an open and creative way in public. Like sincerity, authenticity is not merely an identity technology, but, at the same time, also a moral regime that gives rise to an ethical vocabulary. The ethical vocabulary of authenticity differs from the ethical vocabulary of sincerity by detaching honesty from commitment to social roles and orienting it instead to the truthful expression of a unique inner self.

Since prolificity is a more recent identity technology than sincerity and authenticity, it lacks as of yet an elaborate ethical vocabulary. As indicated by the ascription of authenticity to the openly mendacious Donald Trump, the semantics of the age of authenticity is still often paradoxically applied to describe prolificity phenomena. However, as in the case of Donald Trump, the inappropriateness of this semantics is becoming increasingly tangible.

In this essay, I propose to conceive of honesty under conditions of prolificity as *true investment* in one’s profiles. Similar to role commitment, profile investment is not oriented inward but outward. It consists in behavioral, emotional, and intellectual alignment with a profile. However, as outlined earlier, profiles differ significantly from traditional roles. Importantly, unlike roles that are mostly ready-made and often seemingly “naturally” acquired (such as traditional gender roles), profiles are in need of constant curation. They contain, so to speak, an element of authenticity, because of the importance of personal choice in their curation. Unlike in authenticity, however, profiles are not mere external expressions of an inner self. Their validation does not come from within but emerges in social validation feedback loops.

The academic peer review process is a good example of the constitution of true profile investment in distinction from authentic insistence on originality. Academics today are typically required to revise their “self-expressions”, i.e., their journal articles or book manuscripts, in response to input from anonymous peers. What is more, these peers are in turn expected to formulate their input not simply authentically, but in accordance with certain standardized parameters for academic reviews. This very paper, for instance, is produced on the basis of a rather strictly determined format and explicitly communicated guidelines regarding its contents. As an academic working under conditions of proficility, I, the author, am expected to embrace the prescribed format of this paper and incorporate, at least partly, the revisions that will be suggested by my peers. Even though the academic industry still uses—like the analysts of Donald Trump—a vocabulary of authenticity that celebrates originality, it demands in practice the formation of a professional identity that emerges through elaborately institutionalized social validation feedback loops. Guided by the peer review process and other social validation mechanisms, academics curate professional profiles that they eventually become truly invested in. Importantly, this true investment can only take shape *a posteriori*, i.e., after receiving *a priori* unpredictable feedback. True investment cannot be based on pure “originality,” but relies on adaptation to peer input and the formation of a self-image in accordance with socially established communication platforms and their standardized communication patterns. Here, honesty lies in behaviorally, emotionally, and intellectually living up to the profiles one is curating and displaying once they have been validated.

The Ascription of Inauthenticity and Authenticity to Social Media Use

As mentioned, the inappropriateness of ascribing an “ethics of authenticity” to proficility phenomena is evidenced by the often-paradoxical effect of this ascription: profiles, like those of Donald Trump, can easily appear as both authentic and inauthentic. The same paradox resurfaces in more general assessments of profile curation on social media: On the one hand, a plethora of both academic and non-academic texts describe social media use as tendentially, if not essentially, inauthentic. On the other hand, successful profile curation on social media is often attributed to the virtue of authenticity.

Academic literature tends to express a suspicion that social media threatens or diminishes authenticity. An overview of several studies on the subject by Kaily Marlow is titled “Social Media and The Negative Impact It Has Had on The Authentic Self” (Marlow, 20210). Referring to a study by David C DeAndrea and Joseph B. Walther (2011), Marlow states that “online, people tend to portray themselves in the way that they wish to be seen rather than the way they actually are as a form of impression management” and that “people will alter these different portrayals of themselves based on the audience receiving the content to make themselves more digestible for the viewer.” For Marlow, this means “that it is near impossible to achieve an authentic presentation online due to the inconsistencies between a user’s self-image and the way others perceive them.” Consequently, she concludes that “the ways in which we construct our self-presentation

online provide almost no room for authenticity” (Marlow 2021). In a very similar vein, another survey of academic literature on online platforms by Leah Brown (2020), titled “Social media has helped in manufacturing deceptive and inauthentic identities online,” declares that social media in general facilitates “the spread of deception and inauthenticity of a person’s identity”. Surveys like those by Marlow and Brown seem to suggest a broader consensus in media studies about a more or less fundamental incompatibility between social media use and the formation of an authentic identity: social media use, it is indicated, seriously undermines a person’s ability to become authentic, or, more precisely, is a breeding ground for inauthenticity.

In sometimes even more dramatic fashion, countless non-academic texts decry the loss of authenticity caused by social media. Here are some randomly chosen examples: *The Standard*, a “high school student news site of The American School in London,” features the headline “Social media limits our authenticity” (Meidar, 2023). *The Mentor*, “the student news site of Manhattan High School”, states very similarly: “Social media creates inauthenticity among teens” (Elliott 2019). The online edition of the journal *Entrepreneur* promises to explain “Why Everyone and Everything on Social Media Is Fake” and warns its readers that “social media—the first mass medium to allow people to publicly share their thoughts, feelings and lives with others—has become dangerously inauthentic” (Deutsch, 2018). Social media, it seems, inevitably infects (almost) all its users with the virus of inauthenticity.

To reiterate: As Marlow (2021) put it, inauthenticity on social media is apparently due to the fact that “people will alter these different portrayals of themselves based on the audience receiving the content to

make themselves more digestible for the viewer”. This is to say that social media profiles are curated in social validation feedback loops with a general peer—and that this kind of profile curation is the root cause of the inherent inauthenticity of social media identity. Somewhat surprisingly, however, another body of academic literature, accompanied by a large number of non-academic texts, does not find profile curation on social media through social validation feedback loops essentially inauthentic at all. To the contrary, for many observers, success on social media, hinges on the display of authenticity. From this perspective, social media does not threaten authenticity or makes inauthenticity inevitable. Instead, social media is regarded as facilitating and enhancing authenticity. It is suggested that those who thrive in the digital world do so precisely because they know how to make use of its potential to become and be truly authentic.

An academic paper by Jung Ah Lee and Matthew S. Eastin (2021), for instance, concludes that “public perceptions of the authenticity of social media influencers (SMIs) are a key driver of the latter’s persuasiveness as brand endorsers.” Referring to an empirical study on some of the most popular German social media influencers published by the consultancy firm Batten and Company (batten.admin 2021), the online edition of the German newspaper *Frankfurter Allgemeine Zeitung* goes a step further and reports on social media influencers that not just their perceived but their actual “authenticity is the primary reason for their success” (Anton 2021, my translation). The authenticity of influencers is described in this way: They “only do advertisement deals fitting their image. With a lot of personal content, they create proximity so that their followers can identify especially well with their accounts. A clear brand core generates promises that the influencers keep.” (Anton 2021, my translation). The same “need to be authentic”

on social media is emphasized by the marketing agency Yakkety Yak in an essay on its website titled “Authenticity in Social Media: Being Authentically YOU” (2019). Similarly, another marketing agency named Aspiration Marketing introduces potential customers on its website to the means of “Achieving Authenticity In Social Media Marketing” (2023). Especially in the area of marketing, the semantics of authenticity is applied to describe, and prescribe, effective self-presentation on social media. In this social sphere at least, social media is not regarded as a forum of inauthenticity but rather celebrated as a novel space where authenticity will be rewarded and thrive.

The contradictory descriptions of social media as fostering inauthenticity and as promoting authenticity show, in my view, not as much the bias of the respective observers as the inadequacy of the authenticity semantics, which is applied in a rather unreflected manner. Surely, it seems odd that “advertisement deals fitting their image” or the intentional creation of “proximity so that their followers can identify especially well with their accounts” are regarded as signs of the *authenticity* of influencers. Rather than acts of authentic self-expression, such behavior is better explained in the context of the curation of effective profiles by means of social validation feedback loops. However, it is almost equally misleading to label the activity of “portraying oneself in the way one wishes to be seen,” of which Marlow (2021) speaks, as *inauthentic*. More precisely, such an activity is, from the start, *non-authentic*, or *post-authentic*. Everyone knows that different online platforms (such as academia.edu, Tinder, or eBay) require individuals to present “different portrayals of themselves based on the audience receiving the content to make themselves more digestible for the viewer” (Marlow, 2021). The necessity, however, to curate highly diverse profiles on highly diverse social media, is not, as the

marketing agencies talking about “authenticity” rightly imply, equal to being inherently deceptive or fake. To be guided in one’s self-presentation on social media not primarily by an elusive “inner self” but instead by the structures of the platform one uses and by the feedback one receives is not *per se* dishonest. Identity formation under conditions of prolificity generates ways of being honest or dishonest that differ from the possibilities of being honest or dishonest under conditions of authenticity, as well as from the possibilities of being honest or dishonest under conditions of sincerity. Under conditions of prolificity, identity work consisting in choosing “advertisement deals fitting one’s image” or creating “proximity so that followers can identify especially well with one’s social media accounts” is neither honest in the sincerity sense of commitment to traditional roles nor honest in the authenticity sense of being true to one’s “inner self,” but it may well be honest in the prolificity sense of being truly invested in one’s (social media) profiles.

True Investment

To analyze honesty beyond authenticity and inauthenticity (and beyond sincerity and insincerity), we can return to the “weirdly honest liar” Donald Trump. During his election campaign in 2016, Trump explained how his core electoral catchphrase “Drain the Swamp” was coined: “Funny how that term caught on, isn’t it? I tell everyone: I hated it! Somebody said, ‘Drain the swamp.’ I said, ‘Oh, that’s so hokey. That is so terrible.’ I said, ‘All right. I’ll try it.’ So, like a month ago, I said, ‘Drain the swamp.’ The place went crazy. I said, ‘Whoa. Watch this.’ Then I

said it again. Then I started saying it like I meant it, right? And then I said it, I started loving it.” (Lopez, 2016)

There is no reason to believe that Trump is dishonest in this description. However, at the same time, it is a description of non-authenticity. Trump does not claim to be the originator of his catchphrase and, what is more, he also admits that he originally did not like it at all. He could not identify with it and perceived it as “hokey.” Only after he tried it out, and after its enthusiastic validation by his general peer—his voters—he began liking and eventually “loving it.” This confession of love testifies to a moment of true investment. Even without any claim to originality, Trump’s identification with the slogan “Drain the Swamp” is yet just as honest as, let’s say, the eventual identification of an academic author with the revisions suggested to her paper by anonymous peer reviewers. She might have at first not liked to do these revisions, and only wrote them down like she meant them. But when the paper got cited a lot, she started loving it.

Under conditions of prolificity, honesty is detached from authenticity and acknowledges the importance of social validation feedback loops. This very acknowledgment of the contribution of others to one’s identity formation makes the true investment of prolificity less narcissistic—less self-centered—than the authentic urge to be first and foremost “true to oneself”. In fact, the common accusation that social media use is not just inauthentic, but also narcissistic (see, for instance, Sodha, 2021) can be regarded as just another variation of the misplaced projection of an authenticity framework to prolificity phenomena. With the ascription of absolute priority to one’s inner self, authenticity is probably more prone to narcissism than prolificity. In other words: while prolificity can of course also be dishonest and

pathological (further analyses of Donald Trump's behavior could provide numerous examples of prolific dishonesties and pathologies), inauthenticity and narcissism are "vices" closely tied to a moral regime of authenticity: inauthenticity is regarded as the failure to be authentic, whereas narcissistic obsession with oneself may appear as a kind of over-the-top authenticity. As opposed to pathologically self-centered narcissism, true investment in one's profile, however, goes along with a keen perception of the sensitivities and concern for the reactions of others.

A good example of the pursuit of true investment on social media under conditions of prolificity is the semi-biographical short movie *Love and Revenge* (2021) by the young Saudi Arabian filmmaker Anhar Salam. The protagonist of the movie is an urban Saudi Arabian teenage girl named Doody. Doody is a fan of pop music and the singer Rihanna. Inspired by pop culture and social media, she curates her own avatar on Instagram, a quite sexualized, extraverted blonde woman, and shares it with her peers. Her account, however, is hacked, scandalized by conservative Saudi Arabians, and censored by the authorities.

While, on the one hand, the movie is about oppression, especially of women, in a conservative and authoritarian society, it also centers explicitly on the existential topic of identity formation. Specifically, it can be interpreted as illustrating a young person's struggle with the three different identity technologies of sincerity, authenticity, and prolificity. Based on the experiences of the real-world Doody, the film incorporates some of Doody's poems into the script. Here, Doody expresses her search for personal identity in terms of authentic vocabulary. She says, for instance: "I hope I can hold myself and live with myself more ... I really don't want to be anyone else except

myself.” Statements such as these can easily be understood as a protest against her sincerity-based social environment, which tries to impose traditional feminine gender and family roles on her and severely restricts her self-expression. And yet, the curation of Doody’s avatar is clearly not an authentic endeavour but a highly “artificial” social media construction in the context of social validation feedback loops. Experimenting with a language of proflicity, Doody embraces the virtual world and wants to recreate herself in her avatar to “be lovely and beautiful and free as an image.” “If you want to be free, you just need to raise your phone,” the virtual Doody proclaims.

Doody’s post-authentic freedom does not consist in “raising her voice” in an “intersubjective” context, but in her freedom to curate her appearance in a virtual space constituted by social media—and in the freedom to be truly invested in it. Only in the virtual world, her avatar paradoxically assures her, can Doody “show everyone the real you, show them the real Doody.” Tragically, however, Doody’s desire to be “free as an image” in the virtual reality of her avatar is countered by the moral expectations of both sincerity and authenticity that she is confronted with. In one of her poems, Doody accuses herself of “absolute narcissism” and says: “I’ve begun to see the beauty in sinfulness. The hidden beauty of sins, the scattered beauty, the blooming beauty all around me. It breaks into your heart and revives you”. Doody is ultimately denied true investment in her beautiful and free avatar because it violates, on the one hand, the authentic commandment to be “true to oneself” and appears as the sheer narcissism of being unduly enthralled by one’s self-image. On the other hand, Doody’s avatar is sinful from the perspective of the traditional Islamic society she inhabits. Indulgence in bodily beauty appears from this conservative viewpoint as vanity, and the open display of sensuality, especially in

the form of sexualized femininity, is regarded as evil. While 19th and 20th century novels can often be read as illustrating the struggle for authenticity in a society still overshadowed by a regime of sincerity, Anhar Salem's *Love and Revenge* documents an even more complex struggle for true investment in a profilic self in a society torn between a stubbornly persistent sincerity, a still dominant semantics of authenticity, and an emerging proficity that has already become prevalent on global social media.

Conclusion

The main thesis of this essay is that semantics of authenticity is ill suited to describe and understand proficity phenomena. From the perspective of authenticity, the practice of proficity, for instance on social media, may either appear as inauthentic or as a “weird” and somewhat suspicious new form of authenticity. Instead of judging proficity in terms of authenticity, it is argued, we should first and foremost accept it as non-authentic rather than inauthentic. Secondly, we should develop a new post-authentic conceptual framework to better understand how honesty is reconstituted under conditions of proficity. In sincerity, honesty was tied to the commitment to roles, and in authenticity, it was connected with the pursuit of originality. In proficity, honesty is reoriented away from the inner self and toward one's profiles. Here, it consists in true investment in one's profiles. Such true investment develops during the curation process of profiles and typically involves social validation feedback loops.

The argument for the possibility of a post-authentic form of honesty, especially on social media, should not be misunderstood, however, as saying that this more recent kind of honesty is somehow better, or worse, than authentic or sincere honesty. Neither of the three identity technologies sincerity, authenticity, and proficity gets identity right. They all allow for the construction of selfhood narratives in specific social contexts. In modern society, sincerity may eventually become less convincing than authenticity, but this does not make it right as such. And today, proficity-based selfhood may be increasingly practicable, but that does not make it objectively better than authenticity. Each identity technology creates specific conditions for being honest or dishonest. Dishonesty can be insincere, inauthentic, or it can be fake proficity. None of these ways of being dishonest, however, is intrinsically more or less evil than the others. Importantly, just as different identity technologies can be applied by an individual on different occasions, or perhaps even simultaneously, different forms of honesty and dishonesty can and do coexist. Even in times of social media, people can still be honorable and authentic, but the opportunities for being truly invested—or not—are on the rise.

This paper was realized with the support of a Multi-Year Research Grant on Identity and Gender: From Role to Profile by the University of Macau.

References

Achieving authenticity in social media marketing. (2023, May 31). *Aspiration Marketing*. <https://blog.aspiration.marketing/en/bringing-authenticity-back-into-social-media-marketing>

Anton, J. (2021, December 14). Das sind die wertvollsten deutschen Influencer: Ganz vorne nur Frauen. *FAZ.NET*. <https://www.faz.net/aktuell/stil/trends-nischen/das-sind-die-wertvollsten-deutschen-influencer-ganz-vorne-nur-frauen-17684153.html>

Authenticity in social media: Being authentically YOU • Yakkety Yak. (2019, March 5). *Yakkety Yak*. <https://www.yakketyyak.com/authenticity-in-social-media/>

batten.admin. (2021, November 29). Brands im Ringlicht: Deutschlands wertvollste Influencer. Batten and Company. <https://www.batten-company.com/news/brands-im-ringlicht-das-sind-deutschlands-wertvollste-influencerinnenstudie-zum-markenwert-von-deutschlands-erfolgreichsten-influencerinnen/>

Brown, L. (2020, May 18). Social media manufacturing deceptive and inauthentic identities online. *Debating Communities and Networks XI*. <https://networkconference.netstudies.org/2020Curtin/2020/05/10/social-media-has-helped-in-manufacturing-deceptive-and-inauthentic-identities-online/>

DeAndrea, D. C., & Walther, J. B. (2011). Attributions for inconsistencies between online and offline self-presentations. *Communication Research*, 38(6), 805–825. <https://doi.org/10.1177/0093650210385340>

Debord, G. (1994). *The society of the spectacle*. (D. Nicholson-Smith, Trans.). Zone Books.

Deutsch, B. (2018, April 19). Why everyone and everything on social media is fake. *Entrepreneur*. <https://www.entrepreneur.com/science-technology/why-everyone-and-everything-on-social-media-is-fake/311421>

Elliott, E. (n.d.). Social media creates inauthenticity among teens. *The Mentor*. <https://www.mhsmentor.com/11348/opinions/social-media-creates-inauthenticity-among-teens/>

Fukuyama, F. (2018). *Identity: The demand for dignity and the politics of resentment* (1st ed.). Farrar, Straus and Giroux.

Hegel, G. W. F. (2018). *The phenomenology of spirit* (T. Pinkard & M. Baur, Eds. and Trans.). Georgetown University Press.

Hutzler, A. (2019, January 30). Trump is 'perhaps the most authentic' president: Sims. *Newsweek*. <https://www.newsweek.com/former-white-house-aide-calls-trump-most-authentic-president-1311447>

Klein, N. (2010). *No logo: No space, no choice, no jobs* (10th anniversary ed.). Picador.

Lee, J. A., & Eastin, M. S. (2021). Perceived authenticity of social media influencers: Scale development and validation. *Journal of Research in Interactive Marketing*, 15(4), 822–841. <https://doi.org/10.1108/JRIM-12-2020-0253>

Lopez, G. (2016, December 15). Donald Trump can be weirdly honest about his lying: *The Daily Show* gave a few examples. *Vox*. <https://www.vox.com/policy-and-politics/2016/12/15/13966872/trump-lying-daily-show>

Salem, A. (Director). (2021). *Love and revenge* [Film]. Le Fresnoy Studio National des Arts Contemporains.

Marlow, K. (2021, April 26). Social media and the negative impact it has had on the authentic self. *Debating Communities and Networks XII*. <https://networkconference.netstudies.org/2021/2021/04/26/social-media-and-the-negative-impact-it-has-had-on-the-authentic-self/>

Meidar, G. (n.d.). Social media limits our authenticity. *The Standard*. <https://standard.asl.org/14400/opinions/social-media-limits-our-authenticity/> (Accessed: June 2, 2023).

Moeller, H.-G., & D'Ambrosio, P. (2021). *You and your profile: Identity after authenticity*. Columbia University Press.

Sodha, S. (2021, November 7). Social media fuels narcissists' worst desires, making reasoned debate near impossible. *The Observer. The Guardian*.
<https://www.theguardian.com/commentisfree/2021/nov/07/social-media-fuels-narcissists-worst-desires-making-reasoned-debate-near-impossible>

Taylor, C. (2018). *A secular age* (1st ed.). The Belknap Press of Harvard University Press.

Fill In, Accept, Submit, and Prove that You Are not a Robot: Ubiquity as the Power of the Algorithmic Bureaucracy

Mikhail
Bukhtoyarov
Anna
Bukhtoyarova

'Power' (*Macht*) is the probability that one actor within a social relationship will be in a position to carry out his own will despite resistance, regardless of the basis on which this probability rests.

Max Weber

Introduction: Bureaucratic Machine

Rapid development of digital technology influencing most areas of life has a particularly significant impact on human collective behavior, especially when it involves communication between an individual and

Mikhail Bukhtoyarov, Associate Professor, Institute for the Humanities, Siberian Federal University, Krasnoyarsk ORCID: 0000-0002-0627-3048, email: mikebukhtoyarov@gmail.com

Anna Bukhtoyarova, Independent Researcher, Krasnoyarsk
ORCID: 0000-0003-2445-5002, email: annabukhtoyarova@gmail.com

DOI: <https://doi.org/10.5281/zenodo.14203998>

an organization. Algorithms start playing an increasingly important role especially in those social spheres where human behavior has been regulated by cultural and legal protocols for centuries. Most types of routine interactions involving organizations, authorities and corporate bodies often perform relatively strict algorithmic procedures aimed at functioning with little or no regard to personal characteristics of participating individuals.

Governmental and corporate structures with the primacy of a function over a human connection had been perceived as “machines” long before the actual technology became their essential part. According to Alan Scott, the existence of “machine-like modern organizations and social relations” was recognized as a “machine”, one of the most famous modernity metaphors, at the beginning of the 20th century (Scott, 1997).

Classical meaning of a “bureaucratic machine” was conceptualized by Max Weber (Weber, 1930). He described it as a “living machine” consisting of skilled bureaucracy workers functioning together with a “lifeless machine” comprising the technology itself. Weber saw humanity as “trapped” forever in an “iron cage” striving for efficient means in attaining all objectives with little or no concern for the spirit of humanity” (Hewa and Hetherington, 1995).

Observing the current state of affairs by elaborating on this classical metaphor, we claim that with rapid digital transformation of administrative practices the balance between “living” and “lifeless” (in Weber’s terms) has significantly shifted towards lifeless or humanless bureaucracy changing the nature of interactions within the context. The algorithms behind these interactions between an individual human being and a bureaucratic machine asymmetrically affect the

behavior of the human, imposing machine-like rules and protocols. This type of asymmetry tends to produce an asocial environment that lacks balanced psychological connections and empathy.

The Concept of Algorithmic Bureaucracy

Many of the traditional bureaucratic activities can be seen as essentially similar to algorithms. Rik Peeters and Marc Schuilenburg call digital bureaucratic instruments “the digital offspring of the classic bureaucratic procedure, creating classification through standardized and impersonal processing of data” (Peeters and Schuilenburg, 2018). In this respect algorithms are also very similar to AI-based systems as they normally consist of “step-by-step processes and/or rules processing inputs into outputs” (Jansen and Kuk, 2016). Both discussed tools remind of the classical bureaucratic protocols aimed at functioning highly efficiently but as impersonally as possible. This is arguably one of the reasons why digital algorithms and algorithmic AI have been increasingly employed by bureaucratic bodies. And the wide adoption of AI at the time of writing this paper is limited mostly by its novelty and the relatively high cost of development and support.

At the beginning of the 21st century bureaucracies massively started integrating digital technologies into their workflows, but at that time digital tools “did not yet influence the decision-making apparatuses of these organizations” (Bullock, Huang and Kim, 2022). Since then, for a relatively short period of time bureaucratic bodies have adopted a number of digital tools. And as a result, as Mark Bovens

and Stavros Zouridis state, street-level bureaucracies have moved towards screen-level bureaucracies, and, in some cases, reached the system level bureaucracies (Bovens and Zouridis, 2002).

Twenty years later digital tools based on algorithms have become an essential part of bureaucratic workflows with very few or no ways to access the governmental and corporate bodies in any other manner. Algorithm based decision making is often employed at the entry point of accessing various services provided by authorities and corporations. So the first bureaucrat that the service user “meets” during the procedure is a preset algorithm that analyzes the data in order to meet a claim made by a human or to dismiss it. And in many cases that is the only bureaucratic actor involved in the process. Justin Bullock and Kyoung-Cheol Kim call this actor an “artificial bureaucrat” and argue that modern bureaucracy is heading towards the time when “both human and artificial bureaucrats exist and play important roles” (Bullock and Kim, 2020).

Since the digital algorithms were included in the bureaucratic practices, several new terms appeared in attempts to conceptualize their nature and role in human society.

Rapid emergence of various two- and three-word terms containing “digital”, “information”, “e-” as a part of collocation has recognized a new quality of traditional society elements. *Digital Government*, *e-Government*, *Digital Bureaucracy* and other similar terms are aimed at revisiting well-known concepts by adding this overall beneficial quality. For example, “digital bureaucracy” is introduced as “improvement of good governance through the work of the electronic management system to enhance the four basic good governance indicators” (Abdou, 2021). Many other researchers operating these and

similar terms implement them to describe the tools for performing the required functions more neutrally.

John Danaher introduced the term *algocracy* understood as “a system in which algorithms structure and constrain the ways in which humans act” (Danaher, 2016). He focuses on the threats that such a system can introduce to human society exploring the case when algorithmic systems can be entirely automated. Though the term itself has not become commonly used after its launch it still seems useful for reflecting the consequences of the algorithm’s employment in bureaucratic systems.

Algorithmic bureaucracy as a term focuses on the particular method of performing the bureaucratic procedures rather than just describing the space where they occur. It provides a framework for understanding the impact of algorithms on the bureaucratic routines and broader context of interaction practices in human society.

The most commonly used definition of *algorithmic bureaucracy* was introduced by Thomas M. Vogl and his co-authors. According to them, *algorithmic bureaucracy* represents “a new combined organizational environment in which office workers and algorithms can work together” dealing with a higher level of complexity combining new sources of data, data analysis techniques and tools for data analysis (Vogl, 2019). In 2020 they elaborated on the term identifying five major criteria of *algorithmic bureaucracy*, including levels and ways of collaboration, procedural context, collective intelligence, feedback during decision making, and quality of outcomes. They argue that algorithmic bureaucracy is not a replacement for traditional public administration, but “a transformation of the socio-technical relationship

between workers and their tools, as well as the way that work is organized” (Vogl, 2020).

Vogel’s definition and description represent the view on algorithms as one more digital tool employed in bureaucratic practices. However, there are other perspectives considering algorithms as a more complex phenomenon that would significantly change the bureaucratic practices and as a result the interaction patterns in human society.

Mariavittoria Catanzariti emphasizes that the real challenge of employing algorithms is their extraterritoriality as “algorithmic systems are ubiquitous, along with data that feed machine learning techniques” (Catanzariti, 2021). Traditionally a bureaucratic state is aimed at organizing and managing the distribution of power over and within a particular territory. Data flows processed by algorithms are not limited by any territory signifying that virtual reality is a radically different form territorial sovereignty. Jennifer Daskal argues that “data challenges territoriality doctrine at its core” (Daskal, 2015) as it is borderless and can be scattered everywhere across different countries. Implementing algorithmic bureaucratic tools that are not connected directly to a local municipality can on one hand help to avoid corruption on a lower level but on the other hand it erodes the sense of belonging to a particular neighborhood. Even with all the disadvantages of traditional bureaucracy on the street level it has been working for centuries as one of the building blocks for shaping and maintaining a local community.

One more factor that makes algorithms different from other digital tools for bureaucratic practices is the lack of transparency about the way they work for the public to understand. In this case it does not imply hiding some aspects of bureaucratic activities but rather

impossibility in principle to comprehend the ways high-level algorithms and AI function without advanced competency in this area. The black box problem that has been discussed for years is applied for algorithms in general and especially to AI based ones in particular. And while implementation of such algorithms by bureaucratic bodies could make decisions more efficient, “also they could disguise information inside “black boxes”, preventing citizens from the knowledge on the implications they might have in their own lives” (Fink, 2017). While traditional bureaucracies used to maintain at least an image for decision-making transparency, algorithmic bureaucracies tend to keep most of their internal procedures closed causing confusion and alienation of the service users as most of them are unaware of the ways algorithms work.

Transparency in the case of algorithms can be achieved if external auditors or agencies could “access algorithms assessing their compliance with ethical, legal, and governance standards, regulations and procedures” (Ignacio, 2020). However, the current state of affairs is far from this ideal as bureaucratic bodies are not open to such audits due to various reasons such as security protection, avoiding public criticism and more. Moreover, such audits would require designing tools and procedures capable of coping with the high level of complexity. It also means the application of digital technologies, that could probably lead to the infinite loop of black boxes where the public will still remain unaware or misinformed.

Nevertheless, utilization of algorithms is rapidly becoming a *new normal* in public and corporate administration similarly to commerce, trade and other economic activities. Recognising the undeniable benefits of algorithmic bureaucracy, we claim that it is necessary to

critically approach the limitations and threats of the ubiquitous algorithms implementation.

To summarize, there is an ongoing change of perspective from seeing algorithms and algorithm-based AI as one more tool for bureaucratic practice optimization to viewing it as a vehicle for deep transformation in the public administration and corporate management. There are claims that AI and algorithms are currently reshaping the global order (Wright, 2018) or changing the way of interaction among countries (Klingler-Vidra, 2019). But within the scope of the paper we suggest exploring the impact of *algorithmic bureaucracy* on a user as an individual whose behavior is influenced by performing the numerous routine procedures.

Training an Algorithmic User

- 1 | *Accept cookies, agree with the terms and conditions,*
- 2 | *Register or sign in,*
- 3 | *Fill in the form (Do not forget about the required fields!),*
- 4 | *Submit.*

This pattern with little variations can be found on user interfaces of multiple governmental and municipal services online. Individuals who become the users of such services routinely perform step-by-step operations often without being fully aware of conducted legal actions, and/or without knowing how exactly and where the collected

data can be utilized. Besides their request of a particular bureaucratic service these individuals give their consent on collecting and processing their personal data, analyzing user behavior, being included into multiple databases and transferring other rights. In order to get access to the required service or receive necessary information users become included into the complex data flow that generates not only the data, but also the legal consent. This repeated pattern forms an unconscious habit with everyday behavior automation. That raises the question: do such services convert individuals who follow the imposed algorithms into *algorithmic users*?

As *algorithmic bureaucracy* is increasingly discussed across various domains of social sciences and humanities we believe there is a related concept of *algorithmic user* that needs to be introduced and explored. As the term is currently not present in the literature, in order to formulate the criteria for introducing it we suggest discussing three questions.

Our first question is the asymmetry in the locality of the algorithmic user and ubiquity of the algorithmic bureaucracy. It is highly unlikely that the algorithmic bureaucracy is capable of remaining 'local' in a world where the data markets are dominated by SaaS and PaaS technical solutions. Such technologies are mostly developed by national and transnational IT companies. Application of the global digital tools to the municipal data management produces an asymmetric flow of data. That enables global and nationwide stakeholders to become more capable of imposing standards and managing regulations for the local level. Moreover, it creates the unavoidable presence of such stakeholders in every local transaction and limits the individual in her/his ability to find local, autonomous solutions for their

small-scale, territory-related problems. This ubiquity makes a user ready to perform algorithmic actions required by a service with no regard to a particular territory and/or a time period.

The second question is whether the *algorithmic* way of organizing services with a step-by-step sequence of required actions is influencing the way individuals and groups now expect to have services provided. In other words, whether massive implementation of algorithms is constantly training users to follow particular patterns and to pursue them as a new standard for bureaucratic procedures and likely in a broader context. For defining this type of behavior, the attribute *algorithmic* is more important than, for example, *digital* because it emphasizes the sequential type of actions, and not the environment where they occur.

The third question and the biggest concern is whether the power of algorithms influences the routine behavior of a user/ citizen/ customer (Using, 2017) making him/her perform actions not just without fully understanding of the consequences but without even noticing. In our opinion this mechanism of involvement is in a way similar to microtransactions extensively implemented in the video game industry, where a user performs small impulse purchases inside games (Caetano, 2017). But in the discussed environment a user pays by providing his/her data and transferring the rights to bureaucratic bodies. The interface of e-government platforms is designed of the same elements as the ones of online stores, online media, network games and other services. Ubiquity and habituality of online forms masks legal actions as conventional, insignificant, and even annoying. Routinely pushing the “Agree” button, an *algorithmic user* develops a habit to give consent without legal advice or thorough analysis of the consequences

and to follow a path (called “user path” or “user flow” in IT) embedded into the interface of a platform.

This discussion results in suggesting the following definition:

An algorithmic user is an individual routinely and habitually performing required sequences of actions to comply with the rules and regulations imposed by ubiquitous online services.

While an *algorithmic user* conveniently matches the behavioral expectations of bureaucratic bodies it raises an important question: How is this type of behavior changing the way of interaction within the bureaucratic context?

Describing the user behavior types in terms of their relationship to bureaucratic procedures we suggest applying the classification proposed by John Clayton Thomas who specifies three roles of *public* or, more specifically, *members of the public* in *public management*: citizen, customer, and partner (Thomas, 2017). For the purpose of this paper the term “user” can be considered a synonym to “*member of the public*” to represent an individual, who utilizes the bureaucratic services. The three roles and the interaction models introduced by Thomas can be applicable for analysis of the technologically mediated bureaucracy-user interaction. Observing the impact on user behavior made by the spread of algorithmic bureaucracy all over the world it is worth exploring the changes resulting from employing algorithms for each of the discussed interaction models.

The most common role, according to Thomas, is a *customer* when users “seek a discrete product or service, usually for its personal value to them” (Thomas, 2017). The *customer* role/model is based on

“individualism and the pursuit of self-interest” (Roberts, 2004). This is a classical approach when the administrative procedures are provided as goods or services, sometimes for a fee, similar to the ones from private businesses. Most cases of the individual-to-bureaucratic body interaction fit this model as *customers* come for the particular services for themselves or their family members. The proportions of residents who perform customer-like behavior by submitting “a request for service or a complaint” are as high as 60 to 70 percent or more (Hirlinger, 1992).

Even without introducing any digital tools, processing several types of official claims submitted by *customers* in written form is clearly algorithmic to a great degree. These types of routine requests and complaints involve step-by-step sequences that can be divided into micro actions, presented as workflow diagrams and transformed into algorithms quite easily. The procedures normally start with filling up official forms, continue with paying standard fees and submitting the documents into a particular administrative window and conclude with waiting for a specified period of time to receive a standard document as an outcome. With respect to possible variations this series of steps resembles a classical example of non-digital algorithmic sequence of actions. That is why the transformation of *customer* procedures into the digital algorithm format is the logical first step for the governmental and corporate bureaucracy digitalization.

An *algorithmic user* with his/her habit of performing sequential actions routinely matches the *customer* model of interaction in bureaucratic context perfectly. Having extensive experience in following similar user paths on multiple commercial platforms and applications, an *algorithmic user* expects smooth and predictable transition from one step to another while dealing with governmental and

corporate bureaucratic procedures online. In many cases completing *customer* bureaucratic procedures without involving humans can be very convenient as it saves time and effort if a *customer* is competent in using algorithm-based services. We argue that future algorithm advancement and wider artificial intelligence employment for the *customer* bureaucratic procedures will be growing as this model meets the interests of both users and bureaucratic bodies. It will increasingly produce *algorithmic users* ready to perform their part in an algorithm without interacting with human bureaucrats. With respect to all the benefits of the *customer* model we argue it to be the most dangerous model in terms of its consequences for an individual and human society. An *algorithmic user* proceeding along the algorithmic path from start to finish like a lab rat inside the Skinner box has limited choices and is forced to follow the pattern without questioning it.

Another role model according to J.C. Thomas's classification is a *citizen* that defines the situation when "public managers invite citizens to contribute their ideas on an issue, and share at least some decision-making authority with them" (Thomas, 2017). The proportion of administrative procedures where this role model is applicable is significantly lower in the bureaucratic workflow compared to the ones where a user acts as a *customer*. The *citizen* model is currently applicable almost exclusively for collective actions related to public interest unlike the *customer* model aimed at a private interest.

Public involvement of the users based on the *citizen* model usually intends to legitimize making debatable decisions via open discussions and/or voting for a particular choice. Bureaucratic procedures on group decision making are quite complex to organize and potentially problematic. They usually cannot be presented as step-by-step sequences with predictable outcomes. That is especially the case if

a decision-making process involves open discussions as algorithms are currently not the most efficient way to manage them. Therefore organizing such discussion-based activities requires constant moderation by humans. Though there is a potential for advanced AI technology to replace human moderators in future, currently this type of interaction based on the *citizen* model continues to be complicated and expensive to design and manage.

Algorithmization within the *citizen* model is applicable mostly for the final step of public decision making that is usually selection from the suggested list of several options. Related bureaucratic procedures ensure public involvement with fairly simple actions such as answering questions or choosing variants in an interactive form. An algorithmic *citizen* is significantly less common in the bureaucratic context than an algorithmic *customer*. But as public decision making is often quite important for all the stakeholders the overall expectations for bureaucratic procedures are typically higher.

As their main purpose is to ensure legitimization of a particular publicly chosen variant, it is crucial that they are transparent and trustworthy. The potential situation when algorithms are somehow altering the results of public decision making or manipulating the public opinion in other unethical ways makes the entire model of interaction defective or obsolete. To be seen as an effective tool, a bureaucratic procedure based on the *citizen* model needs to be audited by independent actors in order to prove that an instrument for legitimization is legitimate.

The implementation of digital algorithms for the *citizen* model interaction arguably has a strong potential to grow with the spread of direct democracy involving citizens in regular public decision making. In

this case an *algorithmic user* will form a habit of making choices for public decision making with digital tools. The *citizen* model is basically a combination of a non-algorithmic discussion part and an algorithmic voting part that is ideally based on the first one and requires close public control. As we believe, employing this model involves cognitive presence and informative decision making that could compensate for the pressing behavioral patterns of algorithms.

The third role introduced by J.C. Thomas is a *partner* based on the idea of “co-production” for specific services where “assistance is essential or would be helpful from the public” (Thomas, 2017). The *partner* role means that a user is not just following the procedures but is participating in developing them. That quite uncommon role in bureaucratic context would change the dynamics of individual-to-bureaucracy interaction significantly. Instead of an inferior-to-superior traditional way the relationship would become equal-to-equal for particular situations. The *partner* role is only emerging and is still *terra incognita* in terms of specific bureaucratic procedures necessary to manage the *partner* interaction model effectively.

It is currently unclear what is the potential for user behavior algorithmization within the *partner* model as there are only experimental attempts to introduce administrative practices based on this model. However, as John Alford states, in order to make *partner* role/model work, tasks for the public need “to be simplified as much as possible” as “the easier the task, the higher the likelihood citizens will be able to assist” (Alford, 2009). Such decomposition of complex tasks would probably give opportunities for algorithmization of at least some elements of activities based on a *partner* model. However, we argue that those algorithms would require a high level of flexibility and an *algorithmic user* would have to be competent and creative enough

to participate in co-production. The *partner* model will require active cognitive and emotional involvement and proactive creative approach to procedure performance. Though it is currently unlikely that the *partner* model will become a dominant or even a commonly implemented one it has the strongest potential to overcome the disadvantages of bureaucratic procedure algorithmization or at least to compensate for them.

To summarize the discussion all three user roles/models can potentially employ algorithms for changing interaction within the bureaucratic context. But the impact on the user behavior would arguably be quite different, namely, passive following a behavioral pattern for *customer* model, cognitive involvement in public decision making for *citizen* model and proactive creativity for *partner* model.

Though people currently have limited knowledge about long-term consequences of algorithmic bureaucracy spread, we suggest exploring the changed role of empathy within this context as one of the important factors potentially changing human society.

Administrative Empathy vs Bureaucratic Algorithms

Introducing the ideal state of bureaucracy, Weber quoted Goethe to describe a structure that implies “specialists without spirit, sensualists without heart; this nullity imagines that it has attained a level of civilization never before achieved” (Weber; 1930). This state is often referred to as a *bureaucratic machine* perceived as “a complex decision-making tool in public organizations, which seeks to control and

protect the public interest by providing little room for personal discretion” (Farazmand; 2020). The bureaucratic machine metaphor has been one of the most recognized public images of bureaucracy for centuries but only with the introduction of algorithms is it getting to an extreme point where human specialists are being replaced by genuinely *heartless* and *spiritless* algorithmic tools that represent pure functions.

As J. Ignacio Criado and his coauthors stated, algorithms “have the potential to transform critical dimensions of public sector organizations and people working in them” (Criado, 2020). We argue that one of the most significant differences of algorithm-mediated interaction is the lack of empathy involved in the bureaucratic process.

Before introducing digital tools even with a significant level of administrative procedure algorithmization, none of the bureaucratic systems could be truly “lifeless” or humanless because bureaucratic processes almost always implied some amount of interaction with human employees, especially at the local level. On the one hand it made the system vulnerable to possible corruption and human errors, but on the other hand the bureaucratic service could be “customized” with small negotiations and alterations on demand. As a bureaucratic procedure is becoming increasingly automatic with the use of algorithms, sooner or later a user will have to deal with an algorithm only operating inside the machine without any human feelings and any empathy.

The ongoing development of algorithmic bureaucracy is changing the balance of *administrative empathy* defined by Sofia Ranchordás as acknowledgement “that citizens have different skills and needs, and this requires the redesign of pre-filled application forms, government platforms, algorithms, as well as assistance” (Ranchordás, 2022). One of the most important features of *administrative empathy*

is the recognition that “mistakes made in good faith can be forgiven under limited circumstances, and vulnerable individuals are given second chances to exercise their rights” (Ranchordás, 2022). With algorithms performing decision-making automatically, *administrative empathy* is becoming rare.

From one hand elimination of most humans in the bureaucratic process could significantly decrease “personal subjugation, nepotism, cruelty, emotional vicissitudes, and capricious judgment” (Bennis, 2017), but on the other hand shifting balance of *administrative empathy* with automation of decision-making in governments and corporations could cause an individual to feel isolated, misunderstood and sometimes helpless. Being a unique human feature, administrative empathy provides some empowerment to an administrative service user making him/her believe that an individual has some control over decision making through possible communication with another person. In the case of an ultimate algorithmic procedure the process of individual-to-bureaucracy interaction is a channeled sequence of entering data and giving consents without any chance to emotionally influence the process. Humans inside this system are rejected as non-rational beings. Ironically, asking an algorithmic user to prove that he/she is not a robot algorithmic bureaucracy in a sense makes a human act robotically.

Dealing with new interfaces of the bureaucratic services requires new skills from individuals. Routinely repeating a sequence of operations while accessing a service an individual becomes a skilled *algorithmic user*. Similarly to Edtech applications implementing algorithms, governmental and municipal web services perform an educational function of training individuals to follow and memorize particular patterns. However unlike most EdTech and commercial systems

bureaucratic tools are not adaptive, as their goal is not to develop individuals. They simply train humans in the most ultimate behaviorist manner. Such systems place their users into the environment with strict 'playing by the rules' relations based on rational premises. There is no right to make a mistake and no room for human emotions or individual characteristics.

To summarize an algorithmic service optimizes human behavior in order to adjust to a rational machine algorithm. The ongoing disappearance of administrative empathy from bureaucratic procedures can potentially have a significant impact on how such procedures are perceived and accepted.

Discussion: Ways Out of the Iron Cage

In order to overcome the negative consequences of the increasing algorithmization of bureaucracy we suggest several potential scenarios for solutions. Some of them are already being developed and included into the bureaucratic services.

The first and the most common solution is feedback tools and research of customer satisfaction. There are multiple research papers published on this topic (Kim, 2005; Ali, 2017; Li and Shang, 2020). The limitation of this approach is its recursive nature: the users answer survey questions, fill in more forms, submit answers, and give their consent. This type of feedback does not involve empathy and constructs one more layer of algorithmic procedures.

The second scenario could be an independent community-driven platform for reviewing and evaluating bureaucratic platforms and procedures. This solution requires public awareness and understanding the issues and limitations of the algorithmic bureaucracy and its services. Currently civil society organizations are developing initiatives on monitoring and evaluating public services, but the ubiquity of algorithmic bureaucracy makes these voices unheard.

The third scenario would be the diversification of models for interaction within bureaucratic context by developing activities based on *citizen* and especially *partner* models in order to compensate for the disadvantages of the *customer* model.

In our opinion the most radical scenario could be the right of an individual to construct and redesign bureaucratic procedures, adapting them to the individual needs and characteristics, involving both digital and non-digital interactions following the principles of *partner* interaction model. This approach could raise the user-service relations to a different level of involvement. The most significant limitation is that development of such highly adaptive, user-centered services requires highly sophisticated and resource consuming solutions.

The big question of the algorithmic bureaucracy is whether ‘user friendly’ or even empathic bureaucratic procedures can be feasible in the sociotechnical system that seeks optimisation and rationality.

This paper was conducted with the support of the Southeast European Solidarity Fellowships for Threatened Scholars and Digital Society Lab, Institute for Philosophy and Social Theory, University of Belgrade.

References

- Abdou, A. M. (2021). Good governance and COVID-19: The digital bureaucracy to respond to the pandemic (Singapore as a model). *Journal of Public Affairs*, 21(4), e2656. <https://doi.org/10.1002/pa.2656>
- Alford, J. (2009). *Engaging public sector clients: From service-delivery to co-production*. Springer. <https://doi.org/10.1057/9780230235816>
- Ali, M., Asmi, F., Rahman, M. M., Malik, N., & Ahmad, M. S. (2017). Evaluation of E-service quality through customer satisfaction (A case study of FBR E-taxation). *Open Journal of Social Sciences*, 5(9), 175–195. <https://doi.org/10.4236/jss.2017.59013>
- Bennis, W. G. (2017). *Beyond bureaucracy*. In *American bureaucracy* (Ed.) (pp. 3–16). Routledge. <https://doi.org/10.4324/9781315082707>
- Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, 62(2), 174–184. <https://doi.org/10.1111/0033-3352.00168>
- Bullock, J. B., & Kim, K. C. (2020). Creation of artificial bureaucrats. In *Proceedings of European Conference on the Impact of Artificial Intelligence and Robotics*. <http://dx.doi.org/10.34190/EAIR.20.001>
- Bullock, J. B., Huang, H., & Kim, K. C. (2022). Machine intelligence, bureaucracy, and human control. *Perspectives on Public Management and Governance*, 5(2), 187–196. <http://dx.doi.org/10.1093/ppmgov/gvac006>
- Caetano, R. G. F. (2017). *Main drivers for microtransactions as impulse purchases in e-commerce* [Master's thesis, ISCTE – Instituto Universitário de Lisboa]. <http://hdl.handle.net/10071/15309>
- Catanzariti, M. (2021). Algorithmic law: Law production by data or data production by law? In H.-W. Micklitz, O. Pollicino, A. Reichman, A. Simoncini, G. Sartor, & G. De Gregorio (Eds.), *Constitutional challenges in the algorithmic society* (pp. 78–92). Cambridge University Press. <https://doi.org/10.1017/9781108914857>

Criado, J. I., Valero, J., & Villodre, J. (2020). Algorithmic transparency and bureaucratic discretion: The case of SALER early warning system. *Information Polity*, 25(4), 449–470. <http://dx.doi.org/10.3233/IP-200260>

Danaher, J. (2016). The threat of algocracy: Reality, resistance, and accommodation. *Philosophy and Technology*, 29(3), 245–268. <https://doi.org/10.1007/s13347-015-0211-1>

Daskal, J. (2015). The un-territoriality of data. *The Yale Law Journal*, 125(2), 326–559.

Farazmand, A. (Ed.). (2023). *Global encyclopedia of public administration, public policy, and governance*. Springer Nature. <https://doi.org/10.1007/978-3-030-66252-3>

Fink, K. (2018). Opening the government's black boxes: Freedom of information and algorithmic accountability. *Information, Communication & Society*, 21(10), 1453–1471. <http://dx.doi.org/10.1080/1369118X.2017.1330418>

Hewa, S., & Hetherington, R. W. (1995). Specialists without spirit: Limitations of the mechanistic biomedical model. *Theoretical Medicine*, 16(2), 129–139. <https://doi.org/10.1007/BF00998540>

Hirlinger, M. W. (1992). Citizen-initiated contacting of local government officials: A multivariate explanation. *The Journal of Politics*, 54(2), 553–564.

Ignacio, C., Valero, J., & Villodre, J. (2020). Algorithmic transparency and bureaucratic discretion: The case of SALER early warning system. *Information Polity*, 25(4), 449–470. <https://doi.org/10.3233/IP-20023>

Kim, T. H., Im, K. H., & Park, S. C. (2005, August). Intelligent measuring and improving model for customer satisfaction level in e-government. In *International Conference on Electronic Government* (pp. 38–48). Springer Berlin Heidelberg. https://doi.org/10.1007/11545156_6

Klingler-Vidra, R. (2019). AI super-powers: China, Silicon Valley, and the new world order. *International Affairs*, 95(2), 485–486. <https://doi.org/10.1093/ia/iiz004>

Li, Y., & Shang, H. (2020). Service quality, perceived value, and citizens' continuous-use intention regarding e-government: Empirical evidence from China. *Information & Management*, 57(3), 103197. <https://doi.org/10.1016/j.im.2019.103197>

Peeters, R., & Schuilenburg, M. (2018). Machine justice: Governing security through the bureaucracy of algorithms. *Information Polity*, 23(3), 267–280. <https://doi.org/10.3233/IP-180090>

Ranchordás, S. (2022). Empathy in the digital administrative state. *Duke Law Journal*, 71(6), 1341–1389. <https://scholarship.law.duke.edu/dlj/vol71/iss6/4>

Roberts, N. (2004). Public deliberation in an age of direct citizen participation. *The American Review of Public Administration*, 34(4), 315–353. <https://doi.org/10.1177/0275074004269288>

Scott, A. (1997). Modernity's machine metaphor. *The British Journal of Sociology*, 48(4), 561–575. <https://doi.org/10.2307/591600>

Thomas, J. C. (2017). *Citizen, customer, partner: Engaging the public in public management*. Routledge.

Using digital methods to improve citizen engagement in local government. (2017). *Roundtable Digital Citizen Engagement Discussion: Creating digital impact*. https://novoville.com/wp-content/uploads/2018/12/Roundtable_Digital_Citizen_Engagement-Discussion_Notes_June2017.pdf

Vogl, T., Seidelin, C., Ganesh, B., & Bright, J. (2019, June). Algorithmic bureaucracy. In *Proceedings of the 20th Annual International Conference on Digital Government Research* (pp. 148–153). <https://doi.org/10.1145/3325112.3325244>

Vogl, T. M., Seidelin, C., Ganesh, B., & Bright, J. (2020). Smart technology and the emergence of algorithmic bureaucracy: Artificial intelligence in UK local authorities. *Public Administration Review*, 80(6), 946–961. <https://doi.org/10.1111/puar.13282>

Weber, M. (2001). *The Protestant ethic and the spirit of capitalism* (T. Parsons, Trans.). Routledge. (Original work published 1930).

Weber, M. (1947). *The theory of social and economic organization*. Oxford University Press.

Wright, N. (2018, July 10). How artificial intelligence will reshape the global order. *Foreign Affairs*. <https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-will-reshape-global-order>

A Balanced Approach to Artificial Intelligence in Government Decision-Making Processes

Aleksandar Jazić
Ljubiša Bojić

Introduction

The essential pillars of democratic governance in a successful and functional system at all levels unfailingly include the participation of citizens in the local decision-making processes. The health of any democratic governance structure is often reflected through the level of transparency exhibited by the government, the willingness of the government to actively involve the citizenry in the decision-making procedures, and the facilitation of public voices being heard (Warren, 1992).

These parameters of democratic health, while being fundamental, also represent persistent challenges. For instance, the active

Aleksandar Jazić, Senior Research Fellow, Institute of International Politics and Economics, Belgrade ORCID: 0000-0002-5613-2614, email: jazic@diplomacy.bg.ac.rs
Ljubiša Bojić, Senior Research Fellow, Institute for Philosophy and Social Theory, University of Belgrade, Belgrade; Institute for Artificial Intelligence Research and Development of Serbia, Novi Sad ORCID: 0000-0002-5371-7975, email: ljubisa.bojic@ifdt.bg.ac.rs

DOI: <https://doi.org/10.5281/zenodo.14204046>

involvement of citizens in public policy discussions and decision making, while a cherished democratic ideal, is fraught with logistical and financial implications. Participatory mechanisms, such as referendums, town hall meetings, public consultations, and alike, although might seem the epitome of democratic engagement, are weighed down by significant constraints. These constraints range from the financial and logistical costs to broader issues like voter turnout and the quality of citizen participation.

Facilitating citizen participation in decision-making processes, then, is akin to walking a tightrope. Failure to strike a balance can distort the democratic ethos, skewing decision-making processes, fostering public disillusionment, and ultimately leading to a waste of potentially beneficial fiscal, human and temporal resources (Smith, 2018).

Compared to this predicament faced by governments globally, the advent of Artificial Intelligence (AI) in the governance process provides a promising ray of hope. Seen from a state's perspective, the strategic use of AI portends a revolutionary shift towards enhanced operational efficiency across governmental agencies. With innovative data analysis capabilities requisite for various complexity levels, AI can augment the creation, passage and implementation of government policies, thereby surpassing the capabilities of traditional systems (Bol et al., 2020). The potential of AI extends to digitizing government tasks, which allows for streamlining of mundane tasks, enabling public workers to focus their energies on critical tasks that require human involvement, intervention, deliberation, and analysis.

Enhancing citizen experiences, AI-infused systems can contribute to real-time communication between the government and its citizens, facilitating quicker and more efficient functioning of government

entities. Moreover, AI's ability to gather, manage, and analyze large data volumes can allow governments to analyze potential scenarios related to decision-making processes and resource allocation, thereby replacing many policy development assumptions with actual data-backed insights (Desouza, 2018).

Contrarily, scholars expressing caution towards this rapidly expanding sphere of AI in governance underline several concerns. They suggest that an unbridled expansion of AI incorporation in governance and decision-making might inadvertently risk redundancy of certain job roles, potentially leading to a future society marred by job insecurities and alienation (Brynjolfsson and McAfee, 2014; Arntz, Gregory and Zierahn, 2016). Hence, this relentless march of technology, if not managed responsibly, might render subjective human judgment irrelevant, the consequences of which are yet unclear but potentially profound. Furthermore, AI's potential innate biases could undermine the principles of transparency and fairness, while careless mishandling of citizen data could open the doorway to misuse of private and sensitive information, compromising privacy rights and raising ethical concerns (Zarsky, 2016).

Steering between these opposite poles of optimism and apprehension requires a balanced and pragmatic approach. The application of AI in government functioning should be seen as a citizen-centric democratic tool rather than a purely technical, problem-solving instrument, thereby requiring the public and government officials to work in harmony in defining and harnessing AI's usage. It is proposed that the introduction of AI should align with, and serve to accentuate, the existing governance systems rather than disrupt and replace them. In this environment, the transparency of governments in handling

citizen data and establishing trust is paramount. A diligently considered approach is thereby needed, which also underlines the potential risks of machine learning biases and human redundancies. The society should perceive AI as an aid, augmenting the existing decision-making processes rather than a standalone entity usurping the human intervention from crucial civic decisions (Merh, 2017).

Taking these arguments into account, our objective herein is to navigate through the intriguing landscape of AI in government decision-making spheres. We aim to strike a balance between its transformative potential and the necessity of preserving human agency in crucial decisions, thereby catalyzing a context-sensitive, efficient, and citizen-centric AI governance model that effectively serves our democratic institutions.

Artificial Intelligence and Public Services

Over the course of time, there has been a global shift among governments towards incorporating technological innovations in their operations, spurred predominantly by the evolving needs and demands of their citizens. Prominent among these considerations have been the digitization of public activities and a heightened impetus for citizens' involvement in the cycle of public policy—its development, implementation, and subsequent evaluation. This transformation is inextricably linked with the concept of e-government, which has emerged as a potent solution catering to the aforementioned demands.

More than a mere paradigm shift, e-government signifies a strategic recondite of public governance. It amplifies effectiveness and efficiency in government operations, imbuing the traditional bureaucratic setup with an enhanced focus on processes and procedures leading to the provision of quality public services. Another key aspect intrinsic to e-government is its potential to ameliorate decision-making processes. By leveraging technology, e-government can substantially augment the utilization of available information for informed decision-making while facilitating seamless and efficient communication among various governmental offices. These benefits add up to foster an atmosphere of transparency and efficiency, thereby enhancing public trust in government services. As such, there has been a paradigm shift in how governments approach their operations, with artificial intelligence (AI) being adopted as a significant component of the technological solutions, given its multitude of advantages (Duberry, 2022. 14-15).

Considering public services at the local level, local governments undeniably play an indispensable role. They are tasked with the provision of critical services that citizens rely upon for their everyday administrative needs. The authoritative competence conferred by law and the fiscal capacity of the local self-government dictates the number and quality of these services. These include, but are not limited to, significant domains such as social assistance, health care, tackling unemployment, housing provisions, and education. Ethnographically considered, these services cater to the societal fabric's different strata, each uniquely, pluralistically, and democratically impacting a cross-section of citizens in their respective lives, thereby being inextricably intertwined with citizens' life quality.

The type of public services provided run a complete gamut of categories, including administration, social affairs, education, health, safety, and infrastructure, striving to assuage citizens' diverse and dynamic needs. Considering the pivotal role of local self-government bodies, their primary task remains the provision of these services efficiently and effectively. Decisions within the public sector demand constant vigilance and scrutiny. They need to be justifiable, thereby emphasizing the accountability of public institutions towards the citizens they serve, reflecting responsiveness in local governance (Geske and Leyer, 2022).

Considering the application of AI in public services, several municipalities in Nordic countries serve as ample examples. As of 2019, these municipalities started integrating AI into their digital strategies. The scope and extent of the usage of AI varied among these governments, ranging from attempts at an organization-wide application across different departments to focused attempts at forming AI-based structures grounded on political foundations. While some governments spared no effort in creating conducive conditions for introducing AI, it was observed that, in many cases, AI technologies were not fully utilized, often due to reluctance or hesitation among certain employees or politicians.

Detailed examples of AI's practical usage in these municipal contexts are extensive. Notable among these are systems redirecting calls to designated clerks, AI systems for sorting incoming emails, the utilization of AI chatbots for citizen inquiries, and deploying robots for executing repetitive tasks. Many of these municipalities have leveraged AI chatbots on their official websites that provide automated responses to frequently asked information by citizens, thereby reducing human intervention and increasing efficiency. The City of Copenhagen uses

specific AI algorithms to direct citizens who call the municipal call center to a particular employee. Similarly, Helsinki developed algorithms using text analysis for sorting out citizens' written inquiries, aiming to reduce the manual handling time involved for employees. However, while these instances show the potential advantages of integrating AI in governmental services, they also raise urgent questions on AI's extent of usage, the significant influence it can have on decision-making processes, and the perils of bias or discrimination it may induce in processing citizen inquiries that need conscious, guided discussions and scrutiny (Nordic Municipalities' work with Artificial Intelligence, 2019: 16–17). As such, while AI seems promising, its application requires careful consideration, balancing risks and benefits, and ensuring a citizen-centric approach.

Artificial Intelligence and Its Influence on the Decision-Making Process

Our initial discussion laid out how the facilitation of Artificial Intelligence (AI) in public service domains has considerable potential to transform citizens' lives for the better. It outlined current areas where citizens can exploit AI to navigate through everyday administrative nuisances and to procure essential information with relative ease.

However, the evolution of AI is not without its set of challenges. Significant knowledge boundaries exist, creating roadblocks that restrict establishing stable communication channels within and between different stakeholders involved in AI development. This disparity often

stems from the skewed distribution of authority and power that can shake the foundations of mutual trust and undermine the principles of democratic governance (Warren, 1992).

The magnitude and complexity of such challenges call for a multi-faceted strategy. Incorporating AI in public administration systems warrants prompt and precise policy coordination, which needs to permeate various social levels and administrative sectors. Authorities need to adapt promptly and reconceive their *modus operandi* to establish new procedures that can house the advancements of AI. Efficient governance echoes the productive orchestration of stakeholders and their collaborative efforts, with AI standing to catapult these collective governance procedures to loftier tiers.

We live in an era of unprecedented social and political changes, which propels the call for innovative solutions and forward-thinking strategies. AI carries the potential to usher in significant enhancements to democratic processes, instilling citizens with the required motivation to partake in this wave of change with more involvement and enthusiasm. This citizen-centric approach to engagement is not just beneficial for society but stands as a crucial prerequisite for fostering trust in AI technologies and the institutions effectively managing AI. Governments can play a pivotal role by organizing educational programs and debates centred around the positive aspects of AI, facilitating a better understanding of its features among citizens.

As we delve into the 21st century, our urban landscapes grapple with formidable challenges that have a bearing on their regular functioning. Managing sprawling cities with ever-increasing populations, the continual swelling of resource demand, and, most critically, the widening chasm of socioeconomic differences among citizens,

encapsulate some of the predicaments that today's modern cities encounter. AI comes into the picture as a potential problem-solver, a beacon of hope that could chart the road map for future city development (Yigitcanlar et al., 2021).

However, unchecked or irresponsible innovation has the potential to give rise to unforeseen issues, distracting from ongoing attempts to tackle existing challenges. Responsible urban innovation, hence, becomes pivotal in uncovering and addressing pressing urban challenges dominated by complex problems, uncertainty of solutions, and associated risks. Proposed innovations must incorporate a responsibility clause that places priority on public and environmental values over mere technological advancement. Economic factors are crucial in shaping innovation, but they cannot sideline environmental and social ones in the development and implementation of urban innovations.

Local governments hold an advantageous position in this narrative—to view AI as an instrument that can foster stronger relationships with their citizens and enhance their functioning and efficiency. AI's advantages for local governments span enhancing administrative efficiency, hastening task completion, automation of routine decisions, effective management of repetitive tasks, and the relegation of error margins to a minimum. With these merits, local governmental departments such as infrastructure management, environmental protection, customer services, cybersecurity, and performance analysis could significantly benefit from AI implementation (Sigfrids et al., 2023).

Despite this optimistic view, numerous constraints persist, including inevitable issues of data bias, substantial investment requirements, potential risks of automation, regulatory and ethical concerns, and informational gaps about AI's practical implications. However, local

governments express an eagerness to maximize AI's potential use, notwithstanding their organizational structures that may lack the necessary support for the comprehensive integration of AI (Yigitcanlar et al., 2022).

This situation has given rise to two main approaches: the cautious “wait-and-see” approach, where local governments observe AI with keen interest for the ideal time to jump onto the AI bandwagon, or the rather hands-on, experimental approach, where governments implement AI right away and iron out potential wrinkles during the process.

Yet, both strategies bring to focus the necessity for local authorities to lay out regulatory guidelines, improve AI technology's accessibility, increase public understanding of AI, and encourage public acceptance of AI technologies. They need to define a specific direction for AI integration and adopt best practices of AI management for successful application and overcoming both technical and non-technical issues that may arise along the way. A successful AI application hinges not just on technical know-how, but also on rigorous leadership support. As we continue to navigate the potential of AI in public service and iron out these issues, the full embrace of AI's capabilities in local governance becomes a captivating prospect (Ojha et al., 2021).

**Artificial Intelligence,
Decision-Making Process,
and E-Democracy**

As we delve into the 21st century, we notice that the structure of institutional power is undergoing an overhaul. With economic growth

becoming less uniform and much more fragmented, the conventional role of central government institutions is witnessing a decline. The traditional responsibilities for resource allocation and territorial security, which were the prerogatives of national governance, are gradually shifting to cities and urban areas (Valladão, 2018).

This shift in power dynamics is not happening in isolation. It is intricately connected to the exponential growth of technology. Technological advancements have made it possible for cities—the new hubs of economic activities—to become more interconnected. This interconnectivity, powered by emerging technologies, has fostered the development of independent capacities that allow urban centers to make sovereign decisions without the oversight of central state authorities. In essence, national governments are witnessing a contracting of their institutional sphere of influence, also evidenced in their declining fiscal capabilities (Valladão, 2018).

This decentralization has sent ripples across the political arena, affecting not just the central governments, but also the many organs that are part of the political machine including parliaments and political parties. The clear dividing lines that once marked the boundaries of power are blurring, triggering a redefinition of roles and a rethink of hierarchical structures (Valladão, 2018).

One of the most significant hallmarks in this new political landscape is the emergence of Artificial Intelligence (AI) and its impact on the decision-making process. E-democracy, a concept that is broader than e-governance, represents a shift to a democratic system where people can participate in the decision-making process online. This implies that citizens have the ability to directly influence decisions

that affect them, overcoming barriers of distance and logistics to make their voices heard (dos Reis and Melão, 2023).

Information and communication technologies are the pillars upon which e-democracy stands. And with their continued evolution, they are playing a vital role in strengthening democracy. E-democracy is an enhanced form of democracy where political institutions and subjects communicate with the electorate through the internet, social networks, and mobile technologies. This ensures a channel for dialogue that is uninterrupted and accessible to all (dos Reis and Melão, 2023).

An analysis of e-democracy reveals three main axes along which it expands: basic conditions, e-government, and e-participation. For basic conditions to be satisfied, a well-developed infrastructure promoting access to digital media and unrestricted internet access is a prerequisite. E-participation evolves from mere information dissemination to active involvement of citizens in the decision-making process (dos Reis and Melão, 2023).

Illustrative examples include Portugal, where participatory platforms have redefined the government-citizen interaction, and Belgium, where the establishment of CitizenLab¹ is a testament to how AI technologies can be harnessed for political participation and decision-making support. Utilizing AI, CitizenLab not only encourages citizen engagement but also translates these engagements into categories that aid decision-making.

1 CitizenLab is a Belgian civic tech company which creates platforms for local governments where citizens engage.

E-democracy can also spur governments' actions to become more responsible and transparent. By providing avenues for online deliberation and consultation, they offer citizens the opportunity to express their opinions and influence policy decisions, helping to nurture trust with citizens. Additionally, AI aims for equitable treatment of all citizens, promoting transparency while also initiating discussions on privacy and ethics (dos Reis and Melão, 2023).

Simultaneously, the novel approach of e-government integrates information and communication technologies with the public sector. Rapid technological progression provides citizens with platforms to express their views and contribute to national development. The rise of e-government has brought about a raft of new administrative services making everyday lives easier, and the provision of these services further enables governments to align their functioning with citizens' needs, such as convenience, personalization, and ease-of-use (Rahmadany and Ahmad, 2021).

However, as we make strides in using AI in decision-making and governance, the pertinent issues of privacy, autonomy and fairness come to the fore. It prompts a careful consideration of the manner in which AI is implemented and used. This interaction requires a balanced partnership between people and machines. Notwithstanding our limitations, it is paramount to leverage AI in areas where human capacity can be bolstered (Phillips-Wren, 2012).

The primary function of e-government acts as a conduit for active participation of the citizenry in the decision-making process. The complementary forces of e-democracy and e-government work in tandem to foster a stronger relationship between governments and citizens. This bond can confer greater responsibility and

accountability to the government, therefore increasing public faith in its operations and cultivating a healthier democratic environment (Rahmadany and Ahmad, 2021).

The Hazards of Excessive Dependence and Unrestricted Application of Artificial Intelligence in Decision- Making Procedures

It is an undeniable fact that artificial intelligence (AI) has permeated virtually all aspects of our daily existence. Its applications are seen with increasing regularity in both the private and public realm, from enhancing the functionality of our homes to driving data-centric decisions in business, hinting at how radically this technological revolution is set to reshape our world.

Nevertheless, it is crucial to remain cognisant of the potential dangers that are associated with an over-dependence on AI, in particular when used without appropriate moderation and regulation in decision-making contexts, especially within governmental frameworks. This discourse focuses largely on the identification and exploration of such challenges, including the issue of human redundancy in administrative processes, ethical dilemmas that arise from machine-driven decision-making, and the broader implications these might present for the democratic system and civic involvement.

AI indeed brings a degree of efficiency and accuracy into decision-making procedures that is often difficult for humans to achieve consistently. However, as AI grows in its capacity and penetrates further into decision-making frameworks, it is impossible to ignore the concerns raised about the potential for human roles in these processes to become increasingly redundant. Brynjolfsson and McAfee (2014) have warned about the increasing likelihood of job roles, previously thought to be safe from automation, becoming obsolete. Such a scenario could prompt the sudden onset of several major societal problems, including mass unemployment and escalated societal inequality. A lack of regulation could ultimately result in administrative procedures that are stripped of human intervention and subjective judgement, leaving AI to make critical decisions.

Further adding gravity to these concerns is the tremendous growth that the AI industry has been experiencing. The International Federation of Robotics predicts that worldwide spending on AI systems will increase from just over \$37.5 billion in 2019 to nearly \$98 billion by 2023 (Fernandes, 2019). This rapid expansion could potentially stimulate a crisis characterized by job insecurity and escalating unemployment rates. This assertion aligns with a McKinsey Global Institute report, which estimates AI could displace around 15% of the global workforce by 2030 (Chui et al., 2017).

In conjunction with consideration of human obsolescence is the ethical aspect, another facet of AI incorporation into decision-making processes which cannot be overlooked. In conventional systems, humans have been responsible for making ethical judgements and exercising discretion in decision-making. Despite advances in the development of ethical systems within AI, there remains substantial

ambiguity and uncertainty surrounding how such ethical conundrums can adequately be addressed.

AI systems inherently lack the ability to exercise the kind of discretionary judgment that humans can—this is a significant complication, not least in terms of establishing accountability. This is a particularly severe ethical concern in sectors where the decisions made have far-reaching consequences, such as criminal justice or social welfare. Key questions about the transparency and fairness of AI-driven decision-making are raised, along with concerns about inadvertent bias, and potential misuse of sensitive information (Zarsky, 2016).

The potential risks of an excessive dependence on AI within governmental systems underscore the need for a balanced approach. As Bostrom (2014) suggests, technology, no matter how advanced, should serve as a tool that assists humans rather than replacing them. As highlighted by Schumpeter (1942), technological innovation does not follow a strict linear progression, implying that finding the ideal balance between human expertise and AI is a delicate process. This balance must be achieved while keeping in mind the significant benefits of AI, alongside the need for precautionary measures against potential risks.

Within a democratic system, active participation by citizens is a fundamental tenet. Thus, entrusting AI systems with more decision-making power could potentially lead to a decrease in such engagement. Citizen roles in a democratic system should not be confined to merely providing input for AI systems to analyze. Instead, AI should serve to enhance and complement human involvement in the democratic discourse (Bol et al., 2020).

In achieving this challenging equilibrium, the primary objective should be to develop a visionary approach to government where AI serves as a digital aid, augmenting human capabilities in decision-making rather than replacing them. As suggested by Bostrom (2014), our ultimate goal should be to capitalize on the impressive potential of AI while ensuring that this technology respects and upholds essential human values and needs,

Conclusion

The accelerated integration of artificial intelligence (AI) into governmental activities and procedures has the potential to revolutionize democratic systems on a global scale. However, as noted in the preceding sections of this paper, this technological revolution brings with it a series of intriguing challenges and risks, requiring an equally transformative shift in our approach to technology integration.

The objectives of this study, as summarized in the introduction, were to strike a balance between artificial intelligence's transformative potential and the necessity to preserve human agency in the decision-making process. The aim was to stimulate a debate that catalyzes a context-sensitive, efficient, and citizen-centric model of AI governance. This transformation would effectively serve a wide range of stakeholders involved in the democratic process, ranging from lay citizens to high-ranking government officials.

The methods used in analyzing artificial intelligence's role in modern governmental functions unravelled the multidimensional aspects and implications of its implementation, showing that AI is more than

a simple problem-solving tool. It is precisely this realization that has led to an understanding of how redundant human roles may become in administrative procedures, leading into further discussions on ethical dilemmas that arise from machine-led decision-making processes. Aside from the significant ramifications that this realization holds for democratic systems and civic involvement, a lack of human intervention could lead to the unchecked misuse of the technology, bringing with it the potential to compromise civil rights and personal freedoms.

However, it is equally pertinent to bear in mind the limitations of this study. While an attempt was made to provide a comprehensive overview of the impending AI disruption, a fully conclusive exploration remains elusive due to the rapid progression of the field, and constant technological advancements, as well as evolving legal and regulatory frameworks. Even as the findings of this study draw attention to the potential pitfalls and challenges associated with AI integration, a measure of uncertainty is inherent due to the nascent nature of the field.

The study opens up avenues for future research inquiries. One of the most compelling of these relates to the direct influence of AI on the processes of e-democracy. An investigation into the technologies propelling e-democracy would not only provide a clearer picture of the advantages and risks posed by AI but also inform a proactive strategy for its integration. Moreover, in-depth examination of specific case studies where AI has been embedded in governmental functions could deepen understanding of the practical implications and outcomes of such integration.

Exploration of countermeasures to avoid human obsolescence is another crucial research interest to be pursued. It is also noteworthy to

explore how a human-AI partnership could be cultivated in a way that neither is the human judgement compromised nor is the potential of AI not fully capitalized. Future work should also aim to investigate the ethical consequences of machine-driven decision-making processes, focusing on ways to maintain transparency, fairness and uphold the right to privacy whilst harnessing the efficiency and capabilities of AI.

The future of an AI integrated governmental framework is not too far down the road. It thus becomes imperative that rigorous scholarly attention be given to shape an environment wherein AI supports, rather than supplants, human involvement in democratic systems. The goals that this paper aimed to set out may seem challenging in the current context, but it is clear that they are not impossible to attain. Recognizing the great potential of AI while simultaneously addressing its challenges is the crucial step toward a future where AI brings about enhanced decision-making processes, increased citizen engagement, and improved public service delivery in a balanced and responsible manner.

It is, therefore, fair to conclude that as society moves forward, the objective should always be to use AI as a beneficial tool for humanity, combining human expertise with machine abilities to enhance democratic functioning and ultimately, better serve society. This conclusion, perhaps encourages us to step back and consider the broader significance and ramifications of our current path, and through reflection, envisage a future where technological advancement and human ingenuity merge to create a more efficient, democratic, and inclusive society.

The paper presents findings of a study developed as a part of the research project “Serbia and challenges in international relations in 2024” financed by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, and conducted by Institute of International Politics and Economics, Belgrade, during the year 2024.

This paper was realized with the support of the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, according to the Agreement on the realization and financing of scientific research.

References

- Andreasson, U. & Stende, T. (2019). *Nordic municipalities' work with artificial intelligence*. (Analysis 03/2019). Nordic Council of Ministers. <http://doi.org/10.6027/NO2019-062>
- Arntz, M., Gregory, T., & Zierahn, U. (2016). *The risk of automation for jobs in OECD countries: A comparative analysis* (OECD Social, Employment and Migration Working Papers No. 189). OECD Publishing. <https://doi.org/10.1787/5jlz9h56dvq7-en>
- Bol, D., Giani, M., Blais, A., & Loewen, P. J. (2020). The effect of COVID-19 lockdowns on political support: Some good news for democracy? *European Journal of Political Research*, 60(2), 497–505. <https://doi.org/10.1111/1475-6765.12401>
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.
- Chui, M., Manyika, J., & Miremadi, M. (2017). *What AI can and can't do (yet) for your business*. McKinsey & Company. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/what-ai-can-and-cant-do-yet-for-your-business>
- Desouza, C. K. (2018). *Delivering artificial intelligence in government: Challenges and opportunities*. IBM Center for the Business of Government.
- dos Reis, J. C. G., & Melão, N. F. R. (2023). E-democracy: Artificial intelligence, politics and state modernization. *Proelium*, 8(11), 404–422.
- Duberry, J. (2022). *Artificial intelligence and democracy: Risks and promises of AI-mediated citizen-government relations*. Edward Elgar Publishing.

Gesk, T. S., & Leyer, M. (2022). Artificial intelligence in public services: When and why citizens accept its usage. *Government Information Quarterly*, 39(3). <https://doi.org/10.1016/j.giq.2022.101696>

IDC. (2022). *Worldwide spending on artificial intelligence systems will be nearly \$98 billion in 2023, according to new IDC spending guide*. IDC. <https://www.idc.com/getdoc.jsp?containerId=prUS49670322>

Merh, H. (2017). *Artificial intelligence for citizen services and government*. Ash Center for Democratic Governance and Innovation.

Phillips-Wren, G. (2012). AI tools in decision-making support systems: A review. *International Journal on Artificial Intelligence Tools*, 21(2), 123–135. <https://doi.org/10.1142/S021821301250003X>

Rahmadany, F. A., & Ahmad, M. (2021). The implementation of e-government to increase democratic participation: The use of mobile government. *Jurnal Studi Sosial dan Politik*, 5(1), 22–34. <https://doi.org/https://doi.org/10.19109/jsssp.v5i1.8552>

Schumpeter, J. A. (1942). *Capitalism, socialism, and democracy*. Harper and Brothers.

Sigfrids, A., Leikas, J., Salo-Pöntinen, H., & Koskimies, E. (2023). Human-centricity in AI governance: A systemic approach. *Frontiers in Artificial Intelligence*, 6, 5–7. <https://doi.org/10.3389/frai.2023.976887>

Smith, G. (2018). *Democratic innovations: Designing institutions for citizen participation*. Cambridge University Press.

Valladão, G. A. A. (2018). *Artificial intelligence and political science* (policy paper, p. 15). OCP Policy Center.

Vogl, T. (2021). Artificial intelligence in local government: Enabling artificial intelligence for good governance in UK local authorities (Working paper). *Oxford Commission on AI and Good Governance*.

Warren, M. (1992). Democratic theory and self-transformation. *American Political Science Review*, 86(1), 8–23. <https://doi.org/10.2307/1964012>

Yigitcanlar, T., Agdas, D., & Degirmenci, K. (2022). Artificial intelligence in local governments: Perceptions of city managers on prospects, constraints, and choices. *AI and Society*. <https://doi.org/10.1007/s00146-022-01450-x>

Yigitcanlar, T., Corchado, M. J., Mehmood, R., Li, R. Y. M., Mossberger, K., & Desouza, K. (2021). Responsible urban innovation with local government artificial intelligence (AI): A conceptual framework and research agenda. *Journal of Open Innovation: Technology, Market and Complexity*, 7(1), 1–16. <https://doi.org/10.3390/joitmc7010023>

Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic roadmap to examine efficiency and fairness in automated and opaque decision-making. *Science, Technology, and Human Values*, 41(1), 118–132. <https://doi.org/10.1177/0162243915601988>

The Business of Umbilical Cord Biobanking in Serbia: Biocapital(ism) and Symbolic Geographies of Health

Ljiljana Pantović
Jelena Čeriman

Introduction

The storage of stem cells from umbilical cord blood for private, individual purposes has emerged as a significant social phenomenon over the past two decades. The technology of cord blood storage offers the potential for stem cells to be used in the treatment of various illnesses, creating new possibilities for medical intervention and cure. The storage of stem cells obtained from post-birth umbilical cord blood (UCB) has emerged as a significant area of scientific research and medical advancement. This process involves the collection and preservation of UCB stem cells for various purposes, such

Ljiljana Pantović, Research Fellow, Institute for Philosophy and Social Theory, University of Belgrade, Belgrade ORCID: 0000-0002-2018-9643, email:

ljiljana.pantovic@ifdt.bg.ac.rs

Jelena Čeriman, Research Fellow, Institute for Philosophy and Social Theory, University of Belgrade, Belgrade ORCID: 0000-0002-3848-5926, email:

jelena.ceriman@ifdt.bg.ac.rs

DOI: <https://doi.org/10.5281/zenodo.14204195>

as scientific investigations or personalized medical treatments. Biobanks, which serve as repositories for biological materials, can be categorized into two types: public and private, depending on their collection methods and utilization practices. Public biobanks primarily rely on voluntary donations and supply umbilical cord blood for medical requirements, while private biobanks collect and store individuals' cord blood for a fee, ensuring accessibility for their own future medical needs. Private biobanks have expanded globally, crossing national borders to offer their services to parents seeking to store cord blood for potential medical use.

Although a growing body of scholarship has examined UCB biobanking (Beltrame, 2020; Beltrame and Hauskeller, 2019; Cooper, 2011; Gottweis and Petersen, 2008; Hauskeller and Beltrame, 2016; Jeong, 2014; Katz et al., 2011; Mitchell and Waldbly, 2010; Murata, Kawabe, Hatta, Maeda, and Fujita, 2022; Waldbly, 2006), the existing literature predominantly concentrates on the phenomenon of biocapitalism (Beltrame and Hauskeller, 2019; Helmreich, 2008; Sunder Rajan, 2006; Venkatesan and Peters, 2010) within advanced capitalist societies. Consequently, there is a noticeable lack of research that critically explores the emergence and engagement of UCB biobanking within post-socialist contexts with only a handful of studies (Krajnović, Stojković, and Babić, 2016; Krstić, 2012; Velikonja, Erjavec, and Knežević, 2021). By addressing this research gap, we can gain a more nuanced understanding of the multifaceted implications and potential challenges associated with UCB biobanking, thereby contributing to a more comprehensive global discourse on this topic.

From the second half of the 20th century, we can map the rise and development of a new form of technology, biotechnology, which has put into question the key notions and ways we understand life, nature and

what is changeable and unchangeable by humans. This new form of technology has been described by theorists of science and technology studies as an “emerging science” (Latour, 1993). This means that there is no unified understanding of the dangers and benefits of these new forms of technology (Waldby, 2002). What is also important to point out is that this uncertainty surrounding biotechnologies does not arise because technology and science are outside the socio-cultural and political realm. On the contrary, they are a crucial part of it. Biotechnologies are shaped by current political and cultural frameworks as much as they in turn shape those frameworks themselves. According to Waldby (2002), biotechnology destabilized and changed the naturalized connections between the body, parts of the body and human identity and social systems. Furthermore, biotechnologies became a new way to analyze relations between the body, biopower and biopolitics (Lepetić, 2014).

The term biocapitalism refers to the process of inserting capitalist market values and logic to biological substances, such as tissues, organs, stem cells through biotechnological processes. Stefan Helmreich (2008) argues that biocapital and biocapitalism builds on Foucault’s notion of ‘biopower’ to include the governance of cells, molecules, genomes, and genes. Helmreich (2008, 463) defines biocapital as the “contemporary unions of biological science with profit-oriented enterprise”. Specifically, biocapital is a novel kind of capital that inserts stem cells and genomes into projects of product making and profit-seeking. Private biobanking of human tissue and the faith in the regenerating powers of stem cells is becoming a central topic in understanding the impact of biotechnologies on human life.

With this discovery came the possibility of cryopreserving and storing cord blood for both further clinical research and for medical

treatment. In essence, what is being done is that the cord blood is being effectively taken “outside of time” in both the biological and historical sense so they can be used later. The first cord blood banks were established as parts of public blood centers in New York and London in the early 1990s. Similar to public blood banks, cord blood is donated voluntarily and used for both research and the treatment of others. The first private biobank was built in Connecticut by the Bio-cyte Corporation in 1993, at the same time as the establishment of public cord blood biobanks (Appleby–Arnold, 2012).

Unlike public cord blood biobanks, private banks operated with a different logic. These are for profit institutions which offer their facilities to store individual cord blood that can be used solely for the treatment of the “account holder”: that is, the infant and the infant’s family from whom the blood was extracted. The “account holder”, or rather the future parents are expected to pay an annual fee for this service. In this manner the storage of cord blood storage can be seen as a kind of future health safety deposit box. Private biobanks charge a fee to store UCB exclusively for the benefit of their client. Private biobanks have, in the past three decades, embraced the global market and crossed national borders. According to the data collected by a non-profit foundation aimed at promoting and educating parents about UCB—Parents guide to cord blood, there are 214 registered private biobanks in the world with storage facilities in 55 countries but operating in 93 countries worldwide through affiliates. A biobank can be geographically located in one country, but its clientele is usually international. What is being offered to the consumer – the parent – is the technology to store the cord blood of the child for a finite amount of time (between 20 and 50 years) if either the child or a sibling develops an illness for which stem cells in the cord blood might become the

basis of a cure. Unlike public UCB biobanks, that function along similar practices as blood banks and rely on discourses of donation, gift giving and solidarity (Waldby 2006), with private biobanks the treatment of UCB as a biological commodity is clear.

According to Igor Kopytoff (1986), things themselves are not intrinsically saturated with meaning; it is the social and cultural relationships that they are connected in that provide meaning. When the umbilical cord is cut during childbirth in most cases it is thrown away as biological waste after birth. The example of the change in status of umbilical cord blood from singularized waste to “expensive singularization” (Kopytoff, 1986) highlights the fluidity and change of this tissue’s significance. In the late 1980s, cord blood was revalued as an alternative solution to bone marrow transplants in treating blood diseases and thus transformed from an “abject tissue” (Waldby and Mitchell, 2006) into a form of “biocapital” (Helmreich, 2008). What was discovered is that cord blood is a source of hematopoietic stem cells that can be used to essentially regenerate or rebuild one’s immune system. Cord blood is extracted during birth, after the cord has been clamped. Due to the relatively small amount of cord blood that is taken from a single birth, these types of transplants are predominantly used for the treatment of children. Unlike bone marrow transplants, cord blood is more flexible, easier to match, and the harvesting procedure is less invasive (Waldby, 2006).

Rather than treating UCB private biobanking as either a commodity or gift, Waldby and Mitchell suggest viewing it as a “biological investment” (Waldby and Mitchell, 2006) and in that sense falling within the framework of biocapital and biocapitalism. Storing UCB in private biobanks is a form of future oriented property. To preserve the child’s

cord blood solely for personal use is seen as a form of biological insurance, and it operates according to neoliberal principles of private insurance which offers personalized risk-management services to contradict uncertainties of the future (Waldby and Mitchell, 2006). Other scholars have also noted the future oriented private biobanks operating through “regimes of hope” (REF) in still uncertain medical usage of stem cells, in opposition to present oriented public biobanks that operate through “regimes of truth” and trust in current scientific and medical achievements (Brown and Kraft 2006; Martin et al. 2008; Beltrame, 2014). Private biobanks capitalize “hope” of parents by mobilizing “promissory expectations” for future medicine in a neoliberal market (Martin, Brown, and Turner, 2008).

The aim of this article is to understand how private biobanking, as a form of biocapitalism (Sunder Rajan, 2006; Venkatesan and Peters, 2010), is taking shape in post-socialist Serbia. In order to contextualize the study and theoretically conceptualize the subject of the research and methodological framework that follows, we interpret socio-historical, political, and cultural factors in post-socialist Serbia that shape the attitudes of parents on the subject matter. This chapter ends with analysis and concluding discussion.

Contextual and Theoretical Conceptualization of the Research: Biobanking in Serbia

The dismantling of former Yugoslavia and the subsequent civil war left a major impact on the health care systems of the former member states. During the civil war (1991–2000), all the hospitals in Serbia

lacked basic supplies and experienced immense delays in medical staff salaries (Tosic, 1992). Due to the civil wars in the 1990s, sanctions and inflations, and general political instability, the country's health care system faced a substantial blow (Tosic, 1992) from which it is only now slowly recovering (Becker, 2009).

However, according to Krstić (2012), Serbia and Albania are the only two Balkan countries that do not have public biobanks even though Serbian medical officials state that the country is the leader in the region when it comes to stem cells and their storage. Serbia has received support from the European Investment Fund for the opening of a public biobank in Kragujevac. The biobank would be named "The Center for Excellence," but its construction has not commenced yet (B92, 02/14/2014). According to some news articles there were over 12 international biobanks operating in Serbia in 2015 (Newsweek Serbia, 05.04.2015) but only 3 could be found on the official Business Registry. In 2019 one of the more prominent biobanks working in Serbia, Cryosave, lost their license to store cord blood in Switzerland and work in Serbia. Subsequently the biobank declared bankruptcy (parentsguidecordblood.org/June 2019). This news was a shock to a lot of Cryosave clients in Serbia (Novaekonomija, 2019) and most of the parents who stored their cord blood with this bank filed lawsuits that are still ongoing (N1 2019). On the other hand, all the cord blood that was stored by Cryosave in Switzerland was moved to another biobank in Poland (N1, 2019).

In most cases, biobanks are only present in Serbia through their online presentations, call centers and information offices in the capital city—Belgrade, but not through physical presence on the territory of Serbia, neither public nor private, which clearly indicates division between periphery and the center and reveals regularities of local

geographies of health (and power). Thus, the advertised biobanks can be categorized as multinational global biobanks, accessed by clients across the globe. Thus, the emergence of private biobanking can be viewed through the lens of the continuing neoliberalization and privatization of healthcare in post socialist countries (Pantović, 2018, 2022) and the larger shift towards new regimes of care in the anthropocene (Pantović and Zarić, 2022).

Rare research on this topic in Serbia indicates that stem cell research and their storage in Serbia are not related to the EU accession, given that the stem cell research was ongoing even before Serbia's official declaration of pro-European policy (Krstić, 2012). However, stem cells are used as a promotion of the (civilizational) progress of Serbia (in relation to Europe) (ibid., 2012.). All the research conducted in the social sciences in this time period indicate that the issue of the use of biotechnology for the preservation of stem cells can serve for the examination of the distribution of power and the establishment of control over the body and the health of people in a society (Krstić, 2012; Lepetić, 2014). Therefore, it is rather an action that includes and prescribes a new concept of parenting and parental care for children's health. Krstić argues that biotechnological possibilities have conceptualized a new meaning of health, which becomes the necessity of investing in one's own future and future of the family, i.e. children, with the help of preservation of stem cells. Such a new concept of health prescribes how a successful parent should behave, that is, what good parenting actually means (Krstić, 2012). The ethics of good parenting have a particular manifestation in Serbia due to the high presence of familist ideology or familialism in Serbian society. Familialism or familist ideology sees the ideal of the family in a union whose primary function is the care of a child whose needs always come before the

needs of the parents (Tomanović, 2017; Čeriman, 2021). Next to this, contemporary parenthood in Serbia includes combining of norms regarding child's development as an autonomous person, with the need to protect children from risks posed by contemporary society (Tomanović, 2010).

An important aspect for understanding the latter is the country's turbulent past, especially the 1999 NATO bombings, that also had a toll on the population's (perceptions of) health. This past has led to the widespread perception that Serbian children are at elevated risk of cancer and that the Serbian health-care system cannot adequately manage the problem. Due to the civil wars in the 1990s, sanctions and inflations, and general political instability, the country's health care system faced a substantial blow (Tosic, 1992) from which it is only now slowly recovering (Becker, 2009). One of the major events that shaped the current healthcare climate, as well as people's perceptions of their health, was the NATO bombing of Yugoslavia in 1999. The use of those weapons exposed the population to increased cancer rates, according to some medical experts in the region (Jovanović, Petković, and Čikarić, 2012). Throughout the ensuing years, and especially on the anniversary of the bombings, Serbian dailies write about the impact that the bombing had on cancer rates. The head of pediatrics and hematology of the University Children's Clinic in Belgrade stated recently in the widely read daily "Danas" (11/02/2015) that Serbia has had a substantial increase in cancer diagnoses among children, which he and his colleagues link to the usage of around 45 tons of depleted uranium. "In the year 2000 there were on average 200 new cases on a yearly basis, by 2013 and 2014 that number has gotten up to 350 per year which is a significant increase" said the doctor for this daily newspaper. The doctor ended his interview by stating that Serbia

does not have the capacity for the treatment of all those children and that the future is in stem cell treatment, which means “Serbia is in dire need of a public umbilical cord biobank because that is potentially the savior for the gravest of diseases” (Danas, 02/11/2015). Furthermore, Serbia’s 2001 participation in the World Values Survey showed high esteem for the family, family authority and sacrifice for children (Ćeriman, 2019: 403).

Familialism in an institutional context is manifested in connection with expensiveness and unaffordability of the institutional support to health of the general population, so it is presumed that families will take responsibility for the care of their members rather than leaving that responsibility to the government. Existing analysis of institutional ethics of care for the health and social security of the population in Serbia point to the dehumanizing and bureaucratic approach in the first contact of citizens and frontline workers, and mistrust and unprofessional behavior of employees in state institutions, which is highly represented if the focus is on the positions of vulnerable categories of the people (ethnic or national minorities, women etc.) (Ćeriman, Hrnjak, Radoičić, Nedeljković, 2022). Furthermore, public services are nominally generally available, but they are of low quality, which is especially true for healthcare protection, which is at least declaratively provided to everyone by the state. However, there are waiting lists for certain health services, as well as additional formal and informal payments (Pantovic, 2016) for them due to the low budget allocations by the state. Next to it, research on the quality and accessibility of various social services in Serbia show that they often do not address the real needs of citizens, and that system also needs assistance in order to realize all the prerogatives of the state that protects and helps its citizens (Milutinović, Bojanić, Ćeriman, et al., 2016; Milutinović,

Bojanić and Čeriman, 2016). Therefore, Serbia can be classified in the group of “incongruent social regimes” (Lendvai, 2009), and that position is characterized among other things, through the primary reliance on the family as the main actor of social care, which is also the dominant locus of the solidarity expressed through kinship (Čeriman and Vučković Juroš, forthcoming). In such a situation, the emphasis is placed on the responsibility of parents who should make informed decisions because the health of their children depends on them. Krstić argues that “a person who decides to keep the stem cells, at the same time is a informed person, familiar with medical progress; a health-insured person, a financially wealthy person and, what is most important, a person who not only takes care of his family at the moment, but also thinks about their (and therefore his) future.” (Krstić, 2012: 160). “The refusal (or failure to do so) of depositing the child’s stem cells represents exclusively the personal fault of the parent or “personal failure” of the parent.” (ibid.: 162)

Our research seeks to examine the business of private umbilical cord biobanking in Serbia and its interconnections with biocapitalism and symbolic geographies of health, through the exploration of narratives about good parenting, in order to gain insight into this global phenomenon from a local perspective. Based on the presented theoretical framework for the purposes of this research, we have developed a definition of good parenting which is expressed as: health-insured and financially well-off parents who are well informed and familiar with medical progress and who think about the prospects of the family whose primary function is care for child’s health that always come before the parents’ needs, Such parents make informed and independent-from-the-state-support decisions about their child’s health and take full individual responsibility for them.

Therefore, the analysis of parent's narratives in this article will be guided by the following research questions:

- 1 | How are parents informed about this emerging technology and their role in ensuring children's health?
- 2 | How do future parents, potential biobanking clients, conceptualize new emerging technology in the local context of a post-socialist society of Serbia?

The analysis of parents' narratives will be compared with the content analysis of online biobanking presentations in order to determine similarities and differences in understanding the concept of good parenting.

Therefore, the content analysis of biobank presentations will be guided by the following research questions:

- 1 | How do biobanks present themselves online and in real-life settings to attract potential Serbian clients?
- 2 | What meaning do they attach to the concept of good parenting regarding children's health and how do they perceive their own role in fulfilling that parenting goal in the local context of a post-socialist society of Serbia?

The initial hypothesis of the research is that in a sample of materially well-off parents whose family life takes place in the risky society of Serbia, which poses a threat to children's health, the narrative of parents will contain the same elements as biobanks presentations on the personal responsibility of parents to protect children's health through the use of biotechnology.

Data that is presented and interpreted in the analysis are collected in exploratory ethnographic research realized during 2016/2017, as part of a larger study on privatization and consumer practices in maternal care in Serbia (Pantović, 2019). The primary data collection technique includes in-depth interviews with 6 pregnant women on their attitudes on UCB biobanking. After attending different private birthing classes, one woman was randomly asked for an interview on her opinions on biobanking. The interviews were transcribed and coded using Nvivo qualitative software for the purposes of a narrative analysis. The research followed good research practice, and all of the interviewees signed consent form and the entire larger research went through and obtained ethical review board approval from the University of Pittsburgh¹.

Due to the Covid19 pandemic additional data on biobanking in Serbia were collected online and is based on news articles in Serbia on UCB biobanking and the web page presentation of active biobanks in Serbia in 2021. According to the foundation Parents guide to cord blood², there are currently four banks actively operating in Serbia. This information provided the selection criteria for the analysis of the internet presentations of these four banks. These webpages were converted into pdf files and analyzed using MAXQDA software for a content analysis of online data.

1 The larger project was part of Dr. Pantović doctoral research at the University of Pittsburgh, USA.

2 <https://parentsguidecordblood.org/en/family-banking/serbia> accessed last 31.05.2023.

Analysis of parent's narratives:

Parents in Serbia receive information about the emerging technology through various channels. Since there are no private or public biobanks within the country, parents depend on representatives from international biobanks, primarily located in Belgrade, to keep them updated. These representatives have a crucial role in actively disseminating information to parents. Additionally, the majority of pregnant women in Serbia acquire knowledge about this new technology through online sources, social media, privately operated birthing schools, and pregnancy fairs. These platforms serve as valuable avenues for parents to gather information, ask questions, and receive guidance regarding emerging technology and its potential impact on their children's well-being. To illustrate how parents are informed about umbilical cord blood (UCB) biobanking, two ethnographic vignettes will be presented: one from a maternity/pregnancy fair held in Belgrade in 2016, and another from a private birthing school where UCB biobanking was discussed.

Baby Mixer Festival

The Baby Mixer Festival, an event heavily advertised on social media, unfolded in the vibrant neighborhood of Savamala, Belgrade. Taking place at Mixer House, a central venue known for hosting trendy gatherings, this festival targeted the young, indie rock alternative subculture of Belgrade, often referred to as hipsters. While Mixer House typically hosts its summer Mixer Festival, catering to this subculture, the Baby Mixer emerged as a spin-off event with a focus on parenting. This

marked the festival's second occurrence, commencing in the morning around 10 am and extending until the evening at approximately 8 pm.

Transformed into a bar and concert venue, Mixer House encompassed a spacious loft setting for the festival. Upon entering, attendees encountered an array of booths showcasing various young craft vendors offering handmade baby paraphernalia, including blankets, clothing, toys, and other enticing items. Adjacent to these stalls, a section featured more upscale booths, with the first one belonging to the German umbilical cord biobank, BioSave. Their booth was adorned with a conspicuous banner depicting a cartoonish pregnant woman dressed in green, exhorting prospective parents to participate in their promotional cord blood storage program. To incentivize engagement, they held a raffle for a maternity pillow, alluringly encouraging attendees with the motto "*Be carefree.*" (*Budi bezbrižna*). I spoke to the representative of BioSave, Tamara.

Possessing a background in German language studies, Tamara had been employed by BioSave for the past five years. While initially working as a translator, she now assumed the role of one of their principal agents, actively recruiting expectant parents and dispensing medical advice both in person and through the company's pregnancy blog. Tamara shared insights into the bustling nature of their involvement during such events. The festival attracted a multitude of parents who exhibited a keen interest in investing in the future health of their offspring. According to Tamara, these parents often hailed from well-paying occupations, with some being foreigners or employed by foreign companies and institutions, adding to the cosmopolitan ambience of the Baby Mixer Festival.

Pregnancy Schools

Women are encouraged by the public health care system, and in many cases by their families, to attend birthing classes when they enter their 28th week of pregnancy, in order to prepare for labor psychologically and physically. In the city where fieldwork was conducted in 2016, there is only one state-run Škola za trudnice (literally “School for pregnant women” or birthing school), offered in the primary care center two blocks away from the maternity hospital. Nurse-midwives run this school, and there are between 15 and 20 pregnant women per group. These classes are completely free and are covered through state health insurance. That said, as with the prenatal check-ups, women also have two private options available to them when it comes to these types of classes, two private schools both associated and financially supported by international biobanks. In these two schools there are posters, pamphlets, fliers, and other merchandise regarding UCB biobanking. At the start of each program in the private schools a representative of the supporting biobank gives a promotional lecture about the benefits of investing in UCB biobanking and encourages the women who attended the school to store UCB upon childbirth. Occasionally, the representative would also provide a quick presentation before the start of another lecture, such as the vignette below.

The lecture that was scheduled in the private pregnancy school was on baby handling and breastfeeding. This lecture was given by a retired midwife, whom the organizer had spoken highly of. There were 17 women in attendance sitting in a circle in the room. Before introducing the midwife, Marijana, the organizer, wanted to provide them with some information about the offerings of Seracell, a German biomedical center operating in Serbia. A representative from Seracell joined

us in the circle and began discussing what set their center apart from traditional biobanks in the country.

With a warm and informative tone, the representative explained that Seracell distinguished itself by not being a traditional bank but rather a medical center. She emphasized the higher standards of preservation at Seracell, with cleanliness rating at A+ compared to the B+ rating of even the best hospitals in Serbia. The entire staff at Seracell was well trained and professional, ensuring that not only were cells saved but also that the center had an in-house team equipped to handle the cells when needed. The Seracell representative further reassured the women that only samples meeting the necessary criteria for future usability would be stored. Additionally, she highlighted the financial aspects, mentioning that payment would only be required once the sample was successfully stored and confirmed. The representative concluded by stating that Seracell's current promotional price for cord blood storage was 1790 euros.

Following the representative's departure, the midwife, Sara, chimed in enthusiastically. She shared her personal experience of using cord blood storage for her grandchild and praised Seracell, expressing her recommendation. The endorsement from Sara carried weight among the women, who nodded in agreement.

These two vignettes highlight the two main in-person channels that are present in Serbia when it comes to providing information about UCB biobanking. It is noticeable that the primary target audience are pregnant women, future mothers. Specifically, women who live in large urban areas and of higher education and income. Based on the most recent UNICEF Multiple Indicator Cluster Survey (MICS) conducted in Serbia, only 16% of women in Serbia attend pregnancy

classes³ and 26% of them are in urban areas, 26.5% are highly educated and 28.9% of them have higher than average income. Considering the location where baby festivals are held and the information about the demographic of women, we can conclude that the information about this emerging technology is primarily targeting higher educated, higher income, urban pregnant women.

Interviews

After attending private pregnancy classes, we spoke with 6 pregnant women about their perceptions on UCB biobanking. Following the logic of the priceless child's health (REF), parents usually choose biobanks that offer possibilities of discounts and monthly payments as a way of negotiating the price of private biobanking. This does not mean that biobanking is widely accessible to all Serbian citizens, but that it is something seen as desirable and worth investing in by both upper classes and by the urban middle classes. One woman's comment puts the issue of price into perspective by comparing the costs of cord blood storage with other expenses.

What do you have after the wedding pictures, memories, and you spend a lot of money...what happens if you do not invest in cord blood and your child gets sick? The child dies...Some things are priceless; I would sell my kidney on the black market if needed all for my child's health. (First time mother, 32 years old, urban).

³ This sample does not distinguish between public and private classes.

Other women also raised the topic of better to store and “throw the money” than risk having a sick child or family member. This woman was particularly interesting as she took advantage of both the fairs and the classes to gather information about UCB.

We signed up during a fair and got a discount of €150. We're paying for 12 installments and ok. I insisted on it because my mother died of lymphoma at the age of 30 and lymphoma is treated with stem cells. I always say I hope it's the best €2,000 wasted in my life. (First time mother, 33, urban)

The emphasis on individual responsibility and taking an active, preventative role in the future health of one's child is placed in the backdrop of general distrust and uncertainty of the health care system and overall qualities of life in Serbia. One woman told us that:

The doctors here are corrupt...The state of our hospitals is just terrible! Imagine, God forbid something happens I am sure they do not have the resources to save my child. We saved it, but in Germany. I have no confidence in Serbian companies. Most will tell you that if the money turns out to be wasted, it's certainly better than God forbid one day they need cells, and you haven't saved them. It's not a small amount of money, but it's not too much if you think and believe that one day it can help your child or even you, your husband, another child. (First time mother, urban, 31 years old)

Other women who attended the pregnancy classes explicitly stated that the main reasons why they decided to invest in biobanking

for their newborns is because of the (potential) effects of the NATO bombings on their offspring.

I was at the hospital the other day. I passed by the oncology and radiology wing, and I saw hallways and floors packed with emaciated children with no hair. When that type of cancer happens to someone there is no chance they can beat it. This is sadly our post-NATO reality. Then you really get scared because you have a small child and want another soon...I am certainly saving some money for stem cells; it is a great investment. (first time mother, 25 years old, urban)

For these women, saving money and storing cord blood is seen not as prevention from a very slight risk but a prevention from the general uncertainty felt living in Serbia. In this sense storing of cord blood as intimate tissues in out-of-the country storage facilities transform social and everyday life problems living in Serbia and turns them into individual pathologies that they can react on and prevent. Because their child's health is seen as priceless, and the health context of Serbia is seen as insecure it is their responsibility to protect their child's health. One mother pointed out the importance of the future oriented aspect of storing UCB and her general trust in technology and medicine. She had already stored cord blood during her previous childbirth and plans on storing again.

We have been guided by the fact that we are leaving them in storage for at least twenty years, and we do not know in what direction, at what speed and how medicine will progress and

what will be possible with them then compared to today. (Second time mother, stored previously as well, 33 years old, urban).

The desire to store this intimate tissue outside of the borders of Serbia means that parents have certain ideas and understanding of the places where this tissue will “migrate” to. Parents have certain preconceived notions of the countries where the facilities of biobanks are located. In almost all of the discussions dealing with choosing the “right” biobank culturally embedded notions about various countries across the globe are present. Several women voiced distrust in storing cord blood in China or Greece. “We saved it, but in Germany. I have no confidence in Serbian companies.” By stating that, they are drawing on shared cultural stereotypes of Germany as precise, secure, trustworthy, and scientific. Similarly, another woman noted the same notions about Switzerland.

We decided to go with Future Health. They have a representative office in Belgrade, but it is a Swiss company. Stem cells are guarded for 25 years. The Swiss are Swiss, after all. Plus, their certificate networks preserved stem cells with public banks around the world, so you can access the registry and take another sample, if the preserved one is not suitable for treatment. It's different, at least as far as I know. (First time mother, urban, 25 years old)

Through these symbolic geographies mothers are left with a sense of security and faith in this process as they themselves do not have the opportunity to visit the biobanks themselves but rather get

certificates confirming their storage. This means that in actuality that have no security that this expensive biological insurance is being shipped and stored properly, their only security lies in the imagined geography of the country where the biobanks are located. Thus, as it is not the whole body that is taking part in this global mobility the usage of symbolic geographies ascribed to the locations of these biobanks provides an additional level of security and information for the parents who are entrusting, for them, the priceless future health of their child.

Content Analysis of Online Biobanking Presentations

Initially, the research plan envisioned further interviews with representatives of all currently operating UCB biobanks in Serbia. As a result of the Covid19 pandemic, supplementary information regarding biobanking in Serbia was gathered online, utilizing news articles from Serbia concerning UCB biobanking and the official web page presentations of active biobanks in Serbia in 2021. According to the Parents Guide to Cord Blood foundation, Serbia currently has four actively operating banks. This data served as the basis for selecting and analyzing the online presentations of these four banks, which were converted into PDF files using the MAXQDA web collector tool and subjected to content analysis using MAXQDA software. The main themes noticeable in all the biobanks' marketing materials are the notions of safety, trust in science and technology, and the importance of geographic location of the biobanks. These themes are globally

present, but coupled with narratives of parents we can see that these themes have particular meanings in local contexts.

Our analysis was guided by the following questions: How do biobanks present themselves online to attract potential Serbian clients? What meaning do they attach to the concept of good parenting regarding children's health and how do they perceive their own role in fulfilling that parenting goal in the local context of a post-socialist society of Serbia?

The Price of Pricelessness

While some of the women we spoke to describe the possibility of buffering potential future health risks of their children as priceless, commercial biobanks operating in Serbia have put price tags on their services. Based on the data available online all four biobanks charge a similar amount to store core blood for 20 years (1880 to 1980 euros). Below we have presented a table with the list of all of the biobanks in Serbia, locations in Europe where they store the cord blood and the price for 20 years of storage. This amount can either be paid in full or in instalments ranging from 3 to 6 months with the biobank or by taking out a commercial loan from a financial bank without interest for up to 48 months. In all of the four cases analyzed if the parents choose to pay in instalments, they have to pay a portion up front. For example, on their website Biosave has a calculator indicating that if the parents want to store the cord blood of one unborn child and want to store the UCB in Germany for twenty years, they have to pay 1880 euros in

total. They can choose to pay in 12 monthly instalments of 110 euros per month with 564 euros upfront.

According to the available public data from 2021 it was estimated for a living wage in Serbia one needed around 115 000 dinars or 978 euros⁴. This is the income level that is considered sufficient to meet basic needs and maintain a modest standard of living within a specific geographic area. It is designed to ensure that individuals and families can afford essential expenses such as housing, food, healthcare, transportation, education, and other necessities. The concept of a living wage considers factors such as the cost of living, local economic conditions, and the specific needs of different household types. To put this information into perspective, according to the Statistical Office of the Republic of Serbia the average net salary in March 2023 was 85,485 Serbian Dinar (RSD) or approximately 730 euros⁵. This is below the estimated living wage two years ago and the prices have only gone up in the past two years.

What this information shows us is that for the average parent, even with two incomes paying monthly instalments for storing cord blood that as some mothers noted above is probably “wasted” can be treated as a form of luxury.

4 <https://cpe.org.rs/vesti/poslednji-podaci-plata-za-zivot-u-srbiji-114-752-dinara/> accessed May 31, 2023

5 <https://www.stat.gov.rs/en-US/oblasti/trziste-rada/zarade> accessed May 31, 2023

Table 1: Biobanks operating in Serbia, storage location and price.

BIOBANK	STORAGE LOCATION	PRICE FOR 20 years
BioSave	Poland and Germany	1880 Euros ⁶ for Germany and 1980 euros for Poland ⁷
Vita34	Germany	1980 Euros ⁸
Future Health	United Kingdom and Switzerland	1890 Euros ⁹
Family Health	United Kingdom	1980 Euros ¹⁰

How do Biobanks Present Themselves Online to Attract Potential Serbian Clients?

There are several themes that can be traced in all four-web presentation of biobanks that indicate how these banks represent themselves online to potential clients in Serbia. Given that cord blood stem cells and stem cells in general are still considered as a form of emerging biotechnology, a key feature in all of the websites is the emphasis on science and specifically future oriented science.

6 <https://biosave.rs/cene/kalkulator-cena-maticnih-celija> last accessed May 26th, 2023

7 <https://biosave.rs/cene/kalkulator-cena-maticnih-celija> last accessed May 26th, 2023

8 <https://www.vita34.co.rs/maticne-celije-cena/> last accessed May 26th, 2023

9 <https://futurehealthbiobank.com/rs/cenovnik-cuvanje-maticnih-celija/cuvanje-iz-krvi-pupcanika/> last accessed May 26th, 2023

10 <https://familyhealth.rs/cene-cuvanja-maticnih-celija/> last accessed May 26th, 2023

I believe with all my heart in the present and future possibilities of stem cells. The world we live in is increasingly important to the creation of electric goods, apartments, and cars, but we seem to forget the most precious things, our children. (FutureHealth)

To have a large scientific team that is the guarantor of the scientific approach (BioSave)

At Family Health, we process stem cells with modern Toticyte technology that produces 3 times more stem cells, patented by the English stem cell bank Cells4Life. When transporting samples to a bank abroad, we use state-of-the-art technology to maintain their maximum vitality by controlling the temperature. We care about every detail, because it is important for the success of therapy (Family Health)

Another key theme present on Sall the websites is the emphasis on safety and reliability of the specific biobank. These themes probably emerged as crucial for the Serbian context after the affair with the now non-solvent biobank CrySave that used to operate in Serbia and filed for bankruptcy. For example, BioSave in the frequently asked questions section addresses specifically the question of what happens to the stored cord blood if the biobank stops working:

When signing a contract with Bio Save for storing stem cells, the contract is also signed by the director of FamiCord and with his signature guarantees that all the terms of the contract will be fulfilled. Also, if there is a problem at the storage

point at FamiCord laboratory, all laboratories (10) have mutual bankruptcy insurance and take the samples for storage if one laboratory goes bankrupt. The client has double insurance.

Others, such as Vita34 put the emphasis on the longevity of their bank stating it is the oldest private UCB bank in Europe, founded in Germany with many years of experience. It is with Vita34 and with the other that another guarantor of safety and reliability becomes the symbolic imagery of the countries where the biobanks are located and more important according to the laws of those countries.

We process and store in Germany. Vita 34 performs the entire process of preparation, processing, and long-term storage exclusively in Germany (original emphasis) in accordance with the most rigorous medical standards and German legislation.

While Vita34 overtly signals the symbolic linkages of the company with Germany, others also signal the symbolic importance of the location where the cord blood will be stored employing maps and national flags on their websites.

Biobanking and Good Parenting

Dear Parents. The letter P in this salute is deliberately capitalized. Although grammatically wrong, I want to point out to you

the importance of the role that stands before you. The role is lifelong and the most important in life...Stem cells from the umbilical cord are a unique treasure that is found in the umbilical cord, and it is only possible to collect it during childbirth and never again. If umbilical cord blood is not collected during childbirth, it, together with the placenta, becomes medical waste and disappears irreversibly. There are few things in life that can be both treasure and waste. The decision is yours alone! (Family Health)

The quote emphasizes the significance of good parenting and the role it plays in the context of children's health. The text specifically focuses on the value of stem cells from the umbilical cord, which can be collected during childbirth. The biobank highlights the uniqueness of these stem cells and stresses that they can only be obtained at that specific moment. If not collected, the umbilical cord and placenta become medical waste, irreversibly lost. This biobank's perspective suggests that good parenting involves recognizing the importance of preserving valuable resources for the health of children. By encouraging parents to make the decision to collect umbilical cord blood, the biobank implies that responsible parenting includes proactive measures to ensure potential health benefits for their children in the future. Moreover, the quoted text lacks sufficient context regarding the scientific validity and actual medical benefits of collecting and preserving umbilical cord blood. By simply referring to the stem cells as a "unique treasure," without providing concrete information on their potential uses in medical treatments, parents may make decisions based on incomplete or misleading information.

Conclusion

The possibility of storing cord blood in international banks can be seen as an example of global capitalism through which the local, intimate, and personal—such as human tissue and the relationship between parent and child—becomes global. These are for profit institutions which offer their facilities to store individual cord blood that can be used solely for the treatment of the account holder: that is, the infant and the infants from whom the blood was extracted. The logic of future health investment that is promoted by these biobanks globally personalized risk-management service—must be understood in particular local context by the consumers: the parents.

The risks that come in the late modernity, on the one hand, decline importance of the welfare state, and on the other increase importance of the market economy. Thus, the emergence of private biobanking in Serbia can be viewed through the lens of the continuing neoliberalization and privatization of healthcare in post socialist countries (Pantović, 2018, 2022) and the larger shift towards new regimes of care in the anthropocene (Pantović and Zarić, 2022). Biobanks converts anxieties of global social, economic, and political context of everyday life into individualized pathologies that one can insure against within the local context of a country undergoing drastic neoliberal transformations – Serbia.

According to the theories of late modernity (Bek, 2001), the organization of the life course in a modern society is flexible, and therefore there is an open-future perspective, that is, the possibility of choosing one's own life path. However, although modern society provides

individuals with more opportunities, at the same time it exposes them to numerous risks. According to Beck (Bek, 2001) the nature of these risks is different in comparison to risks in societies from previous epochs. Beck argues that uncertainties and challenges in modern society to the greatest extent come from the development of science and technology and that they are transferred to the sphere of everyday life and manifested in relations within families. New circumstances demand new responses from parents, and new forms of stability that would give them a sense of purpose. Issues with which parents face are consequences of a complex post-socialist transformation in Serbia, as well as of the processes of globalization and modernization (Tomano-*vić*, 2010). Modern parenting therefore includes combining the norm on the development of a child as an individual, i.e. autonomous person, with a need to protect children from the risks of modern society (*ibid.*). An important aspect for understanding the latter is the country's turbulent past, especially the 1999 NATO bombings that also had a toll on the population's (perceptions of) health. This past has led to the widespread perception that Serbian children are at elevated risk of cancer and that the Serbian health-care system cannot adequately manage the problem. In such a situation, social norms are placed on the responsibility of ("good") parents who should make informed decisions because the health of their children depends on them.

Our research is exactly focused on the examination of private umbilical cord biobanking in Serbia and its interconnections with biocapitalism and symbolic geographies of health, through the exploration of narratives about good parenting, in order to gain insight into this global phenomenon from a local perspective. Results show that parents, potential clients of biobanking, interpret living in Serbia as a health risk for children and believe that they can safeguard them

against through the usage of biobanking services. These attitudes are supported by biotech companies in their presentations, shaping the notions of good parenting in Serbia through emphasizing the investment and responsibilities of parents for a future wellbeing of their offspring. Keeping stem cells in Western Europe and outside of Serbia thus becomes a futuristic project that prescribes a new concept of the modern family, i.e. good parenthood, parental care and childcare.

This paper was realized with the support of the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, according to the Agreement on the realization and financing of scientific research.

References

- Appleby–Arnold, S. (2012). “It’s security, but not insurance”: Cord blood banking practices between bio–cosmopolitanism and gift economy. *Antropologija*, 12(3), 171–179.
- Bek, U. (2001). *Rizično društvo*. Filip Višnjić.
- Beltrame, L. (2020). Anchoring hopes in a regime of truth: The legitimizing use of cord blood–derived products in Italy. *New Genetics and Society*, 39(2), 173–190. <https://doi.org/10.1080/14636778.2020.1732971>
- Beltrame, L., & Hauskeller, C. (2019). Assets, commodities, and biosocialities: Multiple biovalues in hybrid biobanking practices. *Tecnoscienza: Italian Journal of Science and Technology Studies*, 9(2), 5–32.
- Ćeriman, J., & Vučković Juroš, T. (2023). From gender re–traditionalizations to anti–gender mobilizations: Care for family in Serbia and Croatia. *East European Politics and Societies and Cultures* (Forthcoming). <https://doi.org/10.1177/088832542311615>
- Ćeriman, J. (2021). Modeli rodne socijalizacije dečaka u porodicama u savremenoj Srbiji. *Etnoantropološki problemi*, 16(1), 245–270. <https://doi.org/10.21301/eap.v16i1.10>
- Ćeriman, J., Hrnjak, J., & Radoičić–Nedeljković, A. (2021). Institutional ethics of care in Serbia during the COVID–19 pandemic: A case study on the effects of the lockdown measures on girls and women trafficking survivors. *Philosophy and Society*, 33(4), 895–909. <https://doi.org/10.2298/FID2204895C>
- Cooper, M. E. (2011). *Life as surplus: Biotechnology and capitalism in the neoliberal era*. University of Washington Press.
- Gottweis, H., & Petersen, A. (2008). *Biobanks: Governance in comparative perspective*. Routledge.
- Hauskeller, C., & Beltrame, L. (2016). The hybrid bioeconomy of umbilical cord blood banking: Re–examining the narrative of opposition between public and private services. *BioSocieties*, 11(4), 415–434. <https://doi.org/10.1057/s41292-016-0017-1>

Helmreich, S. (2008). Species of biocapital. *Science as Culture*, 17(4), 463–478. <https://doi.org/10.1080/09505430802519251>

Jeong, Y. (2014). Scientific motherhood, responsibility, and hope: Umbilical cord blood banking in South Korea. *New Genetics and Society*, 33(4), 349–369. <https://doi.org/10.1080/14636778.2014.946001>

Katz, G., Mills, A., Garcia, J., Hooper, K., McGuckin, C., Platz, A., Rebull, P., Salvaterra, E., Schmidt, A. H., & Torrabadella, M. (2011). Banking cord blood stem cells: attitude and knowledge of pregnant women in five European countries. *Transfusion*, 51(3), 578–586. <https://doi.org/10.1111/j.1537-2995.2010.02954.x>

Krajnović, D., Stojković, T., & Babić, S. (2016). Ethical aspects of the stem cells' use as healthcare technologies in personalized medicine. *Engrami*, 38(1), 5–18. <http://dx.doi.org/10.5937/engrami1601005K>

Krstić, M. (2012). Zdrava porodica i uspešno roditeljstvo: Antropološka analiza internet medija o čuvanju matičnih ćelija u Srbiji. *Antropologija*, 12(3), 139–170.

Latour, B. (1993). *The pasteurization of France*. Harvard University Press.

Mitchell, R., & Waldby, C. (2010). National biobanks: Clinical labor, risk production, and the creation of biovalue. *Science, Technology, and Human Values*, 35(3), 330–355. <https://doi.org/10.1177/0162243909340267>

Milutinović Bojanić, S., Čeriman, J., & Pavić Zentner, V. (Eds.). (2016). *Siromaštvo, ruralnost, rod: istraživanje funkcionisanja sistema socijalne zaštite u ruralnim oblastima Srbije*. Centar za etiku, pravo i primenjenu filozofiju; Institut za filozofiju i društvenu teoriju Univerziteta u Beogradu.

Milutinović Bojanić, S., & Čeriman, J. (Eds.). (2016). *Ruralnost i rod: dostupnost i prepoznavanje programa socijalne zaštite*. Centar za etiku, pravo i primenjenu filozofiju.

Murata, M., Kawabe, K., Hatta, T., Maeda, S., & Fujita, M. (2022). Current status of umbilical cord blood storage and provision to private biobanks by institutions handling childbirth in Japan. *BMC Medical Ethics*, 23(1), 1–7. <https://doi.org/10.1186/s12910-022-00762-3>

Pantović, Lj. (2016). Buying a connection: Private practice in public health care—the case of a Serbian maternity hospital. *Anthropology of East Europe Review*, 34(1), 25–37.

Pantović, Lj. (2018). Not-so-informal relationships: Selective unbundling of maternal care and the reconfigurations of patient-provider relations in Serbia. *Comparative Southeast European Studies*, 66(3), 371–391. <https://doi.org/10.1515/soeu-2018-0030>

Pantović, Lj. (2019). *Private within the public: Negotiating birth in Serbia* (Doctoral dissertation). University of Pittsburgh. <https://d-scholarship.pitt.edu/35994/>

Pantović, Lj. (2022). Providing a service or seeking a favor? The role of private prenatal care on the continuity of care in Serbian public maternity hospitals. *Birth*, 49(3), 506–513.

Pantović, Lj., & Zarić, Z. (2022). Care in the Anthropocene: Acting with compassion. *Filozofija i društvo/Philosophy and Society*, 33(4), 881–894.

Sunder Rajan, K. (2006). *Biocapital: The constitution of postgenomic life*. Duke University Press.

Tomanović, S. (2017). Roditeljstvo između familizma i individualizacije: primer Srbije. In S. Ignjatović & A. Bošković (Eds.), *Individualizam* (pp. 162–181). Institut društvenih nauka.

Tomanović, S. (2010). Odlike roditeljstva. In A. Milić, S. Tomanović, M. Ljubičić, N. Sekulić, M. Bobić, V. Miletić–Stepanović, & D. Stanojević, *Vreme porodica: sociološka studija o porodičnoj transformaciji u savremenoj Srbiji* (pp. 177–194). Institut za sociološka istraživanja Filozofskog fakulteta; Čigoja štampa.

Velikonja, N. K., Erjavec, K., & Knežević, M. (2021). Knowledge, awareness, and attitudes toward umbilical cord blood biobanking. *Open Access Macedonian Journal of Medical Sciences*, 9(E), 702–709.

Venkatesan, P., & Peters, M. A. (2010). Biocapitalism and the politics of life. *Geopolitics, History, and International Relations*, 2(2), 100–122.

Waldby, C. (2002). Stem cells, tissue cultures, and the production of biovalue. *Health*, 6(3), 305–323.

Waldby, C. (2006). Umbilical cord blood: From social gift to venture capital. *BioSocieties*, 1(1), 55–70. <https://doi.org/10.1017/S1745855206060040>

Part III

Democracy, Security and Technology

Standardization for High Impact Technologies: Why Citizens Should Care

Claudia Roda
Rebecca Stone
Antonio Kung

Introduction

Standardization efforts have been vital to technological progress for over a century and their societal impact is complex and multi-form (Timmermans and Epstein, 2010). Today, organizations like the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC), and the Institute of Electrical and Electronics Engineers (IEEE) continue to play an essential role in developing standards for a vast array of technologies, from software and hardware to artificial intelligence and the Internet of Things. As these technologies increasingly require that standardization embraces the complete socio-technical system of their development and deployment, standard setting objectives have moved well past

Claudia Roda, Full Professor, American University of Paris, Paris

ORCID: 0009-0009-1054-8049, email: croda@aup.edu

Rebecca Stone, American University of Paris, Paris

ORCID: 0000-0002-5275-9890, email: rebecca.johnson.stone@gmail.com

Antonio Kung, Trialog, Paris ORCID: 0000-0003-3809-8325, email:

antonio.kung@trialog.com

DOI: <https://doi.org/10.5281/zenodo.14204383>

the original objective of interoperability, aiming at improving quality and safety, promoting stakeholder communication and collaboration, protecting consumers, and other such objectives.

For high impact digital technology, this enhanced role of standardization plays an important social role as it may influence accessibility, safety and, the level of control end-users (including governments) have on technology. While the issue of technology governance as a site of economic and political power, and often injustice, has been widely discussed (e.g., Bradshaw and DeNardis, 2019; DeNardis and Musiani, 2014; McIlwain, 2017; Paris, 2021), our analysis focuses on standardization processes and, in particular, the processes established through the work of official Standard Setting Organizations (SSO¹, e.g., ISO, IEC, IEEE, ITU, ETSI). The development process for these standards is different from standards that emerge from consortia or informal interest groups, or that are established thanks to the widespread adoption on the market of a particular product (*de facto* standards), because it follows clearly articulated governance processes accepted globally. Distinct from the vast literature about standards' support for interoperability, innovation, and trade competition (e.g., Baron and Kanevskaia, 2021; Blind et al., 2011; Farrell and Saloner, 1985; Shapiro and Varian, 1999; Tasse, 2000), this paper considers the social impact of standards and asks whether and how, standard setting may act as an instrument that reinforces societal injustice and bias or may instead foster citizens' protection and empowerment. The document is organized as follows: we first briefly describe what standardization is and how it is governed (section II).

¹ Standard Setting Organizations are alternatively referred to as Standard Developing Organizations (SDO)

We then discuss how standardizations and legislation have become increasingly entangled and propose that, given the important role of standardization in our interconnect world and the significant impact that technology has on society, we should aim to implement *principled standardization processes* (section III). The following two sections are used to exemplify what such principled processes would require. In section IV, we consider the scenario of a hypothetical technology—involving a digital twin of an electronic security system, which incorporates a body-worn camera—and its use in law enforcement. Section V discusses the challenges involved in standardizing such a technology by following principled process. In section VI we highlight the optimal circumstances for principled standard development and the actions that may need to be taken to achieve it. A list of abbreviations is provided after the conclusions.

Standardization Processes

Standardization bodies highlight standards' crucial input for achieving innovative outcomes including in the case of emerging digital technologies with high societal impact. Modern standardization is a relatively new instrument in the array of processes societies have been using to establish rules and guidelines on how to live together. While the standardization of weights and measures dates back thousands of years, standardization processes, as we know them now, are more recent and took their current structure roughly at the time of the industrial revolution. It is in the eighteenth century, with the development of industrial machines, that the need arose to have

interchangeable parts and therefore to agree between different makers, how these could be assembled. As it still often happens, standards would develop within a company, then spread to the industry, to be eventually adopted at the national or international level. In the nineteenth century, as the number of companies and their competition increased, spontaneous or de facto standardization processes became more cumbersome and the need for institutions capable of encouraging and steering standardization efforts became clear. The first national standards body was established in London in 1901: the Engineering Standards Committee, which would later become the British Standards Institution (BSI, n.d.). Since then, many other national and international standardization bodies have been established. Some of them focus on specific technologies (e.g., the World Wide Web Consortium, also known as W3C, for the development of Web standards), others, like the ISO, span a vast variety of products, technologies, and methodologies.

The ISO compares standards to “formulas that describe the best way of doing something [whether] making a product, managing a process, delivering a service or supplying material.” (ISO, n.d.) Standardization can be described as a voluntary process that brings together a range of stakeholders to create consensus over the object of the standard. Standards have no legal authority other than references in laws, contracts, or procurement processes and are not enforced other than through these references because they are voluntary. Given this definition, standards may appear to have little social and economic consequence but, as we will discuss in the next section, their impact on trade, the economy and society is extensive and often underestimated.

SSOs develop standards by first establishing technical committees and working groups—which normally include engineers, researchers,

and often public officials. These technical committees and working groups put forth proposals for the standard to be developed. Proposals undergo discussions, adjustments, compromises, and eventually a final version of the standard is agreed upon. This process, ideally transparent, can extend over a period ranging from a few months to several years. Decision-making can involve majority votes or consensus-based agreements.

SSOs may work at the national, regional, or international level. International SSOs play a key role in developing and promoting standards that have global relevance and applicability. Examples of international SSOs include the ISO, the IEC, and the International Telecommunication Union (ITU). International SSOs provide a platform for collaboration among participating countries, industry experts, and other stakeholders from around the world. Their technical committees and working groups are composed of representatives from national standards bodies, industry associations, academia, and other relevant entities. National and regional² SSOs, on the other hand, are responsible for standards development and adoption within their state or region. Each country typically has its own national standards body, such as the American National Standards Institute (ANSI) in the United States or the BSI in the United Kingdom. National and regional SSOs participate in the work of international SSOs and contribute to the development of international standards by providing expertise, input, and feedback from the perspective of their constituencies. Furthermore, national and regional SSOs often adopt international standards

2 Such as the European Committee for Standardization (CEN) in Europe, the Pacific Area Standards Congress (PASC) in the Asia-Pacific region, and the Pan American Standards Commission (COPANT) in the Americas

as their own or use them as a basis for developing their own standards. They may also provide guidance on the implementation and enforcement of international standards at the national or regional level. International, regional, and national SSOs are expected to maintain a working relationship to ensure harmonization and consistency in standardization efforts. Indeed, they often exchange information, share best practices, and collaborate on matters such as standards development, conformity assessment, and capacity building. This relationship helps foster international cooperation, promote interoperability, and facilitate trade by establishing a globally recognized framework for standards and conformity assessment. Governance rules and guidelines for national, regional, and international SSOs are set by SSOs themselves or by governments and international organizations. For example, while the World Trade Organization (WTO) does not directly engage in standard setting, it plays a significant role in relation to standards through its agreements and principles that govern international trade, also providing a mechanism for resolving trade disputes between member countries. In particular, the WTO Agreement on Technical Barriers to Trade (WTO, 1995) addresses the impact of technical regulations, standards, and conformity assessment procedures on international trade and establishes principles for standard development.

Overall, standards have become an essential component of technology governance and adoption that, in addition to supporting trade and the safe use of technology, often helps companies in enhancing internal processes, facilitating communication, bringing the attention of higher management to potential risks, and staying updated with industry best practices.

The Relationship between Standards and Legislation

This section discusses some of the consequences of standardizations and legislation becoming increasingly entangled as policy makers pursue transparency on technical requirements, buy-in from industry stakeholders, and removal of trade barriers.

The activity of SSOs, nowadays, extends to aspects of technology that go far beyond ensuring interoperability. Modern standards influence a wide range of technology-related organizational activities including governance, consumer protection, social responsibility, and innovation. Modern standards, also play a crucial role in the global economy addressing issues that cannot be tackled within the governing structure of individual States. As new technologies increasingly impact social systems (the question of privacy is a good example of the impact of technology on society) and standards have moved from the realm of technical interoperability to also include issues of corporate or organizational positioning within the social structure, the question arises of the relationship between standardization processes and public policy. Standards, in fact, are regularly used both in non-legislative³ and legislative⁴ acts. For example, compliance with certain standards may be a prerequisite for access to public grants, or standards may be referenced in law. With respect to use in legislative acts,

3 In a guide addressed to policy makers, the ISO/IEC provide a synthetic explanation of the use of standards to support public policy (ISO/IEC, 2015).

4 For a short but complete explanation of how standards come into play in several major domains of law, see (Coglianese, 2023)

references to standards in law have grown from occasional to systematic⁵. In 2019, “federal regulations contain[ed] more than 23,000 incorporations of private standards by reference” (Bremer, 2019: 316–317). In 2023, the same database⁶ contains 26,525 references.

The United Nations Industrial Development Organization (UNIDO) highlights the “*circular relationship between standards, regulations and policies, which feed into an overall concept/public good of good governance. Standards are a voluntary complement to regulation, which have the effect of enhancing efficiency and productivity. These standards inform effective regulations, which can create an enabling environment for innovation and minimize risk for disruptors and investors*” (UNIDO, 2021: 25). Additionally, standards may be used to demonstrate compliance to laws. A case in point is the European Union where “*over the last 30 years, the European Standardisation System has delivered more than 3600 harmonised standards allowing companies to demonstrate compliance with EU law*” (European Commission, 2022). Within the EU, a *harmonised standard* is a standard developed by a recognized European Standards

-
- ⁵ An early example of reference in law, within the consumer space, is the DIN 13164 standard referred to in a 1960s German traffic law mandating first-aid boxes in cars. In the US, during the 1970s, there was a shift in federal policy, towards promoting the use of standards produced by SSOs, discouraging the development of standards by government agencies (Bremer, 2013); since 1995 agencies are required to use existing SSO’s standards unless they are inconsistent with applicable law or otherwise impractical.
- ⁶ <https://www.nist.gov/standardsgov/what-we-do/federal-policy-standards/sibr-database>

Organization⁷ that can be used by organizations to demonstrate that products, services, or processes comply with relevant EU legislation (European Commission, n.d.). In fact, within the EU, the relationship between *harmonized* standards and law is much stronger. In a landmark, and widely discussed, case⁸ delivered by the Court of Justice of the European Union in 2016 *harmonized* standards have been essentially recognized as part of Union law (van Waeyenberge and Amariles, 2017). This convergence, on one hand worries the industry because of the stronger constraints imposed by a legal interpretation of standards (Bertuzzi, 2023), on the other hand, worries civil society organizations because of the possible increased power of private organizations on the legal system (Degallaix, 2018).

At the international level, the significant impact that standards may have on legislations is underscored by the fact that, under the Technical Barriers to Trade (TBT) Agreement, “*Where technical regulations are required and relevant international standards exist or their completion is imminent, Members shall use them, [...] as a basis for their technical regulations except when such international standards [...] would be an ineffective or inappropriate means for the fulfilment of the legitimate objectives pursued. [and] Whenever a technical regulation is prepared, adopted or applied [...] in accordance with relevant international standards, it shall be rebuttably presumed not to create an unnecessary obstacle to international trade.*” (WTO, 1995, Articles

7 CEN (European Committee for Standardization), CENELEC (European Committee for Electrotechnical Standardization), or ETSI (European Telecommunications Standards Institute) are the European Standard Organizations

8 James Elliot Construction case (*James Elliott Construction Limited v Irish Asphalt Limited*, 2016)

2.4 and 2.5, *emphasis added*). Consequently, national measures that impose more stringent protection for citizens may have to be abandoned because they risk being considered as violating trade obligations if challenged before a panel or the Appellate Body. Therefore, under the TBT framework, standards initially designed to enhance product quality and facilitate trade, often are instead used to dissuade countries from implementing more stringent protective policies (Fontanelli, 2011).

The strong connection between standards and regulations has many advantages as it facilitates policy making, supports innovation, and helps trade. However, standards are often developed by private organizations with objectives that are primary commercial, while laws and regulations typically have different primary objectives, including the protection of citizens, social justice, and in general the good functioning of the local, national, or regional legislatures. Given this situation, one could ask whether standards are, in fact, a form of “hidden” regulation that fails to go through proper democratic channels and is adopted into law thanks to the recognized status of SSOs. This question is beyond the scope of this paper but, the fact that standards do have a significant impact on legislation prompts our exploration of the conditions necessary for the development of *principled standardization processes*, i.e., processes that are sufficiently inclusive and transparent to ensure that their impact on legislation will not jeopardize the basic tenets of law-making.

Of course, recommendations and requirements are already in place to guide standardization processes. Examples include the ANSI’s Due process requirements for American National Standards (ANSI, 2022), the regulation on European standardization (Regulation EU 1025/2012,

2012) and the WTO's Principles for the Development of International Standards (WTO, 2000) based on the Agreement on Technical Barriers to Trade (WTO, 1995) and its reviews and updates. Many SDOs follow principles similar to the six WTO's principles: transparency, openness, impartiality and consensus, effectiveness and relevance, coherence, and development. *Transparency* requires that essential information about work programs and standards (whether being developed or finalized) are made accessible to interested parties. *Openness* requires that participation in standard development is open to relevant bodies in a non-discriminatory manner. *Impartiality and consensus* require that the standard development processes don't privilege or favor the interests of particular suppliers, countries or regions. *Effectiveness and relevance* require that standards respond to regulatory and market needs, as well as scientific and technological developments without distorting the global market or giving preference to specific countries or regions. *Coherence* requires avoiding the development of conflicting or duplicate international standards. *Development* requires the effective participation of developing countries in standardizations processes. However, these principles do not seem to be able to prevent serious issues connected to lack of transparency in standard development (Mavroidis and Wolfe, 2017), limited citizens' access to laws that include references to standards often only accessible at significant costs (Bremer, 2019)⁹, uneven distribution of voting power favoring corporate interests, and poor representation of societal stakeholders (European Commission, 2022).

⁹ For the point of view of US SSOs (which is similar to that of many other SSOs) on why standards continue to be behind a paywall, even when referenced into law, see (ANSI, 2011)

As standards seamlessly migrate into law (e.g., in the EU system) and effectively limit the ability of countries to establish more stringent regulations (as it happens under the WTO's TBT agreement), then standardization processes should ensure that the rights of all citizens are respected. In fact, the limited power of States and regions, when confronted with established standards, seems to motivate the need expressed by the European Commission in its recent Strategy on Standardisation document, which notes: "*More than ever, standards do not only have to deal with technical components, but also incorporate core EU democratic values and interests, as well as green and social principles*" (European Commission, 2022).

In order to analyze some of the challenges of such *principled standardization* we consider the standardization of a hypothetical technology and a scenario of its use in law enforcement. The scenario is particularly interesting because, especially in recent times, the adoption of technologies by law enforcement agencies has been the subject of many controversies, and the use of standards in law enforcement regulations¹⁰ makes it easy to identify societal challenges of standard definition.

In section IV we present the technology—a digital twin for an electronic security system that incorporates a body-worn camera—and

10 Examples of the use of standards in law enforcement include the British National Police Chiefs' Council guidelines on police response to electronic security systems (NPCC, 2022), which contains references to over ten British standards (BS), four European standards (BS-EN), and one ISO standard (BS-EN-ISO). As another example, UK and EU forensic service providers are required to comply with EN ISO/IEC 17025 "General requirements for the competence of testing and calibration laboratories" (EC OJ L 322, 2009; UK Secretary of State, 2018)

the scenario in which such technology is used in law enforcement. In the following section V, we consider the challenges associated with the standardization process.

Digital Twin for Law Enforcement Body- Worn Cameras

In their most basic definition, digital twins (DTs) are “*digital informational construct about a physical system, created as an entity on its own and linked with the physical system in question.*” (Kritzinger et al., 2018, p. 1016). According to Kritzinger and colleagues, a defining characteristic of DTs is a bidirectional data flow between the physical object and the digital object so that the digital and physical objects might act as controlling instances of one another. DTs allow to record, control, and analyze real-time situations¹¹. In our case the physical system is a security system composed of a Body-Worn Camera (the focus of our analysis) plus other elements such a car, a weapon, etc.

Body-worn cameras (BWCs) are wearable devices, typically affixed to the front of security operators’ clothing or uniforms, enabling the recording of visual and audio data. They have the capacity to capture footage and sound in close proximity to individuals during various face-to-face interactions, e.g., during traffic stops, on public transportation, or within private premises like homes and shops.

¹¹ An important application of DT is the simulations of physical systems, however, here we will not discuss the use of DT for simulation

In the past few years, there has been a push for law enforcement agencies to adopt BWCs, sometimes with little or insufficient effort to establish policies regarding their usage. In the US a \$75 million program initiated by former President Obama, led to the distribution of 50,000 BWCs in locations throughout the country (Wright and Brown, 2020). In Australia, the Queensland Police Service were already deploying 2,700 units by 2017 and numbers have skyrocketed around the country since then (Murphy and Estcourt, 2020). After an experiment conducted in 2016 for a period of two years (Ministère de l'Intérieur, 2018), in 2021, the French Ministry of the Interior deployed 30,000 BWC to the National Police and Gendarmerie with an investment valued at €15 million (securityWorldMarket.com, 2021). In 2023 the Royal Canadian Mounted Police started the deployment of 10,000–15,000 BWCs to contract and federal police officers engaged in community interactions throughout Canada's rural, urban, and remote areas (Royal Canadian Mounted Police, 2023b). These are just a few examples showing the widespread adoption of BWCs. BWCs are often considered as capable of enhancing investigations, crime prevention, and public safety, although their use, especially if not guided by sound policies, raises several concerns.

For our purpose of exemplifying the challenges and opportunities in the standardization of technology with high societal impact, it will suffice to concentrate on the basic use of BWCs by law enforcement, that is, the recording and storage of video footage. It should be noted, however, that BWCs can now be integrated with real-time facial recognition technology, or other biometric analysis, which enable the automatic identification of citizens present on the scene of the recording. Other tools exploit the large amount of BWCs footage for specific discoveries (e.g., citizens' networks and associations)

or predictive policing (Haskins, 2019). These more advanced applications of BWCs are already in use in several jurisdictions and continue to raise significant concerns about systemic and biased inaccuracies, surveillance, and the consequences that their use may have on our democracies and the respect of human rights (Dodd, 2023; Doffman, 2019; Gataveckaitė, 2023; Privacy International, 2029; Ringrose, 2019; Turner Lee and Chin, 2022; Williams and Kind, 2019). We choose here, to contain our use case to the basic use of BWCs because this gives us a chance to discuss the characteristics of standardization processes on a well-established technology supported by the still evolving technology of digital twins.

In the scenario of figure 1, the physical system is composed of a law enforcement agent (LEA), their BWC, and car (LEA Vehicle). This system is coupled with a DT monitoring the state of the physical system and communicating with the police headquarters. The DT keeps a record of the state of the physical system by receiving real-time information such as battery levels, activation/deactivation of devices, information transmitted, movement, etc. Based on this information, the DT verifies the correct and compliant functioning of the physical system and, in case of problems, may intervene on it (e.g., activating devices) and/or communicate with the police headquarters.

The scenario starts with the headquarters requesting the LEA to intervene at a crime location, the request is registered by the DT. The LEA activates the light bar and siren; the vehicle informs the DT of this activation; it also provides it with continuous location information. Once the crime scene is reached, the agent opens the car door. The information about the changed status of the door is communicated to the DT that informs headquarters of the arrival of the LEA

on the scene and sends a signal to the BWC to activate it, informing headquarters of the activation. As the agent manually deactivates the camera, the DT is informed. The DT detects a Possible Protocol Violation and communicates it to headquarters. Headquarters confirms, to the DT, the Protocol Violation. The DT informs the LEA of the protocol violation and reactivates the BWC locking it in the active state. The scenario continues through other possible events such as the LEA closing the car door, opening the trunk, etc.

As shown by the example, DTs in law enforcement could serve systems and subsystems related to camera surveillance, equipment tracking, and data management of equipment and actions providing an avenue for transparency, accountability, and oversight. In turn, this is considered to bolster public trust and safety through enhanced monitoring and analysis of potential incidents by perpetrators and law enforcement alike. The DT could also manage real-time alerts for camera malfunctions or lost equipment. These provisions allow analytics and reporting measurements to improve police and law enforcement operations and decision-making based on data. Resource optimization would also be possible through improved allocation and deployment of personnel for more effective and efficient law enforcement actions.

Standardizing a system of this type presents several challenges ranging from the governance of the system and its data to technical specifications. These are discussed in the next section.

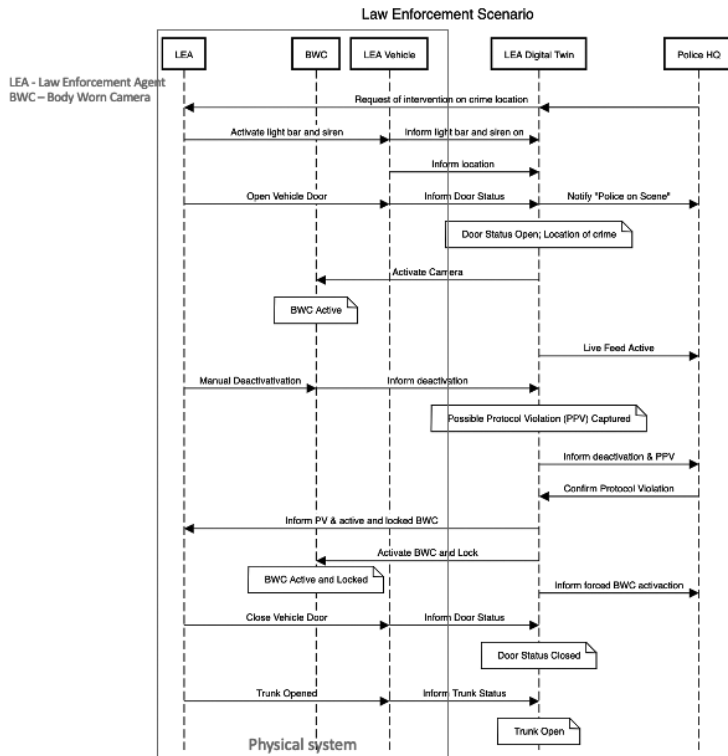


Figure 1: Sample Scenario for Digital Twin of Law Enforcement Agent with Body Worn Camera

Standardizing DT for Law Enforcement BWC

DTs have the potential to support numerous industries, from manufacturing to healthcare and public safety (Attaran and Celik, 2023), but their implementation still presents several challenges (Sharma et

al., 2022) and their use raises concerns about data privacy and security (Holmes et al., 2021; Wang et al., 2023). Lawful basis for the collection of personal identifiable information (e.g., camera images, ID information, location information) must be dynamically established while the distributed processing nature of the system may pose additional issues of data transfer, including the definition of controls and possibly contracts delimiting data usages. DT may introduce significant security challenges as the DT itself may represent a point of vulnerability, and exchanges between the physical system and DT must be secured and monitored to ensure that the states of the physical system and its DT are consistent. Additionally, information may be much more easily accessed and used by a malicious third party, from the DT representation of the physical system than from the physical system itself.

Ethical issues have also been discussed with respect to DT applications in healthcare system (Huang et al., 2022) and an issue raised by Mittelstadt's appears particularly relevant also to our use case. Mittelstadt argues that DT of patients "*will undermine patient's authority and autonomy in the doctor-patient relationship by providing an alternative, more 'objective' account of the patient's health*" (Mittelstadt, 2021, p. 1). In our scenario, the DT records the LEA's actions creating a narrative about the events at the crime scene that is normally considered to be objective. This narrative, however, may not coincide with the LEA's narrative. Similarly to the healthcare patient situation, the DT's narrative may undermine the agent's authority and autonomy in the citizens-agent relationship by providing an alternative, more 'objective' account of agents' behavior. In other words, just like the DT of a patient is considered more capable of giving a reliable account of a patient's health, so the DT of a security system is considered to give a more reliable account of the agent's behavior. While this may be the

case in some, or even most, situations, there may be other situations (e.g., something happening outside of the view of the camera) when the officer would be better in recounting the events and their own behavior. It is unclear whether these situations would be easy to identify.

These privacy, security, and ethical issues make developing and implementing standards for DT challenging. Existing standards include the ISO 23247 series (ISO 23247, 2021) on DT in manufacturing, and the IPC 2551 for DT product, manufacturing, and lifecycle frameworks. However, the public descriptions of these standards do not indicate that they address privacy or other policy, and ethical issues related to their application in a variety of domains.

BWC, on the other hand, are a more mature technology than digital twins. The BSI has developed, in collaboration with security and privacy groups—including the Home Office, the Metropolitan Police, and Big Brother Watch—a standard for the deployment and use of BWC, BS 8593:2017 (BSI, 2017). However, in many countries, regulations and guidance, helping to interpret how specific legislative frameworks may apply to BWC use, is not well established and heterogeneous. In Europe, the Law Enforcement Directive (LED) (European Parliament, 2016) establishes regulations for safeguarding personal data in the case of processing activities aimed at crime prevention, investigation, detection, or prosecution. While national laws are responsible for providing additional specifications and details, the LED promotes consistent implementation of data protection principles in the realm of law enforcement throughout the European Union. Under articles 13–17 of the LED data subjects have the right to be informed about the processing of their personal data, access their data, and have their personal data rectified or erased, albeit with great limitations

as compared to the GDPR¹² and often only through the competent supervisory authority. Other rights in the LED include the right not to be subject to decisions based solely on automated processing and to obtain human intervention in case such decisions are authorized by law (CNIL, 2021; European Parliament, 2016; Quezada-Tavárez, 2021). Similar high-level data protection laws, ensuring various degrees of safeguard, exist in several countries across the world. Following the fast adoption of BWCs described in the previous section, specific directives and guidelines for law enforcement use of BWCs are being developed. Examples of existing regulations include the policy statement issued by the US Department of Homeland Security (2023), the Model Governance Framework for Police Body-worn Camera Programs in Ontario (IPC Ontario, 2021), the operational manual issued by the Royal Canadian Mounted Police (2023a), and France's application decree supplementing the law, passed in 2018, permitting the use of BWC by the municipal police (Légifrance, 2019).

Given the fast-paced evolution and adoption of technologies and regulations exemplified by this scenario, we can derive that *principled standardization processes* should not only enable compliance with existing regulations, but also ensure that a dialogue is established amongst all stakeholders that may be affected by the standard' impact on future regulations, including technology experts, regulators, civil society organizations, representative from industry (including SMEs) and academia. Additionally, given the different

¹² The General Data Protection Regulation (GDPR) protects citizens privacy encompassing all forms of personal data processing within the Member States, including both public and private sectors. However, it does not extend to processing activities that fall outside the purview of European Union law such as state security or national defense activities.

levels of knowledge of the technology being standardized amongst such different set of stakeholders, dissemination and educational efforts should be established prior, during, and after the development of standards to support active and effective participation in standards development and encourage widespread uptake and correct application.

Our scenario also highlights an issue that often accompanies the introduction of new technologies: the uncertainty around their effectiveness (in our case, uncertainty about the effectiveness of police use of BWCs). Research studies report mixed evidence that BWCs produce the expected benefits such as reducing problem behavior on the side of officers or citizens, or the rate of assaults against police (Davies and Krame, 2023; Ferrazares, 2023; Lum et al., 2020). Unclear is also the effect that these devices may have on officers' work environment, health, and safety, especially when BWCs are used with the objective of monitoring police behavior (Adams and Mastracci, 2019; Hansen Löfstrand and Backman, 2021). In the presence of these concerns and uncertainties, one should first ask whether standardization of the technology should be considered at all. In the case of BWCs, however, the uncertain outcomes of their use have not prevented their wide adoption by law enforcement agencies worldwide. Therefore, unless new legislations prohibit their use, or law enforcement agencies are discouraged from implementing the technology, an effort in standardization seems appropriate. The case study highlights the need for *principled standardization processes* to facilitate knowledge sharing and collaboration among different jurisdictions and agencies so that they can share best practices and lessons learned, with regards to the conditions enabling effective and appropriate use of the technology (BWCs in our case). Empirical research studies report that

establishing and enforcing clear policies about the situations when BWCs should be activated, is a necessary condition for BWCs effectiveness (Ariel et al., 2017; Katz and Huff, 2023; Lum et al., 2020). All research agrees that, when used, BWCs should record the entirety of officers' interaction with the public to avoid missing possibly essential aspects of the encounter. This requires considering how to best balance officers' discretion and their accountability; the questions are, what are the situations when officers should exercise discretion? And can the technology support different policies with respect to when and how BWCs should be used? Burke finds that practicing lawyers view the introduction of police BWCs as a substantial change with wide-ranging effects. For example, depending on how they are used, cameras can serve as a valuable tool in assessing the strength of prosecution cases, leading to early guilty pleas or supporting the defendant's account of events (Burke, 2020).

There are two implications for the definition of *principled standardization processes*. First, the technical system should allow for enough flexibility to accommodate different policies, some of which, may require enforcement "by design"¹³. Second, standard committees should include, in addition to technical experts, representatives of police officers, of the policy makers responsible for setting policies for BWCs use, and of practicing lawyers who can provide valuable insights on their most effective use in legal cases. Additionally, given the impact that law enforcement use of BWCs may have on citizens, and minority residents in particular (Pyo, 2022), civil society organizations representing this population should also be included.

¹³ "by design" is intended as inbuilt in the system workings (an extreme example would be a BWC that can never be turned off).

Another issue recurring with the design and adoption of new technology is related to the consequences it may have on people's privacy and control over their data. In the case of BWCs, Newell notes that, as part of their standard procedures, officers often request personal identifiable information such as name, dates of birth, driver's license numbers, etc., and they may enter private residences and other personal spaces. With the introduction of police BWCs, these interactions and associated information are recorded in video footage, which, in certain US states, can be accessed by the public under open-records laws (Newell, 2021). This is obviously an infringement on the privacy of citizens who appear in the recording, whether as part of the main scene, or bystanders. Common notions of privacy, and most privacy legislations, require that these recordings are not accessible to the public. On the other hand, it is important, for transparency and accountability, that subjects, whose voices or images are on the video, have access to relevant parts of the recording (Quezada-Tavárez, 2021) and, as we have seen, some regulations enable, in restricted form, such access. Regulations may also establish different time limits for the storage of the recordings under different circumstances (e.g., if the recording is being used for an enquiry or as evidence in a trial) or legislations. Finally, technical decisions related to the secure transmission and storage of videos captures and associate information will be important to appropriately support compliant access.

This implies that *principled standardization processes* should enable conversations between technical experts, legal experts, civil society organizations, and interested citizens to ensure that technological solutions provide appropriate support for data protection, data access (possibly to edited versions of the video to protect the privacy of other citizens appearing in the recording), data transmission, and

data end of life. For example, the governance framework for BWCs proposed by the Office of the Information and Privacy Commissioner of Ontario (IPC Ontario, 2021), includes guidelines for all the areas above and was created in collaboration with a variety of stakeholders including the Toronto Police Services Board, the Ontario Human Rights Commission, the Canadian Civil Liberties Association, the Ontario Association of Chiefs of Police, members of the Citizen Lab, as well as taking into account a public consultation.

Finally, in their research Wright and Brown find that legislators regulating BWCs *“should engage in more collaborative efforts that bridge the gap between the police and the community. Similar groups, such as: marginalized racial groups and classes as well as ex-offenders, may be able to identify ways to make the same transition if community activists can go from traditionally negatively socially constructed to more positive social construction”* (Wright and Brown, 2020, pp. 477–478). Similar conditions apply to our case study and to standardization processes in general. These should be designed to bridge gaps existing between a variety of stakeholders, in our case, these include technical experts, regulators, representatives of police officers, practicing lawyers, legal experts, interested citizens, and importantly, civil society organizations representing citizens in general and, in particular, vulnerable groups.

Optimal Circumstances for Principled Standard Development

We have seen that standardization has a clear impact on our societies and, under some circumstances, this impact goes as far as standards

having the same power as law. Whether this convergence between standardization and legislation is desirable or necessary is an open question. However, even in less extreme cases, standards do influence how technology will be used and consequently how it will impact citizens. Through a case study, we have highlighted the challenges of standardization processes and how they may be addressed within the context of the specific case. From this analysis we now highlight some principles that, along with those commonly established and briefly presented in section II, would create optimal circumstances for *principled standard development*.

Foresight: A first takeaway from our case study is that standardization processes for innovative, high impact, technologies should not only enable compliance with existing regulations, but also anticipate societal needs and ethical principles that will be likely addressed in future regulations¹⁴. Standardization committees should therefore include regulators as well as other stakeholders capable of addressing those societal needs and ethical principles.

Representativeness: A second point highlighted by the case study is that the activities of committees ought to foster knowledge exchange, active participation, and collaboration among stakeholders who have different perspectives on, and perhaps objectives for, the technology being standardized, so to enable the sharing of best practices and experiences related to its effective and ethical use. This requires representatives from a variety of groups such as: technical experts, end-users, policy makers, professionals, and citizens who will

¹⁴ Standardizations for artificial intelligence-based technologies, or technologies aimed at supporting the ecological transition will certainly require this type of foresight.

likely be impacted by the technology. Special attention should be given to including currently under-represented organizations and populations, these may be, for example, small and medium-sized enterprises (SMEs), vulnerable populations (e.g., people with disabilities), or experts in the ecological transition. Appropriate representativeness appears to be the objective of the WTO principle of openness. According to the WTO, *openness* requires that “*Membership of an international standardizing body should be open on a nondiscriminatory basis to relevant bodies of at least all WTO Members.*” (WTO, 2000: 2) However the WTO also specifies that countries adhering to the code of good practice (Annex 3 of the TBT Agreement) should participate through a single delegation representing all standardizing bodies in the territory. This may represent a significant restriction if several, diverse perspectives on the technology are to be represented. The definition of openness provided by national and regional SSOs, such as by ANSI¹⁵ (2022) or by the regulation on European standardization¹⁶ (Regulation EU 1025/2012, 2012) are much better suited to support wide representativity as they do not impose similar restrictions enabling experts in technical specifications to work hand-in-hand with policy and ethics experts. This is more likely to result in enough flexibility to accommodate different policies and level of enforcement, ensuring that, not only technical, but also social requirements can be integrated “by design” and the entire life cycle of the technology is supported,

15 The openness criteria are defined as: “Participation shall be open to all parties who are directly and materially interested in the activity in question. There shall be no undue financial barriers to participation.”

16 The openness criteria are defined as: “the technical specifications were developed on the basis of open decision-making accessible to all interested parties in the market or markets affected by those technical specifications;”

through appropriate data sourcing, data security and privacy, correct levels of access, control of data transmission, and data end of life.

Sustaining wider representativity is often challenging (Boström and Hallström, 2013) and requires actions to ensure that weaker stakeholders can sustain participation and that agreements can be reached even when significantly different interests are represented. Some SSOs have been working to implement these actions for some time.

Participation transparency: Assessing who has contributed to a standard, and whether all necessary stakeholders are (or have been) represented, is often not straightforward because, when developing or releasing standards, SSOs don't normally indicate who contributed to the development¹⁷. Providing this information would help stakeholders decide if they feel represented or they should instead offer their own contribution to the standard. A third result therefore is that principled standardization would be facilitated if SSOs systematically published information on who participates in standard development (as ISO does) and this information were included as part of the published standard. Similarly, SSOs don't publish detailed statistics about overall participation, it is therefore difficult for stakeholders to assess whether there is an overall bias in standard development. It is generally reported that large corporates represent the vast majority of participants¹⁸, but no statistic is available to assess which other

17 ISO publishes committee membership during standard development, this information however, is no longer available once the standard is published (although it may be possible to find it in the Web archives)

18 For example, in a recent report, the EU Committee on the Internal Market and Consumer Protection, highlights that “*the effective participation of European SMEs and other societal stakeholders in international*

interests are represented or are missing. ISO publishes statistics on the number of convenorships and secretariats by countries from which emerges a very significant imbalance in their distribution. In 2022, for example, five countries (US, Germany, China, Japan, and the UK) held 61% of these positions, with the US and Germany at 16% each (ISO, 2023)¹⁹. The reasons for such imbalance are numerous, ranging from the costs associate to participation, to language barriers, and they are often out of the control of any one agent, but the availability of detailed data on participation to standard setting would be a first step in identifying unbalanced representation and correcting it. From this, a fourth suggestion is that SSOs regularly publish detailed statistics about participation in standard development, these statistics should include the type of representativity (country/region, civil society interest, commercial interests, etc.) This form of transparency would strengthen the transparency principle already present in most standardization guidelines. Transparency, as defined by the WTO guidelines, requires that all relevant information regarding ongoing work programs, proposed standards, and final outcomes is readily available to interested parties in the territories of all WTO Members. Additionally, according to the WTO, procedures should be implemented to allow sufficient time for written comments, and the details of these procedures should be effectively communicated to ensure widespread awareness (WTO, 2000). In the US, the ANSI's requirements for due process do not explicitly mention transparency,

standardisation continues to be a major challenge, owing to limited resources such as funding” (Bielan, 2022, general remark 28)

- ¹⁹ In a study of potential drivers for national stakeholders to participate in standardization, Blind and von Laer (2022) highlight how, between 2008 and 2016, the number of secretariats held by the US has decreased sharply while increasing for Germany and Asian countries.

however they indicate that SSOs should notify their work on current and new standards and make written procedures available to “directly and materially interested parties” (ANSI, 2022 see 1.5, 1.9 and 2.5). EU Regulation 1025/2012 mostly follows the WTO transparency requirements with the additional condition that work programs should be made available for the wider public (Regulation EU 1025/2012, 2012, art. 3.3.) Transparency, therefore, should ensure that interested stakeholders have access to documents related to standards and to standard development processes and do so in a manner that enables their participation in standard development or use of standards. Unfortunately, several barriers still exist that often prevent real transparency. Kanevskaia notes that *“many standards developed by consortia and informal interest groups may be endorsed through fast-track processes of formal SSOs; in some cases, this would mean that a technical document is directly submitted for the (final) approval, depriving interested stakeholders of opportunities to participate in the earlier stages of standards development.”* (Kanevskaia, 2019: 29). The fact that SMEs, and civil society organizations in particular, still have inconsistent access to standards and standardization processes is often attributed to this lack of transparency.

Widespread education on standardization: Stakeholders with heterogeneous backgrounds generally have different levels of ability to effectively access and participate in the work of SSOs. In our case study, we have highlighted two types of competencies that need to be supported to ensure effective participation; first, knowledge about standard development processes, and the procedures proper to the main SSOs and, second, knowledge about the key technical, legal, and societal facts related to the technology being standardized. While the former enables understanding of standardization processes in

general, the latter enables active participation to the development of standards for a specific technology. A fifth takeaway, therefore, is that, as highlighted in the recent European report on standardization (Bielan, 2022), it is essential to establish educational programs capable of raising awareness, build knowledge, and foster engagement in the standardization processes, while also highlighting the role they play in society. These programs should be addressed to all those susceptible to contributing to standardization processes among academics, students, industry professionals, policymakers, and civil society representatives with particular care to include small and medium enterprises and marginalized stakeholders. ISO's twinning program (ISO, 2021) and the teaching material developed by the Penn Program on Regulation (2023) in collaboration with the National Institute of Standards and Technology, are excellent initiatives in this direction. These initiatives are designed, respectively, for stakeholders already involved in standardization processes and law students. The European Telecommunications Standards Institute (ETSI) offers material for alphabetization programs on standard processes which is addressed to a much wider audience (ETSI, n.d.).

Consistent dissemination and training: A sixth outcome is that, to promote active and effective engagement from diverse stakeholders and facilitate widespread adoption and proper implementation of new standards, it is crucial that standardization processes establish training initiatives in preparation, during, and after the standard development. These initiatives should be designed to address the three different phases. First, to inform the widest audience possible about the subject of the standard to be developed. Second, to provide interested stakeholders (who possibly have diverse backgrounds) with accessible material about the technology so that they can meaningfully

participate to the development. Third, to make the standard and any supporting material available to anyone interested. These training actions should go much further than what is currently available, which normally consists in providing a title or a succinct abstract in the first phase, a request for comments in the second phase, and a final document (often behind paywall) in the third phase. In a recent assessment of European cybersecurity standardization procedures, for example, the authors recommend that “the results of the work of SDOs (standards, technical reports) are made freely available to universities or independent cyber security researchers, as otherwise security research will be hindered” (Cyber Security for Europe, 2022: 38).

Conclusions

By leading standardization processes, stakeholders (e.g., countries, industrial organizations) certainly gain a leadership position in technology development. However, leadership, if intended as gaining more benefits than others, only leads to a zero-sum: more benefits for one stakeholder leads to less benefits for others. If instead, leadership is intended as effectively responding to different needs and coordinating action, then everyone will be able to gain. If a country aims to take a leading role in technology, it should prioritize utilizing technology to deliver tangible benefits for its entire population. Similarly, other organizations should aim at benefiting their own customers and consumers in general. Consequently, the process of standard development should demonstrate a strong and universally acknowledged commitment to leveraging technology for the advancement of public welfare and fairness. This presents an opportunity to establish

exemplary practices, exhibit leadership in these domains, and export expertise to effectively address these objectives. Given the important role that standard development plays in the global economy, supporting trade, ensuring safety, and supporting international dialogue and collaboration, *principled standardization processes* would strengthen standardization by taking into account the democratic principles that guide our societies, balancing innovation with ethical considerations, involving all stakeholders, and ensuring effective and fair standards design.

In addition to the well-established principles for standardization processes, a focus on foresight exercises, extended representativity, participation transparency, widespread education on standardization and consistent dissemination and training, would help ensuring that standardization processes follow a pattern of transparency, inclusivity, and accountability to support the responsible and equitable use of technologies while promoting principles of equity and justice. *Principled standardization processes* do require considering a larger number of stakeholders, providing them with the capacity for effective participation and, sometimes, reconciling very different value systems. However, surmounting the difficulties represented by these requirements may be the only way to effectively align standardization processes with the needs of addressing increasingly complex and multi-faceted problems.

List of abbreviations

ANSI	American National Standards Institute
BSI	British Standards Institution
BWC	Body-worn camera
CEN	European Committee for Standardization
CENELEC	European Committee for Electrotechnical Standardization
COPA/NT	Pan American Standards Commission
DT	Digital Twin
ETSI	European Telecommunications Standards Institute
GDPR	General Data Protection Regulation
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
ITU	International Telecommunication Union
LEA	Law Enforcement Agent
LED	Law Enforcement Directive
PASC	Europe, the Pacific Area Standards Congress
SDO	Standard Developing Organization
SME	Small and Medium Enterprises
SSO	Standard Setting Organization
TBT	Technical Barriers to Trade Agreement (WTO)
UNIDO	United Nations Industrial Development Organization
W3C	World Wide Web Consortium
WTO	World Trade Organization

References

Adams, I., & Mastracci, S. (2019). Police body-worn cameras: Development of the Perceived Intensity of Monitoring Scale. *Criminal Justice Review*, 44(3), 386–405. <https://doi.org/10.1177/0734016818823689>

ANSI. (2011). *Why voluntary consensus standards incorporated by reference into federal government regulations are copyright protected*. <https://share.ansi.org/shared%20documents/News%20and%20Publications/Critical%20Issues/Copyright%20on%20Standards%20in%20Regulations/Copyright%20on%20Standards%20in%20Regulation.pdf>

ANSI. (2022). *ANSI essential requirements: Due process requirements for American national standards*. <https://www.ansi.org/american-national-standards/ans-introduction/essential-requirements>

Ariel, B., Sutherland, A., Henstock, D., Young, J., Drover, P., Sykes, J., Megicks, S., & Henderson, R. (2017). “Contagious accountability”: A global multisite randomized controlled trial on the effect of police body-worn cameras on citizens’ complaints against the police. *Criminal Justice and Behavior*, 44(2), 293–316. <https://doi.org/10.1177/0093854816668218>

Attaran, M., & Celik, B. G. (2023). Digital twin: Benefits, use cases, challenges, and opportunities. *Decision Analytics Journal*, 6, 100165. <https://doi.org/10.1016/j.dajour.2023.100165>

Baron, J., & Kanevskaia, O. (2021). Global competition for leadership positions in standards development organizations. SSRN Scholarly Paper No. 3818143. <https://doi.org/10.2139/ssrn.3818143>

Bertuzzi, L. (2023, March 28). MEPs ask EU Commission to reconsider standardisation approach. *Euractiv*. <https://www.euractiv.com/section/digital/news/meps-ask-eu-commission-to-reconsider-standardisation-approach/>

Bielan, A. (2022). *Report on a standardisation strategy for the single market | A9-0136/2023 | European Parliament*. https://www.europarl.europa.eu/doceo/document/A-9-2023-0136_EN.html

Blind, K., Mangelsdorf, A., & Jungmittag, A. (2011). *The economic benefits of standardization*. DIN German Institute for Standardization. <https://www.din.de/blob/89552/68849fab0eeaaafb56c5a3ffee9959c5/economic-benefits-of-standardization-en-data.pdf>

Blind, K., & von Laer, M. (2022). Paving the path: Drivers of standardization participation at ISO. *The Journal of Technology Transfer*, 47(4), 1115–1134. <https://doi.org/10.1007/s10961-021-09867-w>

Boström, M., & Hallström, K. T. (2013). Global multi-stakeholder standard setters: How fragile are they? *Journal of Global Ethics*, 9(1), 93–110. <https://doi.org/10.1080/17449626.2013.773181>

Bradshaw, S., & DeNardis, L. (2019). Privacy by infrastructure: The unresolved case of the domain name system. *Policy & Internet*, 11(1), 16–36. <https://doi.org/10.1002/poi3.187>

Bremer, E. S. (2013). Incorporation by reference in an open-government age. *Harvard Journal of Law and Public Policy*, 36(1), 131–190. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2127288#

Bremer, E. S. (2019). Technical standards meet administrative law: A teaching guide on incorporation by reference. *Administrative Law Review*, 71, 315–355.

BSI. (n.d.). *Our history—ISO, certification, standards*. <https://www.bsigroup.com/en-GB/about-bsi/our-history/>

BSI. (2017). *Code of practice for the deployment and use of body-worn video (BWV)*. BSI British Standards. <https://doi.org/10.3403/30345718>

Burke, A. (2020). The impact of body technologies upon the teaching of criminal law. *Bond University Centre for Professional Legal Education*.

CNIL. (2021, June 2). “Law enforcement directive”: What are we talking about? Commission Nationale de l’Informatique et des Libertés (CNIL). <https://www.cnil.fr/en/law-enforcement-directive-what-are-we-talking-about>

Coglianese, C. (2023). Standards and the law. *Standardization: Journal of Research and Innovation*, 2(2), 15–20. <https://ssrn.com/abstract=4452726>

Cyber Security for Europe. (2022). *D8.4 Standardisation procedure assessment document*. <https://cybersec4europe.eu/wp-content/uploads/2022/04/D8.4-Standardisation-Procedure-Assessment-Document-v1.0-submitted.pdf>

Davies, A., & Krame, G. (2023). Through an Australian lens: The influence of body-worn cameras on complaints against police—Beyond the numbers. *Policing: A Journal of Policy and Practice*, 17, paad015. <https://doi.org/10.1093/police/paad015>

Degallaix, L. (2018). *The use of standards in legislation and policies*. ECOS.

DeNardis, L., & Musiani, F. (2014, September 15). Governance by infrastructure: Introduction, 'The turn to infrastructure in internet governance.' SSRN Scholarly Paper No. 2730689. <https://doi.org/10.2139/ssrn.2730689>

Department of Homeland Security (DHS). (2023). *Department policy on body-worn cameras (Policy statement 045-07)*. https://www.dhs.gov/sites/default/files/2023-05/23_0522_opa_signed-dhs-policy-on-body-worn-cameras-508.pdf

Dodd, V. (2023, May 16). Ministers looking at body-worn facial recognition technology for police. *The Guardian*. <https://www.theguardian.com/uk-news/2023/may/16/ministers-calling-for-facial-recognition-technology-in-police-bodycams>

Doffman, Z. (2019, January 10). Facial recognition is coming to police body-worn cameras in 2019. *Forbes*. <https://www.forbes.com/sites/zakdoffman/2019/01/10/body-worn-2-0-how-iot-facial-recognition-is-set-to-change-frontline-policing/>

EC OJ L 322. (2009). Council framework decision 2009/905/JHA of 30 November 2009 on accreditation of forensic service providers carrying out laboratory activities. *Official Journal of the European Union*. http://data.europa.eu/eli/dec_framw/2009/905/oj/eng

ETSI. (n.d.). *Education about standardization*. <https://www.etsi.org/education>

European Commission. (n.d.). *Harmonised standards*. https://single-market-economy.ec.europa.eu/single-market/european-standards/harmonised-standards_en

European Commission. (2022). *An EU strategy on standardisation: Setting global standards in support of a resilient, green, and digital EU single market*. <https://ec.europa.eu/docsroom/documents/48598>

European Parliament and Council. (2012). *Regulation (EU) No 1025/2012 of the European Parliament and of the Council*. Official Journal of the European Union, L 316/12.

European Parliament. (2016). *Directive (EU) 2016/680 of 27 April 2016 on the protection of natural persons regarding processing data for the prevention, detection, or prosecution of crimes*. Official Journal of the European Union, L 119.

European Court of Justice. (2016, October 27). *James Elliott Construction Limited v Irish Asphalt Limited*, Case C-613/14.

Farrell, J., & Saloner, G. (1985). Standardization, compatibility, and innovation. *The RAND Journal of Economics*, 16(1), 70–83. <https://doi.org/10.2307/2555589>

Ferrazares, T. (2024). Monitoring police with body-worn cameras: Evidence from Chicago. *Journal of Urban Economics*, 141. <https://doi.org/10.1016/j.jue.2023.103539>

Fontanelli, F. (2011). ISO and CODEX standards and international trade law: What gets said is not what's heard. *International and Comparative Law Quarterly*, 60(4), 895–932. <https://doi.org/10.1017/S0020589311000459>

Gataveckaite, G. (2023, May 31). Green Party digs heels in over Garda body cams amid coalition dispute over facial recognition. *Independent.ie*. <https://www.independent.ie/irish-news/green-party-digs-heels-in-over-garda-body-cams-amid-ongoing-coalition-dispute-over-facial-recognition/a566347508.html>

Hansen Löfstrand, C., & Backman, C. (2021). Control or protection? Work environment implications of police body-worn cameras. *New Technology, Work and Employment*, 36(3), 327–347. <https://doi.org/10.1111/ntwe.12201>

Haskins, P. A. (2019, October 24). Research will shape the future of proactive policing. *National Institute of Justice Journal*, 281. <https://nij.ojp.gov/topics/articles/research-will-shape-future-proactive-policing>

Holmes, D., Papathanasaki, M., Maglaras, L., Ferrag, M. A., Nepal, S., & Janicke, H. (2021). Digital twins and cyber security: Solution or challenge? *2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*, 1–8.

Huang, P., Kim, K., & Schermer, M. (2022). Ethical issues of digital twins for personalized health care service: Preliminary mapping study. *Journal of Medical Internet Research*, 24(1), e33081. <https://doi.org/10.2196/33081>

IPC Ontario. (2021). *Information and Privacy Commissioner of Ontario—Model governance framework for police body-worn camera programs in Ontario*. <https://www.ipc.on.ca/wp-content/uploads/2021/07/model-governance-framework-police-body-worn-camera-programs.pdf>

ISO. (n.d.). *ISO—Standards*. <https://www.iso.org/standards.html>

ISO. (2021). *Guidance on twinning in ISO standards development activities* (6th ed.). <https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100341.pdf>

ISO. (2023). *ISO in Figures 2022*. https://www.iso.org/files/live/sites/isoorg/files/about%20ISO/iso_in_figures/docs/ISO_in_Figures_2022.pdf

ISO 23247. (2021). *Automation systems and integration—Digital twin framework for manufacturing—Part 1: Overview and general principles*. <https://www.iso.org/standard/75066.html>

ISO/IEC. (2015). *Using and referencing IEC and ISO standards to support public policy*. <https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100358.pdf>

Kanevskaia, O. (2019). Governance of ICT standardization: Due process in technocratic decision-making. SSRN Scholarly Paper No. 3400838. <https://doi.org/10.2139/ssrn.3400838>

Katz, C. M., & Huff, J. (2023). The Achilles heel of police body-worn cameras: Understanding the factors that influence variation in activation. *Justice Quarterly*, 40(3), 315–336.

Kritzinger, W., Karner, M., Traar, G., Henjes, J., & Sihh, W. (2018). Digital twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, 51(11), 1016–1022. <https://doi.org/10.1016/j.ifacol.2018.08.474>

Légifrance. (2019, February 27). Décret n° 2019-140 du 27 février 2019 portant application de l'article L. 241-2 du code de la sécurité intérieure et relatif à la mise en œuvre de traitements de données à caractère personnel provenant des caméras individuelles des agents de la police municipale. *Légifrance*.

Lum, C., Koper, C. S., Wilson, D. B., Stoltz, M., Goodier, M., Eggins, E., Higginson, A., & Mazerolle, L. (2020). Body-worn cameras' effects on police officers and citizen behavior: A systematic review. *Campbell Systematic Reviews*, 16(3), e1112. <https://doi.org/10.1002/cl2.1112>

Mavroidis, P. C., & Wolfe, R. (2017). Private standards and the WTO: Reclusive no more. *European University Institute, Robert Schuman Centre for Advanced Studies*. https://cadmus.eui.eu/bitstream/handle/1814/40384/RSCAS_2016_17.pdf

McIlwain, C. (2017). Racial formation, inequality, and the political economy of web traffic. *Information, Communication & Society*, 20(7), 1073–1089. <https://doi.org/10.1080/1369118X.2017.1309442>

Ministère de l'Intérieur. (2018, June 7). Expérimentation de l'emploi des caméras mobiles par les agents de police municipale. *Ministère de l'Intérieur*. <http://www.interieur.gouv.fr/Archives/Archives-des-actualites/2018-Actualites/Experimentation-de-l-emploi-des-cameras-mobiles-par-les-agents-de-police-municipale>

Mittelstadt, B. (2021). Near-term ethical challenges of digital twins. *Journal of Medical Ethics*, 47(6), 405–406. <https://doi.org/10.1136/medeth-ics-2020-107054>

Murphy, J. R., & Estcourt, D. (2020). Surveillance and the state: Body-worn cameras, privacy, and democratic policing. *Current Issues in Criminal Justice*, 32(3), 368–378. <https://doi.org/10.1080/10345329.2020.1774087>

Newell, B. C. (2021). *Police visibility: Privacy, surveillance, and the false promise of body-worn cameras*. University of California Press. <https://doi.org/10.1525/9780520383175>

NPCC. (2022). *British National Police Chiefs' Council—Police operational advice and security industry requirements for response to security systems*. https://www.policesecuritysystems.com/images/NPCC_Police_Requirements_for_Security_Systems_-_May_2021.pdf

Paris, B. (2021). The internet of futures past: Values trajectories of networking protocol projects. *Science, Technology, & Human Values*, 46(5), 1021–1047. <https://doi.org/10.1177/01622439211012451>

Penn Program on Regulation. (2023). *Voluntary codes and standards: Teaching resources for law and public policy courses*. <https://pennreg.org/codes-standards/>

Privacy International. (2029). Real-time facial recognition should never be coupled with body-worn cameras. *Privacy International*. <http://www.privacyinternational.org/news-analysis/3028/real-time-facial-recognition-should-never-be-coupled-body-worn-cameras>

Pyo, S. (2022). Understanding the adoption and implementation of body-worn cameras among U.S. local police departments. *Urban Affairs Review*, 58(1), 258–289. <https://doi.org/10.1177/10780874211043092>

Quezada-Tavárez, K. (2021). Impact of the right of access on the balance between security and fundamental rights. *European Data Protection Law Review*, 7(1), 59–73. <https://doi.org/10.21552/edpl/2021/1/8>

Ringrose, K. (2019, February 18). Law enforcement's pairing of facial recognition technology with body-worn cameras escalates privacy concerns. *Virginia Law Review (online)*, 105, 57. <https://virginialawreview.org/articles/law-enforcements-pairing-facial-recognition-technology-body-worn-cameras-escalates/>

Royal Canadian Mounted Police (RCMP). (2023a, May 10). Operational manual—Chapter 25.5: Body-worn video. <https://www.rcmp-grc.gc.ca/corporate-organisation/transparency-transparence/policies-politiques/25-5-eng.htm>

Royal Canadian Mounted Police (RCMP). (2023b). Body-worn cameras. <https://rcmp.ca/en/body-worn-cameras>

Security World Market. (2021, May 12). France invests \$17M on huge police body-worn camera deployment. *Security World Market*. <https://www.securityworldmarket.com/int/News/Business-News/france-invests-17m-on-huge-police-body-worn-camera-deployment1>

Shapiro, C., & Varian, H. (1999). The art of standards wars. *California Management Review*, 41(2), 8–32. <https://doi.org/10.2307/41165984>

Sharma, A., Kosasih, E., Zhang, J., Brintrup, A., & Calinescu, A. (2022). Digital twins: State of the art theory and practice, challenges, and open research questions. *Journal of Industrial Information Integration*, 30, 100383. <https://doi.org/10.1016/j.jii.2022.100383>

Tassey, G. (2000). Standardization in technology-based markets. *Research Policy*, 29(4–5), 587–602. [https://doi.org/10.1016/S0048-7333\(99\)00091-8](https://doi.org/10.1016/S0048-7333(99)00091-8)

Timmermans, S., & Epstein, S. (2010). A world of standards but not a standard world: Toward a sociology of standards and standardization. *Annual Review of Sociology*, 36(1), 69–89. <https://doi.org/10.1146/annurev.soc.012809.102629>

Turner Lee, N., & Chin, C. (2022, April 7). Police surveillance and facial recognition: Why data privacy is imperative for communities of color. *Brookings*. <https://www.brookings.edu/research/police-surveillance-and-facial-recognition-why-data-privacy-is-an-imperative-for-communities-of-color/>

UK Secretary of State. (2018). *The accreditation of forensic service providers regulations*. <https://www.legislation.gov.uk/ukxi/2018/1276/made>

UNIDO. (2021). *Standards and digital transformation: Good governance in a digital age*. United Nations Industrial Development Organization. https://www.unido.org/sites/default/files/files/2021-11/Standards%20and%20Digital%20Transformation_Complete_2021.pdf

van Waeyenberge, A., & Amariles, D. R. (2017). James Elliott Construction: A “new(ish) approach” to judicial review of standardisation. *European Law Review*, 6, 705–716.

Wang, Y., Su, Z., Guo, S., Dai, M., Luan, T. H., & Liu, Y. (2023). A survey on digital twins: Architecture, enabling technologies, security and privacy, and future prospects. *IEEE Internet of Things Journal*, 1(1), 1–15. <https://doi.org/10.1109/JIOT.2023.3279554>

Williams, P., & Kind, E. (2019). *Data-driven policing: The hardwiring of discriminatory policing practices across Europe*. European Network Against Racism. <https://raceandpolicing.issuelab.org/resources/36920/36920.pdf>

Wright, J. E., & Brown, H. (2020). What matters more? Police or the community for body-worn camera policy. *Public Administration Quarterly*, 44(3), 462–482.

WTO. (1995). *Agreement on technical barriers to trade*. https://www.wto.org/english/docs_e/legal_e/17-tbt_e.htm

WTO. (2000). *Principles for the development of international standards, guides, and recommendations*. https://www.wto.org/english/news_e/news19_e/ddgaw_05dec19_e.htm

Technology-Facilitated Gender-Based Violence and the Novel: Realizing Women's Rights

Sekai Zhou

Introduction

Modern emerging technologies enable us to connect almost everywhere and at any time. Emerging technologies, such as the internet, advanced wireless technologies, and mobile communication networks, are becoming increasingly indispensable in various aspects of work and everyday life. However, despite the ease of work and life that modern emerging technologies provide, they also present risks to vulnerable and marginalized groups in society. Emerging technology's ability to shape society is of concern, especially to women who are vulnerable to violence.

This paper explores the risks emerging technologies (Facebook) pose towards the realization of the rights of women in the virtual space. Of particular interest is how emerging technologies, through

Sekai Zhou, Independent Researcher, Bulawayo ORCID: 0000-0001-9324-9355, email: sekaizhou2009@gmail.com

DOI: <https://doi.org/10.5281/zenodo.14204541>

technology-facilitated gender-based violence (TFGBV), impact women's lives. Through female-authored novels, this paper demonstrates that TFGBV is a social, structural, and systemic problem that seemingly mirrors non-technology-facilitated gender-based violence (NT-FGBV). While there are attempts to regulate virtual space there are pertinent questions.

RQ1: How emerging technologies, through technology-facilitated gender-based violence (TFGBV), impact women's lives

RQ2: Why do the law and human rights seemingly fail to address TFGBV?

RQ3: Can women-authored novels reveal insights into why TFGBV persists or may be addressed?

Technology-facilitated gender-based violence (TFGBV) is a multifaceted, universal problem that produces devastating consequences. Women, because of gender, race, culture, religion, and others, are subject to violence facilitated through technology. This means women's safety is uncertain. Of greater concern is the silencing of women in digital spaces. TFGBV makes it unsafe and unwelcoming for women to express themselves freely in a world in which digital communication has become one of the primary modes of communication.

I employ a law and literature approach to address the question of justice for women in the face of TFGBV. Law co-exists with other disciplines; therefore, its application is contextual with other disciplines. I explore the intersection between law and literature in how they address the challenges that confront women. How do legal and literary texts represent justice?

I draw from several hybrid research methodologies to gain a greater insight into the challenges confronting African women. In addition, this approach allows for multiple perspectives on understanding a problem. Multiple interpretations of a problem help narrow down the source of the problem. Furthermore, there is flexibility in addressing the research questions. One of the limitations of this design is the risk of incompatible conclusions from various methods. I give deliberate attention to the integral elements of the data gathered. In this case, the several methods I use help show what and how women confront their challenges.

Reading from a law and literature perspective, I attempt to understand the predicament of women in the face of TFGBV. Literature refers to all literary texts, including novels, short stories, prose, and poems. However, for the purposes of this paper, literature is restricted to novels. I use female-authored novels to explore the challenges TFGBV confronts women. The novels do not entirely represent the challenges faced by women but provide a glimpse and a perspective on them. The overall idea is to come up with an understanding of why human rights and law seem inadequate in addressing women's challenges that TFGBV present. This study does not suggest a refusal of human rights but aims to indicate gaps in the concept regarding women in the virtual space.

Relying on the research done to date on TFGBV, this paper examines the studies done on TFGBV. Following this the paper examines the inadequacy of the law and human rights in addressing TFGBV. It explores how women-authored novels can contribute to law and human rights by finding a possible remedy to TFGBV. Finally, the paper proposes that law and human rights must be empathetic responses to TFGBV.

This paper will prove useful for policymakers, civil society organizations, and others interested in gaining a better understanding of TFGBV. In addition, it will respond to Sustainable Development Goal 5, which highlights the need to achieve gender equality and empower all women and girls. Specifically, Goal 5.2, which aims to eliminate all forms of violence against all women and girls in the public and private spheres, including trafficking, sexual exploitation, and other types of exploitation.

Understanding and Mapping TFGBV

TFGBV is a contemporary system of gender-based violence that makes use of digital technologies to cause harm. The increased usage of digital technologies in everyday life sees a heavy proliferation of TFGV. The last two years, with much of people's lives moving online due to the COVID-19 pandemic, witnessed an increase in TFGBV. The root of TFGBV is discriminatory beliefs and institutions that reinforce sexist gender norms. This is no different from other forms of GBV. TFGBV overlaps with race, gender, and class, creating interdependent systems of discrimination. There are a few but growing studies on the comparatively new concept of TFGBV, which include several empirical studies. Most of this research attests to the gravity of TFGBV as a universal problem. In addition, these studies generally focused on high-income countries, which necessitates the need to consider studies of lower- and middle-income countries. Above all, these studies may not have considered the law as an accomplice to TFGBV. This may explain why the law and human rights seemingly fail

to address TFGBV. Another thought on the previous studies on TFGBV is the possibility of the novel as a reservoir of the TFGBV experiences of women. Exploring novels, although fiction, may provide an alternative source of TFGBV experiences of women that may influence law and human rights.

A considerable amount of literature of late has been published to date on TFGBV. The Kaspersky report on digital stalking in relationships (2022) indicates one in four people (25%) have experienced some form of abuse by their partner, although males are less likely to have suffered abuse (19%) than females (30%). Women and girls are more vulnerable to TFGBV. According to Bailey et al. (2022), for legal and policy decisions to be evocative and successful, they must be constructed in knowledge that focuses on the real-life experiences of members of marginalized communities, particularly those represented in social science findings. In essence, their position although indicating the need for law to be sensitive to the marginalized, it does not recognise that the roots of law are in exclusivity rather than inclusivity. This chapter attempts to address that gap. In another study, Dunn (2020) gives a detailed overview on the studies so far on TFGBV. Dunn echoes the same sentiments as Bailey et al. (2022) that emphasizes the need to explore experiences of the marginalized. She indicates that law can help address TFGBV. In addition, she stresses that more studies are needed on the Global South's response to TFGBV. Although I agree on the need for more studies on Africa, I disagree on the current law as the way to address the scourge of TFGBV. The section on the inadequacies of law responds to this assertion. United Nations Women (2020a) and Plan International (2020) both produced reports at the height of Covid 19 in 2020 that explore the vulnerability of women and girls online. Both reports underscore the role of

patriarchy in making women and girls vulnerable on and offline. In addition, both reports emphasize the danger of TFGBV that results in silencing women and hindering them from public participation. Plan International whose report is a survey of girls and young women of online harassment highlights the need to have legal recourse to address the scourge of TFGBV.

The existing literature on TFGBV ignores the possibility that law, and human rights contribute to the vulnerability of women to TFGBV. For example, the same law that encourages privacy and anonymity is the same that protects the identification of online abusers. In addition, existing literature fails to recognize how other fields of study such as literature (novels) may offer avenues of resolution that are outside the parameters of law and human rights. Most of the research while primary research and reflects the responses of victims of TFGBV, have a possibility of not fully covering the depths and trauma of TFGBV. In contrast, novels although fiction, explores the lived experiences of women without necessarily visiting their trauma as in primary research.

The Inadequacy of Law and Human Rights in Addressing TFGBV

The section demonstrates the inadequacies that exist in the law and the human rights system concerning women. The inadequacy of the law in addressing TFGBV is rooted in its origins and irrefutable link to western modernity. Enlightenment introduces the idea of reason as the principal approach to achieving success and self-realization. The

idea of reason resonates strongly in modern law and human rights. There is a need to consider alternative modernities such as providing an atypical interrogation of society. This means that alternative modernities examine society away from the usual methods and narratives.

Law as a Tool of Social Order

One feature of modern law is the use of the law as a tool of social order in the modern world. This understanding traces back to the Enlightenment era, which facilitated the introduction of law as a tool of social order. Social order became necessary after modernity created social chaos in a bid to accommodate capitalism. Modernity created social chaos with its promises and possibilities of emancipation that capitalism had submerged. Society identifies the rise of capitalism with modernity. It creates tension and exposes the inequalities that exist in society. Law as an act of governmental social control. The law is part of systems of social control. I deduce several assumptions about law and social order. First, that law is regulatory. Second, the world without the law is tumultuous. Third, it presupposes that the law has the aptitude to carry out this mandate. The law is dependent on society. The law depends on society to remain relevant. Without society, the law ceases to exercise the power it has. In addition, modern law is mythic as it depends on other factors to thrive. The mythic view of the law is based on the idea that it is self-grounding, that it has no origins and thrives on its own. It is a myth that something with no origins attaches itself to modernity and becomes the guiding force of the modern social order. The law cannot function on its own without enforcers.

Simply acknowledging the law as a tool of social control without dissecting that declaration leaves questions. For women, the difference between identifying the power behind the law as a social regulator and addressing the challenges confronting them is of crucial importance to their accessing justice. The law is not just for women; it determines their identity, ability, and place in their society. The law is a system of rules; it inevitably embodies tenets, philosophies, and a wealth of meanings. In addition, they write that the tenets of the legal system favour the dominant ideology of a society. Fitzpatrick (2005) describes the origins of law as 'vacuous', 'empty,' and 'layered irony'. This speaks to the mythological aspect of the law, which exists like a fable that the world accepts as credible. If the origins of law are empty or exclude certain parts of humanity, then what fills that vacuum? The male-dominated systems not only exclude women but also fortify their inferior positions in their societies. If the law is a tool of social order, it must be just, fair, and equal, regardless of gender, religious beliefs, race, colour, location, or sexual orientation. Yet, what is tragic is that women in the modern era continue to battle for the law to recognize their challenges and offer a working solution. The struggle should never be for the law to recognize them.

Feminists are alert to the representation of women in all spheres, real or imaginary. Real spheres include law, business, politics, and economics; the imaginary includes films, stories, etc. Postcolonial feminist scholars, Rajan and Park (2005) reaffirm that the theoretical concerns of modern postcolonial feminism are representation and location. Representation and location articulate issues of race, gender, class, nationality, and sexuality, which are specific to the challenges confronting women. The representation and location of women are absent in the law; therefore, the law cannot be responsive to the

challenges confronting them. Women require a law that simply provides justice whenever they require it.

Loss of Morality

Another feature of modern law stems from the loss of morality in pursuit of professionalism. Douzinas and others explain that one defining feature of legal modernity is the separation of legality and morality. Law excludes morality, ethics, and the right to question the legality, validity, and form of law. This explains why the law separates itself from morality to guard itself against questions about how it operates to provide justice. Douzinas and Geary (2005) point out that the dawn of other disciplines as separate from the law, such as sociology and psychology, deprived the modern legal theory of cognitivism and morality. In other words, sociology, which studies the science of society, social institutions, and social relationships, would have been effective as part of the law.

Paradoxically, the law is knit into the everyday functions of society, yet it avoids any attachment to sociological interests. If law detaches itself from sociological interests, it may miss understanding the background of women. This background includes the legacy of the Enlightenment, patriarchy, and other harmful traditions. How then will the law respond to the challenges confronting women?

At the core of the cognitivism debate in law is the idea of truth and falsity and the mental acceptance of morality. The problems that women confront in accessing justice seem to result from the loss of

cognitivism and morality due to modernity. The law is fallible. Even though the law needs to maintain distance from cognitivism and morality, it is fallible and changeable. What is really at stake is the protection of the authority that the law has over people. Examination of the authority of the law threatens its reputation. Legal cognitivism commands attention because of the likelihood of human failure to interpret the law accurately within the jurisdiction it provides. The likelihood of human failure is what the law avoids by denying attachment to cognitivism and morality. With the challenges confronting women juxtaposed with cognitivism and morality, there is a possibility that there is a gap in the law.

Admittedly, the birth of modern legal theory denotes the death of cognitivism and morality. The concerns of modern legal theory are rationalism and positivism. Positivism places emphasis on knowledge established by reason and evidence. Stacy (2001) expounds that positivism entails reading and pursuing the meaning expressed in legal words and texts. To put it concisely, positivism relies on reason and science to determine truth and worth. Stacy (2001) explains that the postmodernist argument against positivism is the ability to conceal power in reason. Reason expels morality as the law reduces it to a set of rules that objectify power. In essence, it is important for positivists to depersonalize power so that the law detaches from moral reflections. Cotterrell (1991) claims that law seemingly influences the political power that the state represents, leading to the loosening of links between law and morality. Cotterrell's interpretation is present in the relationship between modern law and the state. Although modern law refutes this relationship as it insists on the separation of law and state, it is the determination of law that forms the same state and the rules that govern its operation.

Acknowledging and addressing that critical question may reveal possible alternatives to challenges that adversely affect women. To expose the significance of jurisprudence (Douzinas and Geary, 2005), they illustrate that jurisprudence can investigate deeper into the background of various scenarios and determine what influences certain actions, stances, and convictions. In addition, Goodrich (1996) explains that the study of minor jurisprudences serves to interrogate the remains of law (discarded minor jurisprudence such as the court of love), of inventions, and 'fictive laws, of itineracy and fiction, as well as of contingent and local practices.' What Douzinas, Geary, and Goodrich imply is that as much as the law claims to be self-grounding, there remains a seriousness to comprehending the anatomy of the law. The process that interrogates the law cannot use the usual methodology but demands unconventional instruments that may yield different results. Outside the usual legal avenues, literature may provide an unconventional way to interrogate the law. I aver that to comprehend the loss of cognitivism and morality in law, inquiry may need to broaden beyond just law and its application. In fact, from a postmodernist perspective, one single legal theory cannot thoroughly comprehend the workings of the law. Relying on a single legal theory may result in laws that fail to respond to the conditions of women.

The Fallacy of the Universality of Human Rights

In the first place, the perception that human rights are universal is contentious. If human rights are universal, that implies that they are recognisable as claims that human beings can make worldwide by being

humans. The erroneous assumption is all persons are born with these rights by being human. If there were truth in that assumption about human rights, then women would not be confronting grave injustice daily. Questions arise over what undoubtedly purports to be universal since various parts of the world foster a unique appreciation of justice. Philosophically, in a multicultural, multipolar and plural world, how does one establish human rights as universal? Donnelly (1992) deciphers contesting the idea of universality against the alternative ways various parts of the world perceive the system of human rights, such as the Global South, besides economic, social, and cultural rights coveting for self-determination and development rights. Donnelly (1992) writes from a cultural relativism and a new order perspective, which means that the notion of the universality of human rights silences other social orders that exist in other parts of the world.

The debate on the universality of human rights points to the modernist trope that tends to homogenize and turn the world into a village. Human rights, like modernity as a universal system, subsume other narratives that equally have an opportunity to restore human dignity and offer justice. In essence, the claim that human rights are universal is arguable in various parts of the world; different people require different demands from the human rights system. Equally, there is no explicit universal agreement on what justice is and means to everyone, therefore making it difficult to accept that human rights are universal.

Modern human rights as a universal system close off other possible interpretations of what justice represents for women. This reflects the restricted jurisprudence familiar to legal modernity. Ward (2004) highlights that specific circumstances fulfil justice, not 'not in universal absolutes'. Drawing from philosopher and sociologist Jürgen

Habermas, Stacy (2001) explains that moral judgments must take into consideration not only the rules and the norms but also the moral sense of the individual. In other words, both Ward and Stacy underscore the grave importance of discovering the concerned individual and the peculiar circumstances surrounding that concerned individual. The universality of human rights claims inexplicably fails to accommodate various women in its just system. Women's challenges are not general, and they differ from one woman to another and from those of men.

Obsession with the Notion of Human Rights

The obsession with the notion of human rights or the excessive use of human rights language for anything and everything is part of the problem of human rights. Baxi (1998) observes that there is a 'weariness' and "wariness"—a sense that they are clichéd, fatigued, and probably even an impediment to justice instead of an avenue to providing it. This observation, while not probably shared by other scholars, presents a glimpse into the effects of modernity. Ultimately, fatigued human rights may not entirely respond to challenges affecting women online. Nevertheless, human rights appear to have a magical effect that promises to set the world at ease. This explains the obsession with the idea of human rights. Despite how utopian and idealistic they are, the promise they hold appears worthwhile for some sectors of society.

I propose two positions on what fixation with the idea of human rights in the modern era denotes. Firstly, a quick look through various media

reflects that the world has a different view of what entails human rights. There is a departure from the initial idea of human rights. Just like enlightenment shuns the ancient ways for the modern ones, the same evolution seems to occur with modern human rights. In addition, advances in technology have made the human rights space a mine-field. There is an increase in various cyberbullying tactics against, for example, African women, either demeaning their features, identities, or beliefs. Online platforms such as Twitter, Instagram, Facebook, and many others facilitate this abuse. Cyberbullying happens under the guise of freedom of speech and expression. Douzinas (2018) terms this a paradox of human rights,” where a violator of human rights can use them to violate To put it simply, the use of human rights or human rights language can either liberate or oppress the rights of the marginalized. All these abusive online tactics are not only dehumanizing but also take years to get justice. In a way, the obsession with rights suggests their misuse too. However, what is missing from the obsession with human rights conversation is the idea of responsibility. The modern world creates subjects whose concern is more to protect their rights and extraordinarily little about the rights of others. Who is responsible for promoting and protecting the rights of women online and offline? The enlightenment idea of self-realization as an individual journey translates into how, despite the presence of law and human rights, people continue to pursue their rights at the expense of others. In *Asking for It*, Emma is exposed both online and offline to violence from people who have the freedom to do so. Emma cannot identify any of the perpetrators of her online violence. The law allows these perpetrators to practice anonymity online, exposing the gap in justice for women. Literature exposes the lives of others that master discourses may overlook. *Asking for it* opens the reader to a world that would otherwise remain invisible.

Secondly, Douzinas (2007) acknowledges that in Western postmodern societies, the hackneyed phrase ‘I have a right to X’ is used interchangeably with the bland expressions ‘I desire or want X’ or ‘X should be given to me.’ Put another way, modernity blurs the difference between a need and a want such that the two become synonyms. At the core of the overwhelming demand and rapid use, language is the modernity, modernity which emphasises the freedom that individuals exercise. In addition, Ward (2004) explains that modernity provides conditions that stimulate individuals to pursue the re-control of their lives through being vocal and retaking the language that illustrates themselves. Individuals who feel they deserve or demand something may exercise their human rights to obtain what they require. While the use of rights to claim certain demands is laudable, Douzinas’ major worry is the motivation driving those demands and claims to rights. In other words, Douzinas assumes people cannot differentiate between a want and an entitlement. Douzinas (2007) solemnly believes that the tragic failure to separate wants from entitlement takes away the authority that human rights possess. Human rights and the human rights language are specific, applicable, and obligatory, but if used without caution, they can inevitably lead to divisions, grave loss of worth, and miscarriages of justice for women on and offline.

Women-Authored Novels

The choice for women-authored novels is motivated by the need for a woman’s voice in the struggle to understand TFGBV. In addition, the choice is necessitated by the need to find texts that have conversations that use technology. This has proven to be difficult, especially

for novels from southern Africa, where I come from. There are copious novels that show TFGBV, especially in the global north. This finding almost reflects the little research that has generally been done on TFGBV in Africa. The limited presence of TFGBV in women-authored novels in the global south is another area that needs further research.

Women-authored novels describe the conditions and experiences of women, therefore readily providing a female narrative absent in human rights jurisprudence. There is always an assumption that human rights are the solution to world problems, especially in the aftermath of the Second World War. The reality is that there is a need to constantly review and evaluate whether law and human rights do respond to the challenges that confront women. In turning to literature, to women-authored novels, I attempt to understand the experiences and realities of women that could better inform law and human rights.

Novels, although works of fiction capture the social relations of power between men and women. Novels reveal patriarchy as a system for maintaining class, gender, racial, and heterosexual privilege and the status quo of power—relies both on crude forms of oppression, like violence; and subtle ones, like laws; to perpetuate inequality. The novel is a powerful tool to show patriarchy because it paints images in the mind of the reader of the inequalities, discrimination and violence that confront women. In addition, the novel provides a voice to countless women who under the oppressive system of patriarchy would be silent about violence. The female-authored novel, through imagination, persuades the reader to experience being a woman in a virtual space. Imagination is important because it is useful to challenge, subvert or change things we find in our society. Although there are many other factors that influence our imagination, it is critical to note that

the female-authored novels offer insights that challenge the views about women in the virtual space.

Novels reveal gender construction in a patriarchy where males are socially respected while females represent a form of continuity. In addition, literature can provide a deeper understanding of characters and conflicts, which should allow readers to better understand the conflicting social and legal principles at work. It is these relationships that technology displays between men and women. Online women experience subjugation in a patriarchal virtual space. In addition, novels reveal that females' journeys of development are different from males. Rawoski (1983) explains that the art of living for women differs substantially from that of men, and as such, it is impossible. The art of living refers to how women experience life. The art of living for a woman is in recognition of her limitations. This means limitations are routine for any woman, but there are conditions that different women suffer because of where they come from. For example, women in Africa continuously contend with the burdens of colonialism, gender, and race.

Kehinde and Mbipom (2011) write that literature depicts the way different people relate to each other in that society. In addition, Matthewson (2004) argues that literature can elucidate the missteps and injustices of the law and offer persuasive alternate renderings of a just society. Interrogating the forces around which a text is produced lays bare the power relations and inequalities that exist in a system. It is critical to examine texts, or in this case, virtual space, to comprehend the forces around which they are produced and the power relations and inequalities that exist in that system. The woman-authored novel offers that interrogation with particular attention to the challenges that confront women.

Asking for it, a novel by Irish writer Louise O’Niel, includes 19 references to Facebook, a social media service. Emma O’Donovan, the 18-year-old narrator of this novel set in the small Irish village of Ballinatoom, is determined to express her identity (“I am sick of people thinking they know me”). A party puts her behavioural boundaries through a test, but the next morning she can’t recall what she did—or the awful things that happened to her.

The internet plays a nefarious role in the plot and in defining identity for teens accustomed to a “flattering Instagram filter”—yet nothing can filter out her pain. “I feel like someone has taken a blade to my insides,” Emma says after seeing the gruesome images taken of her that night, which have gone public on Facebook. She is overcome with humiliation as she reads the comments, which are “a wildfire, out of control.” The topic of sexual consent is central: the victim-blamers claim she was “asking for it,” that she is a “slut... bitch... skank... whore,” while more sympathetic voices believe she was gang-raped and deserves punishment.

In this novel, technology plays a significant role in Emma’s fate. This book delves into crucial issues such as rape culture, prejudice, and social media abuse, approaching taboo topics with nuance and understanding.

As one reads a novel, they may identify images, realities, absurdities that exist in their real lives. Just like in *Asking for it* readers may identify with the issues that Emma, her parents, school friends struggle with that are common to them every day. Although to an extent, what is in novels is the view of the author to certain issues in the society, it gives the reader an opportunity to reflect on issues the writer raises. Readers are at liberty to accept or deny what they are reading. This

process is important because it persuades the reader to engage with the novel. Literature not merely influences life but also shapes it. Literature is the brain of humanity, and it records and preserves the experience, knowledge, idea of human race. Literature offers the chance to arrive at the point to understand the true meaning and worth of life. In essence, it reveals the importance of empathy both on and offline to human relations.

Empathy in the Virtual Space

Novels reveal that empathy is necessary to facilitate dialogue on issues that are problematic for women in society. Readers not only identify the problems but also imagine themselves in those situations. Moral imagination helps to understand and interpret the social reality of the individual and that of others. On the contrary, modern moral imagination is a battlefield of antagonistic ideas and pictures of the world, discordant inclinations of cognisance, or contradictory forms of culture. Simply put, modernity offers many sources of information likely to influence moral imagination. For this reason, empathy remains fundamental to connecting people and searching for ways to function safely as a society. Sadly, the organization of modern society relies absolutely on law and human rights to determine social order rather than empathy. So far, the role of law and human rights as keepers of social order has proven problematic, especially in addressing the challenges that confront women. Empathy liberates not only women but also humankind.

Empathy manifests as cognitive, emotional, and compassionate. This implies empathy, which identifies with the idea of thinking, being sensitive, and being considerate of others. Jurecic (2011) explains that scientists discovered mirror neurons that ignite when a person acts or feels an emotion and when that person observes someone having the experience. Put simply, empathy is biological and inherent, and our bodies respond to our showing empathy to somebody else. This does not imply that people use empathy in everyday situations. At times, the mindless modernist pursuit of individualistic progression, self-realization, and success rules out the idea of others. Human rights as a system promote the individualistic pursuit of success instead of group success. To date, marginalized groupings struggle to assert their rights through the human rights system because its roots are in individualistic Enlightenment ideals. In contrast, the concept of Ubuntu, a widespread practice in southern Africa, is debatably believed to promote the collective progression and success of a community. This belief in the collective progression and success of a community is contested in other sectors of society.

Studies exist to corroborate that the study of novels encourages the growth and practice of empathy in society. These studies have different variables that separate one from the other, but they all point to the fact that reading literature may contribute to a better society. Mar et al. (2009) presented a study that explores the link between reading fiction and empathy while ruling out individual differences and examining outcomes. The Mar et al. (2009) study results reveal that fiction exposure invokes empathy in readers and improves their social abilities. In addition, the study rules out the idea that certain personalities are susceptible to empathy. People with personalities that are open to experience are most likely to feel empathy whenever they

read because their minds operate that way. However, the study indicates that empathy happens to people who do not necessarily have an open-minded personality. Furthermore, the study notes that exposure to fiction correlates with social support, while non-fiction is associated with loneliness and negatively relates to social support. In other words, reading fiction produces a feeling for the reader that they are a part of the community they read in. In comparison, non-fiction tends to produce a feeling of loneliness with little opportunity for sociability.

Stansfield and Bunce (2014) carried out a study with thirty-three participants in 2014 to test prior exposure to fiction, transportation, and a different measure of cognitive empathy, affective empathy, and helping tendency. The results reveal that exposure to fiction is associated with cognitive but not affective empathy, while the experience of being transported is associated with story-induced affective empathy. Readers who allow fictional narratives to transport them experience a sense of understanding and recognition of the characters. This means that reading fiction encourages cognitive empathy. Cognitive empathy recognizes what the other person experiences. The Harvard Business Review (2020) notes that increasingly there is a need to include a reading of fiction in business schools to produce future business leaders with empathy, theoretical, and critical thinking skills. This underscores the value of empathy not only in the field of human rights but also in the business sector.

What I observe from the research on the relationship between literature and empathy is that despite the many variables that each study pursues, the consensus remains that literature, specifically novels, may produce empathy in readers. It is debatable if reading these

stories does indeed change the way people feel and view the suffering of others in society. This challenge is one of the reasons why some sectors of society do not believe empathy can successfully change the world for the better. I consider that although empathy alone cannot change the world, its absence seems to keep it in constant chaos. The chaos is particularly tremendous for African women, who, as women, always maintain a subordinate position in society. Therefore, the absence of empathy in a system of human rights that seems to order the world is indeed concerning.

Another link between empathy and reading fiction is that it recognizes 'cognitive empathy as the ability to understand the world from another person's point of view and to infer beliefs and intentions, whereas affective empathy refers to the capacity to share another's feelings and emotions.' In essence, cognitive empathy is the aptitude to recognize the reality of another person and to understand the opinions and purposes of that person. Affective empathy is the skill of participating in the emotions and feelings of another person. Novelists sometimes illustrate or attract features of human insight, thought, feeling, or memory in ways that are more complex and accurate than the standard views of their contemporaries, including scientists. In short, novelists, through their stories, explore human areas that were once the preserve of science. For that reason, literature may better articulate the description of human insight, feeling, and memory. The inability to manage and understand human insight, feeling, and memory can lead to conflict. Law and human rights seem to fail to manage and understand women's insight, feelings, and memories.

Novels provide the experiences and realities of women that draw empathy from the reader. The novel gives a glimpse into how humans

interact with one another. In addition, the novel contextualizes the conversations so that readers can recognize the motivations for various human conditions. I contend that what differentiates the universal language of human rights from others, for example, the female-authored novel, is the idea of empathy that these stories invoke in readers. Reading literature makes readers more empathic. Alternatively, reading is not a meaningless function but a productive enterprise towards recognizing oneself and others. Recognizing oneself and others is important to making the physical and virtual worlds a better place.

Empathy Discontents

However, as much as the idea of empathy presents an opportunity for women to get justice, there are some arguments against it. Seagul (2018) admits that empathy is not necessarily the answer to a better life, but it does create a society that works together for safe and fulfilling lives. Empathy is at the centre of prejudice, not kindness. This idea believes that people feel empathy for people who they feel are different from them, as in, they are not in the same situation as they are. Instead of empathy, this scenario creates meta-stereotypes, which are stereotypes from which another group views or judges another.

Secondly, empathy is not necessary to do good. It is possible to do good with low levels of empathy. Religious or social movements could influence this. People do good because it is the right thing to do, not necessarily because deep inside there is a conviction to do

good. Today, the world recognizes human rights as a yardstick for doing good, despite their problems of conceptualization and operations that force others to bend to their will. A novel that inculcates empathy can also act as a tool to spread messages of good without necessarily convicting the heart and mind of the reader. It is possible that empathy can cause immoral choices. The latest research in psychology and neuroscience shows that people express empathy mostly towards those they find attractive and who appear to resemble them and not at all towards those who are dissimilar, distant, or nameless. Empathy, therefore, biases us in favour of individuals we know while numbing us to the plight of thousands.

Why Empathy?

Firstly, empathy is necessary for a sense of belonging. A sense of belonging is a psychological feeling of attachment to something. Women are in a position where patriarchy continuously erodes their sense of belonging. Patriarchal values and institutions strip women of the sense of being human, subsequently relegating them to the margins of society. Reading novels exhibits the inconsistencies that society carries about women. Van Schalkwyk (2018) confirms that patriarchy, for example, othered African women such that they found themselves inferior to belong to society. At the core of the system of patriarchy is the idea of using power to make women inferior, insignificant, and worthless. Van Schalkwyk (2018) highlights that (African) women's narratives dislocate patriarchal accounts about female sexuality as they together subdue binaries of dominant discourses. The novels place women not only at the centre of society but also as part

of that society. Female-authored novels make women visible to the larger society through the highlights of their narratives, along with the dominant ones. The hope is that reading female-written narratives encourages a sense of belonging not only in women but also in men, who realize that women have the same rights to belong in virtual and non-virtual societies.

Secondly, empathy is important because it promotes positive behaviour. The idea of what positive behaviour is a mammoth task to delineate. Rather, positive behaviour in this study is the one that treats women as deserving of the same opportunities and treatment as men before the law. In an ideal world, men and women should be equal before the law. It is also essential to note that having empathy does not guarantee that one will always do good, but it does encourage one to think of doing good. How does one define what is good for others? This is significant, especially noting that human rights are a system that the world presumes ensures consistent good in society. Bethlehem et al. (2017) used real-life scenarios to understand what leads people to be kind and confirm that the act of doing good correlates with empathy. People do well even if there is no benefit for them. Goodness needs no ulterior motives, although in some instances it does happen.

Thirdly, empathy is necessary for cooperation. Cooperation implies people working together and interacting in a manner beneficial to all. In the absence of empathy, individuals fail to work together for the good of everyone because they cannot empathize with the situations of others. McCarthy (1993) observes that in trying to understand others, people recall moments of others understanding them. However, the nature of the patriarchal society that is portrayed in the novels illustrates that women rarely receive empathy from men. What remains

true is that the novel allows the reader to experience the condition of being a woman in a violent and intolerant environment. The hope is that after understanding that experience, the appeal for recognition of the challenges that confront women will receive a favourable response from the systems of justice

Conclusion

Although emerging technologies have made life and communication easier, they come with dangers such as TFGBV, especially for women and girls. There needs to be urgent attention to address the universal scourge of TFGBV, especially on social media applications such as Facebook. From the novels, it is apparent that society needs to practice empathy towards each other, especially women and girls who confront intersecting challenges on and offline. There is a need to pursue a feminist jurisprudence that can address the gaps in law and human rights. The gaps should address the patriarchal nature of the law that seemingly fails to provide justice to women and girls in TFGBV. According to this study, empathy is critical to finding a human connection. The root of TFGBV, ultimately GBV, is the absence of human connection. The absence of viewing women or others as human, therefore deserving of respect and dignity both on and offline.

One of the challenges of this study is the struggle to find as many novels as possible from across all continents that will provide a wide perception of the challenges that confront women on Facebook. The novels, women must write them, with a focus on female characters and their experiences on Facebook. Novels are fictional and likely

regarded as far removed from the reality of women and Facebook. Other schools of thought do not agree that literature can inform the law. They consider law and literature two distinct fields that are better off separated. Therefore, the law and literature approach can be open to argument. Offering empathy as a possible solution to TFGBV may be problematic as it is something that is abstract and immeasurable. Empathy is imperfect but can alter the way society thinks about women and TFGBV.

Further studies are needed to investigate if law and human rights can address the ideas of anonymity and TFGBV. In other words, while anonymity has its advantages, even for women online, it hinders bringing perpetrators of TFGBV to justice. At what point do anonymity and privacy meet without hindering justice? In addition, future studies must investigate TFGBV in developing countries. There seems to be little information on Africa, African women, and TFGBV. Does access to the internet determine the prevalence of TFGBV in developing countries?

References

- Bailey, J., Craig, C., Dunn, S., & Lawrence, S. (2022). Reframing technology-facilitated gender-based violence at the intersections of law and society. *Canadian Journal of Law and Technology*, 19(2), 209–216. <https://doi.org/10.1108/978-1-83982-848-520211001>
- Baxi, U. (1998). Voices of suffering and the future of human rights. In *Transnational Law and Contemporary Problems*, 8, 125–169. https://www.academia.edu/8134847/1998_Voices_of_Suffering_and_the_Future_of_Human_Rights_Transnational_Law_and_Contemporary_Problems_Fall_1998_pp_125_169
- Bethlehem, R. A. I., Allison, C., Van Andel, E. M., Coles, A. I., Neil, K., & Baron-Cohen, S. (2017). Does empathy predict altruism in the wild? *Social Neuroscience*, 12(6), 743–750. <https://doi.org/10.1080/17470919.2016.1249944>
- Cotterrell, R. (1991). *The sociology of law: An introduction* (2nd ed.). Butterworths.
- Kaspersky. (2021). Digital stalking in relationships report. https://media.kasperskydaily.com/wpcontent/uploads/sites/86/2021/11/17164103/Kaspersky_Digital-stalking-in-relationships_Report_FINAL.pdf
- Donnelly, J. (1992). Human rights in the New World Order. *World Policy Journal*, 9(2), 249–277.
- Douzinas, C., & Gearey, A. (2005). *Critical jurisprudence: The political philosophy of justice*. Bloomsbury Academic.
- Douzinas, C. (2000). Human rights and postmodern utopia. *Law and Critique*, 11(3), 219–240. <https://doi.org/10.1023/A:1008935817302>
- Douzinas, C. (2007). *Human rights and empire: The political philosophy of cosmopolitanism*. Routledge-Cavendish.
- Dunn, S. (2020). Technology-facilitated gender-based violence: An overview. Centre for International Governance Innovation. <https://www.cigionline.org/publications/technology-facilitated-gender-based-violence-overview/>

- Fitzpatrick, P. (2005). *The mythology of modern law*. Routledge.
- Fitzpatrick, P. (2001). *Modernity and the grounds of law*. Cambridge University Press.
- Goodrich, P. (1996). *Law in the courts of love: Literature and other minor jurisprudences*. Routledge.
- Jurecic, A. (2011). Empathy and the critic. *College English*, 74(1), 10–27.
- Kehinde, A., & Mbipom, J. E. (2011). Discovery, assertion, and self-realization in recent Nigerian migrant feminist fiction: The example of Sefi Atta's *Everything Good Will Come*. *African Nebula*, 3, 62–78.
- Mar, R. A., Oatley, K., & Peterson, J. B. (2009). Exploring the link between reading fiction and empathy: Ruling out individual differences and examining outcomes. *Communications*, 34(4), 407–428. <https://doi.org/10.1515/COMM.2009.025>
- Mathewson, G. (2004). Challenging the court: Charles Chesnutt's *Marrow of Tradition*. In M. Meyer (Ed.), *Literature and law* (pp. 219–240). Rodopi. https://doi.org/10.1163/9789401201315_015
- McCarthy, M. (1993). Empathy amid diversity: Problems and possibilities. *Journal of Pastoral Theology*, 3(1), 15–28.
- O'Neil, L. (2015). *Asking for it*. Hachette.
- Plan International. (2020). Free to be online? Girls' and young women's experiences of online harassment. <https://plan-international.org/publications/free-to-be-online/>
- Rajan, R. S., & Park, Y. (2005). Postcolonial feminism or postcolonialism and feminism. In H. Schwarz & S. Ray (Eds.), *A companion to postcolonial studies* (pp. 53–71). Blackwell.
- Rawoski, S. R. (1983). The novel of awakening. In E. Hirsch (Ed.), *The voyage in: Fictions of female development* (pp. 48–67).
- Schwarz, H., & Ray, S. (Eds.). (2008). *A companion to postcolonial studies*. Wiley-Blackwell.

Seagull, E. A. (2018). *Social empathy: The art of understanding others*. Columbia University Press.

Stacy, H. M. (2001). *Postmodernism and law: Jurisprudence in a fragmenting world*. Ashgate.

Stansfield, J., & Bunce, L. (2014). The relationship between empathy and reading fiction: Separate roles for cognitive and affective components. *Journal of European Psychology Students*, 9(4), 3–7. <https://doi.org/10.5334/jeps.ca>

Sundaramoorthy, L. (2016). Is the idea of human rights a universal concept? *Merici*, 2(1), 23–29. <https://doi.org/10.22459/M.02.2016.03>

UN Women. (2020). COVID-19 and ending violence against women and girls: Addressing the shadow pandemic. *Policy Brief No. 17*. <https://www.unwomen.org/en/digital-library/publications/2020/06/policy-brief-covid-19-and-violence-against-women-and-girls-addressing-the-shadow-pandemic>

Van Schalkwyk, S. (2018). *Narrative landscapes of female sexuality in Africa: Collective stories of trauma and transition*. Springer.

Ward, I. (2004a). *Law and literature: Possibilities and perspectives*. Cambridge University Press.

Ward, I. (2004b). *Introduction to critical legal theory*. Routledge-Cavendish.

Cybersecurity Culture from the Perspective of Social Cognitive Theory: Examining the Relevance of the Reciprocal Determinism Model

Sladana Ćurčić

Introduction

With the development of sophisticated technologies and the digitization of society, cyber security has become an important issue at the national, societal, and individual levels. In other words, as a result of the growth and spreading of computer networks into whole aspects of life, cyber security is no longer a “special” and “extraordinary” issue but a kind of security much embedded into the everyday realities of people, states, and non-state actors (Cavelty, 2010a). While an increasing number of countries are adopting cyber security strategies, each individual must develop their own strategy to protect herself or himself against cybersecurity challenges in everyday

Sladana Ćurčić, Junior Research Assistant, Institute of European Studies, Belgrade
ORCID: 0000-0001-8276-0983, email: sladjana.curcic@ies.rs

DOI: <https://doi.org/10.5281/zenodo.14204694>

life. Cybersecurity culture is considered that kind of individual strategy. This is why this relatively recent concept in academic literature is gaining more and more prominence in practice and academic research. Its emergence induced academic curiosity about an adequate theoretical perspective for researching and explaining the concept. The reciprocal determinism model of security culture (RD model hereinafter), derived from Social Cognitive Theory (SCT), appears to be one of the potentially valuable frameworks. Its significance relies on the interdependence of three key factors: organizational, psychological, and behavioral, which capture the usual elements of security culture in general and, so, cyber security culture. This paper seeks to, through the systematization of the application of the RD model in cyber security culture research, indicate its importance and potential scope for a deeper understanding of this relatively new concept. At the same time, adequate theoretical framing and a clearer definition of the concept elements will contribute to the possibility of developing and improving cyber security culture on a practical level as well.

The improvement of cybersecurity culture becomes more and more important when we consider the potential cyber risks to which we are exposed on a daily basis, which most often come from the online sphere, which has become an integral part of our lives, whether through work, study, or entertainment. This also applies when it comes to the functioning of critical infrastructure, which is impossible without the Internet and a complex system of electronic communication connections (e.g., healthcare facilities, energy plants, banks, traffic infrastructure, etc.). In all these cybersecurity considerations, humans are the weakest link, as they could be potential targets of cyber attacks or even unknowingly participate in a cyber attack (Von Solms and Van Niekerk, 2013). Positioning this issue in the

broader context of cyber security, it could be said that the common characteristic of different definitions of cybersecurity is that they try to be holistic and to include human aspects (The International Telecommunications Union (ITU), 2008; Von Solms and Van Niekerk, 2013; Schatz, Bashroush and Wall, 2017), which is actually the main difference between the notion of cyber security and information security. Speaking about these differences, it should be emphasized that cyber security goes beyond the boundaries of information security to include not only the protection of information resources but also that of other assets, including the person himself or herself (Von Solms and Van Niekerk, 2013).

This fact that the human factor is decisive for cyber security—people's behavior, awareness, knowledge, beliefs, and attitudes made socio-psychological theories often used in cyber security culture research and cyber security in general. For example, some researchers combined institutional theory and protection motivation theory in exploring security awareness through the lenses of cybersecurity culture (Andronache, 2021). Other authors explored the impact of protection motivation theory and general deterrence theory on active cyber defense (White, 2017). Technology threat avoidance theory (TTAT) is also a valuable framework for understanding individual threat avoidance motivation and behavior, which is a critical component in designing effective cybersecurity solutions for both users and organizations (Carpenter et al., 2019). So it is also used in researching the coping responses of employees to a cyberattack, especially concerning human and emotional aspects of cybersecurity (Stacey et al., 2021).

It can be concluded that a specific field such as cyber security very often requires the integration of theories closely related to cyber

security and socio-psychological theories, which elucidate that very important human aspect. That's why many authors emphasize the need for an interdisciplinary approach in cybersecurity and cybersecurity culture research, combining information systems theories with social science theories like the Theory of Planned Behavior (TPB), Rational Choice Theory (RCT), Protection Motivation Theory (PMT), Theory of Reasoned Action (TRA), and Social-Cognitive Theory (SCT) (Hanna, 2020; Maalem Lahcen et al., 2020; Rohan et al., 2021; Ogden, 2021; Georgiadou et al., 2022). Those are, of course, valuable insights and recommendations for future research, but the state-of-the-art, as evidenced in the literature, shows that researchers usually study cybersecurity culture by assessing its level through surveys without relying on a concrete theoretical framework. For now, SCT, because of the compatibility of its RD model (a triad of factors: organizational (environmental), psychological (cognitive), and behavioral) with usual cybersecurity culture elements, appears to be the most adequate to fulfill this gap. From these insights arises the research question of the paper: What is the theoretical relevance of cybersecurity culture research through the RD model, and what are the potentials for practical improvement of cybersecurity culture by applying this model?

In order to answer the research question, first the theoretical basis in the subject field will be analyzed: academic thematization of cybersecurity, current knowledge, and scope of research on the concept of cybersecurity culture. Then, the basics of SCT and the RD will be presented, along with an overview of their applications in security culture research. Based on the summarization of the literature review and content analysis of existing cybersecurity culture models, the central part of the paper will provide an overview of the cybersecurity culture models based on reciprocal determinism.

Cyber Security: Theoretical Considerations

Social, political, and technological developments that transform our daily lives but also shape the reality of security are, at the same time, key factors in the growing importance of the field of cyber security. Following the introduction of personal computers in the 1980s, when the term cyber-crime was born with the development of cyber-counter-culture,” and then the 1980s and 1990s, when the debate on cyber threats was influenced by the post-Cold War strategic context, it is clear that cyber security actually evolved in line with the information revolution (Cavelty, 2010a). Today, as a result of the growth and spreading of computer networks into whole aspects of life, cyber security is no longer a “special” and “extraordinary” issue but a kind of security much embedded into the everyday routines of people, states, and non-state actors (Cavelty, 2010a).

Speaking specifically about the definition of the term cyber security, many authors tried to provide a more complete and clearer definition by realizing the shortcomings of the existing definitions. Some of them find that the term is used broadly and that its definitions are highly variable, context-bound, often subjective, and, at times, uninformative (Craigén, Diakun-Thibault, and Purse, 2014). They argue that the absence of a definition that captures the multidimensionality and interdisciplinarity of cybersecurity potentially impedes technological and scientific advances by reinforcing the predominantly technical view of cybersecurity (Craigén, Diakun-Thibault, and Purse). So, they provided, as they say, an inclusive, meaningful, and unifying definition

that could fill this theoretical gap: “Cybersecurity is the organization and collection of resources, processes, and structures used to protect cyberspace and cyberspace-enabled systems from occurrences that misalign *de jure* from *de facto* property rights” (Craig, Diakun-Thibault, and Purse, p. 17). Others claim that the goal of cyber security is to secure those that function in cyberspace, whether individuals, organizations, critical national infrastructure, societies, or nations. Relying on that position, they define cyber security “as the protection of cyberspace itself, the electronic information, the ICTs that support cyberspace, and the users of cyberspace in their personal, societal, and national capacities, including any of their interests, either tangible or intangible, that are vulnerable to attacks originating in cyberspace” (Von Solms and Van Niekerk, 2013, p. 101).

A group of authors also conducted an interesting review and content analysis of cyber security definitions classified into three main categories: industry definitions, government and nation-state definitions, and academic definitions, and as a result, they created a new definition comprising the key terms identified (Schatz, Bashroush, and Wall, 2017). According to them, “cyber security refers to the approach and actions associated with security risk management processes followed by organizations and states to protect confidentiality, integrity, and availability of data and assets used in cyber space. The concept includes guidelines, policies, and collections of safeguards, technologies, tools, and training to provide the best protection for the state of the cyber environment and its users” (Schatz, Bashroush, and Wall, 2017, p. 66). In addition to the academic definitions, it should also be mentioned the approach that the International Telecommunications Union (ITU) advocates. It defines cyber security as “a collection of tools, policies, security concepts, security safeguards, guidelines,

risk management approaches, actions, training, best practices, assurance, and technologies that can be used to protect the cyber environment, organization, and user's assets" (ITU, 2008, p.). It could be said that the common characteristic of the above definitions is that they try to be holistic and to include human aspects, which is actually the main difference between the notions of cyber security and information security.

Regarding these differences, it should be emphasized that cyber security goes beyond the boundaries of information security to include not only the protection of information resources but also that of other assets, including the person himself or herself. So, in cyber security, humans are considered potential targets of cyber attacks or even unknowingly participating in a cyber attack, while in information security, reference to the human factor usually relates to the role(s) of humans in the security process (Von Solms and Van Niekerk, 2013). But what is common both to information and cyber security measures is the goal of ensuring confidentiality, integrity, and availability of information (Stoneburner, 2001; ITU, 2008). Briefly, it should also be said that the three main types of cyber threats are cybercrime, cyberterrorism, and cyberwar (Cavelty, 2010b), but cyberbullying could also be added (Kaur and Ramkumar, 2022). Then, regarding cybercrime, according to the UK's Crown Prosecution Service (CPS), there are two broad categories: cyber-dependent (e.g., hacking, malware, denial of service) and cyber-enabled crimes (financial fraud, phishing, pharming, extortion) (CPS, 2019).

The Concept of Cybersecurity Culture

Academic literature, but also the everyday reality we live in, shows that changes and development of the concept of security culture are conditioned by changes in security reality. Transformations in the understanding and manifestations of security culture, especially happen when they are necessary for serious social adjustments (Buluc, Lungu, and Deac, 2018). In this sense, the Chernobyl nuclear disaster, the terrorist attack on the USA in 2001, the development of information technologies, but also pandemics of infectious diseases, are events and processes that represent important milestones in the development of security culture, both as a concept and as a practice. So, the still-increasing development of information and the cyber sphere refers to the security field, in which cyber security culture has gained great importance.

As it was previously elaborated that cyber security is a broader concept than information security, the same can be said for the relationship between cyber and information security cultures. Also, it should be said that the concept of information security culture has been studied for a significantly longer period (Uchendu et al., 2021) and is more established (Reid and Van Niekerk, 2014). Therefore, it is understandable that there is much more knowledge about information security culture, which is sometimes uncritically applied to cyber security culture. But what is the common acknowledgment of many authors in the elucidation of these differences and overlappings is the emphasizing of the human factor as indispensable while studying cyber security culture (Gcaza et al., 2017; Georgiadou et al., 2022; Mwim and

Mtsweni, 2022). An additional explanation of the differences is that information security culture emphasizes behaviors that comply with information security policy, but a cybersecurity culture includes not only compliance with policy but also personal involvement in organizational cyber safety (Huang and Pearlson, 2019). In order to give more clarity, these authors offer a definition of organizational cybersecurity culture, which means “the beliefs, values, and attitudes that drive employee behaviors to protect and defend the organization from cyber attacks” (Huang and Pearlson, 2019, p. 6399). However, some authors still think that cybersecurity culture is an ill-defined problem, especially because of the lack of clarity in the academic community about the definition of cybersecurity culture, the lack of research focusing on measuring cybersecurity culture, and consequently, the absence of a solution for cultivating cybersecurity culture (Gcaza and Solms, 2017). Regardless, what is indisputable is that cybersecurity culture is necessary for encouraging acceptable user behavior in the reality of cyberspace (ITU, 2008). Additionally, other authors think that the process of raising a cyber security culture could function as a self-learning process for organizations by producing valuable insights regarding organizational values, norms, etc. (Karyida, 2017).

In the understanding of cyber security culture, some authors adhere to Schein’s model of organizational culture and argue that cyber security culture consists of several layers: artifacts, espoused values, tacit assumptions, and their respective contents (Reegrđ, Blackett, and Katta, 2019). Actually, these authors are speaking about organizational cyber security culture (according to their understanding of cybersecurity culture as a sub-component of organizational culture), but they remind us that cybersecurity extends beyond the organizational boundaries, and therefore, research on the influence of factors

external to the organization on cybersecurity culture is needed. Further, they state that key practices for developing cybersecurity culture resemble those highlighted in the literature on safety culture: management support, policy, awareness and training, involvement and communication, and learning from experience (Ree-grd, Blackett, and Katta, 2019). Other authors also follow this line of thinking, relying on Schein's model (and adding knowledge as a fourth level) in defining firstly information security culture and then arguing that cybersecurity culture consists of the same components but with a bit of different content due to the different contexts in which the cultures foster (Reid and Van Niekerk, 2014). More specifically, information security culture is cultivated and managed within insulated organizational contexts, which are relatively well controlled environments with relatively predictable user behavior, activity, and profile sets. On the other side, cyber-security culture would be cultivated within a societal environment that would likely be less controlled (Reid and Van Niekerk, 2014).

A comprehensive model of cyber security culture entails two levels: organizational and individual, divided into different dimensions. The organizational level is divided into the following dimensions: assets, continuity, access, and trust; operations; defense; and security governance. Individual dimensions are attitude, awareness, behavior, and competency. Each dimension consists of domains with distinctive application areas and quantifiable indicators (Georgiadou et al., 2020). For Da Veiga (2016), cybersecurity culture should ideally be fostered at all levels, including individual, organizational, national, and international levels. In this context, it is an interesting attempt to define cybersecurity culture from the perspective of strategic culture (Tziarras, 2014). Actually, this author starts from the broader framework

of the changing concept of security and focuses on cultivating cybersecurity culture on a global level through multi-leveled collaboration. So, he defines a security culture of multileveled cybersecurity “as a body of collective (non-state, sub-national, and national) attitudes, patterns of behavior, and beliefs, as well as conceptions of (cyber) security, shaped based on the need to secure multiple referent objects against various cyber threats, which would influence cybersecurity strategies” (Tziarras, 2014, p. 330). Regarding the previous criticism about the lack of methodological solutions in measuring cybersecurity culture, the meta-analysis of cybersecurity culture research shows that not all studies aim to provide a method in which security culture can be measured or assessed but to present a framework or approach in which a security culture can be built and maintained (Uchendu et al., 2021). Moreover, this analysis found that questionnaires, surveys, and interviews are widely used in measuring the level of cybersecurity culture (Uchendu et al., 2021). Of course, as Uchendu and colleagues emphasize, any proposal of a framework at a conceptual or theoretical level should be carefully considered regarding the possibility of assessment in practice. So, they add that there is a need for testing and evaluation of proposed security culture approaches and frameworks to provide real-world evidence of their efficacy (Uchendu et al., 2021).

Finally, it can be said that the operationalization of cybersecurity culture also depends on the specific perspective of researchers and contexts of application, but as previous analysis has shown, most authors agree on some key elements of the concept. Certainly, the integration and communication of different disciplines will potentially contribute to a better and more precise determination of the concept.

Reciprocal Determinism Model of Safety Culture¹

The reciprocal determinism model originates from Albert Bandura's Social Cognitive Theory (SCT), which subscribes to a model of emergent interactive agency (1986). Bandura describes *triadic reciprocal causation* or *triadic reciprocal determinism* as a model in which "action, cognitive, affective, and other personal factors, and environmental events all operate as interacting determinants" (1989, p. 1175). Respectively, he explains it by the fact that persons are neither autonomous agents nor simply mechanical conveyers of animating environmental influences. Rather, they make a causal contribution to their own motivation and action within a system of triadic reciprocal causation (1989, p. 1175). Even though these factors operate as interacting determinants that influence each other bidirectionally, Bandura empathizes that it doesn't mean that the different sources of influence are of equal strength, nor do the reciprocal influences all occur simultaneously. Actually, some influences may be stronger

-
- ¹ In his research on reciprocal determinism, Cooper used the term *safety culture*, so original term was kept, but when referring to the cyber context, it will be used the term *security culture*. Namely, as a key difference between the concepts of safety and security, and thus safety and security culture, the literature emphasizes the origin of undesirable events, which are unintentional in the field of safety (incidents, natural disasters), and intentional in the field of security (actions that are designed and planned to cause harm, e.g. terrorist attacks, cyber threats) (Stanarević, 2012; Mattila, 2013; Sas et al., 2021). However, the greatest similarity is that both safety and security are part of the overall culture of the organization (Sas et al., 2021).

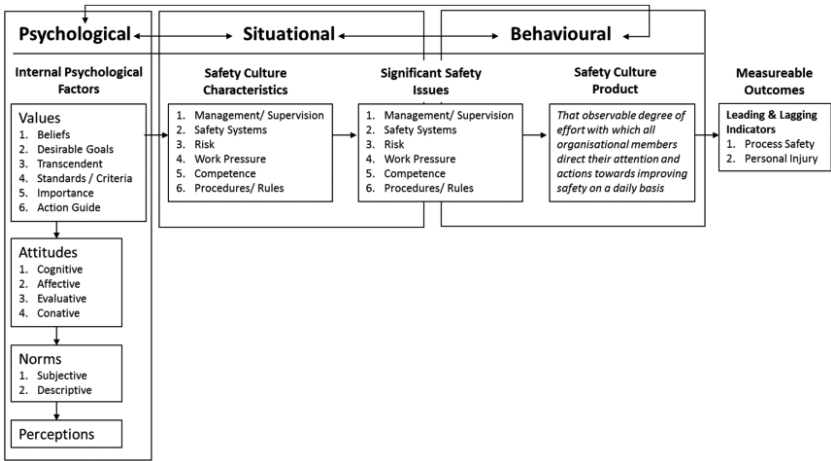
than others, and it takes time for a causal factor to exert its influence and activate reciprocal influences (Bandura, 1999). In other words, this bidirectionality of influence means that “people are both products and producers of their environment” (Wood and Bandura, 1989, p. 362).

This interdependence of behavioral, psychological, and social factors was found to be insightful by some authors as a theoretical and practical framework to measure and analyze safety culture. Thus, Bandura’s model of reciprocal determinism was adapted by Cooper (1994, 1997a, b, 2000, 2002, 2016, 2018), to research safety culture, resulting in what is called the reciprocal determinism model of safety culture. It contains three elements, which encompass subjective internal psychological factors, observable ongoing safety-related behaviors, and objective situational features (Cooper, 2000). As can be concluded from the Picture 1, Cooper defines safety culture as “the product of multiple goal-directed interactions between people (psychological), jobs (behavioral), and organizations (situational)”² and actually sees it as dynamic reciprocal relationships between those elements (Cooper, 2002, p. 32). So, this is exactly where the compatibility between the

-
- 2 Originally, Cooper’s reciprocal safety culture model implied this sequence: psychological aspects—behavioral aspects—situational aspects (Cooper, 2000). Later, as can be seen from the attached picture, the revised model implied behavior at the end. This resulted from his insight that changes in certain procedures and rules (organizational factors) will affect behavior change, which will then increase safety performance and reduce incidents (which will in turn positively affect psychological factors), rather than the changes in core basic assumptions and attitudes (psychological factors) will lead to a behavior change (Cooper, 2016). In his words “The principle is to optimise the situation to optimise the behaviour. In turn, as the desired behaviours become habitual, the various psychological factors will become more positive” (Cooper, 2018, p. 51).

reciprocal determinism model and the nature of safety culture lies. Additionally, the potency of the RD model for analyzing safety culture, according to Cooper, also lies in the fact that it provides a triangulation methodology that allows researchers to holistically examine safety culture as a complex, multi-faceted construct. Thus, the reciprocal relationships between psychological, behavioral, and situational factors can be examined with a view to establishing antecedents, behaviors, and consequences within specific contexts (Cooper, 2000). Specifically, speaking about measurement, he emphasizes that internal psychological factors (i.e., attitudes and perceptions) are assessed via safety climate questionnaires, safety-related behavior is assessed via checklists developed as a part of behavioral safety initiatives, and situational features are assessed via safety management system audits and inspections (Cooper, 2000).

Picture 1: *Cooper’s reciprocal safety culture model (2016).*



Since the core of the safety culture construct is about “proactively managing safety, thinking positively about safety, and behaving

safely” (Cooper, 2016, p. 4), deepening the theoretical basis of the concept and ways of its practical improvement is gaining more and more importance. Many authors recognized the potential of Cooper’s RD model of safety culture and applied it to safety culture research in various contexts, mostly in high-risk industries. Based on Cooper’s safety culture model, Choudhry, Fang, and Mohamed (2007a) have developed a conceptual model of construction safety culture. The model is anchored in three fundamental conceptual categories, namely, safety climate, behavior-based safety, and safety system. The model offers an integrative framework and can be applied to construction projects to maintain and improve construction site safety. It reveals that unsafe conditions can be traced through the site safety implementation and can be rectified (Choudhry, Fang, and Mohamed, 2007b). Consequently, it appeared that Cooper’s model could really be fruitfully applied in the construction industry, so other authors also applied it, trying to even revise and improve the previously offered model by Choudhry, Fang, and Mohamed (Ismail et al., 2009; Alasamri, Chrisp, and Bowles, 2012).

Based upon a modified version of Cooper’s reciprocal determinism model, the study on safety culture in the fire service uses two sets of exogenous variables, labeled Safety Management System and Safety Related Behaviors, to explain a dependent variable called Organizational Safety Climate (Pessemier and England, 2012). Some authors researched safety behaviors among firefighters and the safety culture of the department, relying on SCT and reciprocal determinism but without special reference to Cooper’s model of safety culture (Freaney, 2011). Of course, there are authors who didn’t explicitly use Cooper’s model of safety culture but implicitly supported it and indicated its relevance by studying behavioral, organizational, and

psychological elements of safety culture in their studies, predominantly related to the Occupational Health and Safety (OHS) domain (Fernandez Muniz et al., 2007; Lefranc et al., 2012).

In addition to the evident academic curiosity about the RD model of safety culture, its practical relevance is also reflected in the fact that this exact approach has been officially adopted as a standard by the American Petroleum Institute and the American National Standards Institute (Cooper, 2018). Additionally, while comparing other safety culture models and finding similarities and differences with the RD model, Cooper pointed out that it would appear that the reciprocal model has some general applicability, particularly as it incorporates the underlying features of existing safety culture models and allows both the qualitative and quantitative aspects of safety culture to be explored (Cooper, 2000). Moreover, as he noted, the RD model of safety culture encompasses some models, like the “Total Safety Culture’ model. So, he concludes that the RD model of safety culture has the potential to facilitate future meta-analyses of safety culture research (Cooper, 2000).

**Research Findings:
Application of the
Reciprocal Determinism
Model in Cybersecurity
Culture Research**

Table 1 represents the results of the literature review, which aimed to find cybersecurity culture models based on the social cognitive

framework, specifically the RD model of security culture. Analysis revealed that there is no cybersecurity culture model directly based on the RD model, but there are several cybersecurity culture models that could be labeled as partially consistent with the RD model. This conclusion is derived from the essence of the presented cybersecurity culture models. Specifically, elements of these models are reorganized according to three key aspects of the RD model. Of course, these models do not exhaust all the elements contained in the RD model, and they differ in their terms and meanings, which is expected as a reflection of the specificity of the field of cyber security. What is important is that the elements of the listed cyber security culture models, according to the logic of their meaning, can be classified as psychological, organizational, or behavioral factors. This also means that all authors (implicitly or explicitly through the models used) recognize that cybersecurity culture necessarily implies the interaction of these three levels. Also, as could be noted, a group of almost the same authors explored the same model of cybersecurity culture in different sectors: healthcare systems, remote working, critical infrastructure, as well as the energy sector. This is also important in an effort to empirically test the model and determine its applicability in different contexts.

These results could be considered as a part of the evidence about the relevance of the RD model in cybersecurity culture research. The general use of Social Cognitive Theory in cybersecurity culture research also gives indirect and valuable insights about the potency of the RD model. For example, SCT has been used to study individuals' responses to specific cyber security threats, like phishing. Specifically, SCT was used to examine the influence of the triadic factors of perceived self-efficacy toward antiphishing behaviors, expected negative

Table 1: Cybersecurity culture models based on the reciprocal determinism model.

Authors	Cybersecurity culture model	Industry/field	Consistent with RD model of security culture
Huang and Pearlson (2019)	External influences Societal cybersecurity culture External rules and regulations Peer institutions Organizational mechanisms Cybersecurity culture leadership Performance evaluations Rewards and punishments Organizational learning Cybersecurity training Communications channel Beliefs Attitudes Values (at leadership, group and individual level) Behaviors In-role cybersecurity behavior Extra-role cybersecurity behavior	Organizational context	Partially Psychological aspect Beliefs Attitudes Values Situational External influences Organizational mechanisms Behavioral In-role cybersecurity behavior Extra-role cybersecurity behavior

Authors	Cybersecurity culture model	Industry/field	Consistent with RD model of security culture
Kabanda and Chingoriwo (2021)	Five pillars: Shared national cybersecurity vision and strategy ICTs and related infrastructure Cybersecurity legislation Education and awareness Technology framework and skills	General	Partially Psychological aspect Awareness Situational Shared national cybersecurity vision and strategy ICTs and related infrastructure Cybersecurity legislation Technology framework Behavioral Education Skills
Georgiadou, Mouzakitis and Askounis (2021)	Organizational level Assets Continuity Access and trust Operations Security governance Defense Individual level Attitude Competency Behavior Awareness	General	Partially Psychological aspect Attitude Awareness Situational Assets Continuity Access and trust Operations Security governance Defense Behavioral Competency Behavior
Georgiadou, Mouzakitis and Askounis (2021)		Critical infrastructure	
Gioulekas et al. (2022)		Healthcare	
Georgiadou, Michalitsi-Psarrou and Askounis (2022)		Energy sector	
Georgiadou, Mouzakitis and Askounis (2022)		Remote working	

outcomes from reporting spear phishing emails, and cybersecurity self-monitoring on individuals' likelihood of reporting spear phishing emails. By adding the construct of cyber risk beliefs (CRBs) into the SCT framework, the research model explained the motivational factors that inhibit the reporting of spear phishing (Kwak et al., 2020). In a similar way, Ogden (2021) used Social Cognitive Theory to identify factors that influence human security behavior and best practices for developing a cybersecurity culture while focusing on the relationships between 1) environmental factors, 2) cognitive factors (personal factors), and 3) their mediating effects on behaviors. Research has shown that social proximity, subjective norms, descriptive norms (environmental factors), self-efficacy, knowledge, and experience (cognitive factors) are contributing factors that influence an individual to perform cyber-secure behaviors (Ogden, 2021).

Additionally, SCT is successfully used in exploring cybersecurity awareness, which is closely related to cybersecurity culture (Hanna, 2020). Components of SCT were demonstrated to be important concepts for promoting and fostering desired cybersecurity behavior in organizations across industry domains. Specifically, by assuming that employees are agents and their behavior and learning are directed by the triadic reciprocal determinism model, organizational IT leaders can incorporate adequate SETA strategies (security education, training, and awareness) and foster environments that promote cybersecurity behavior and improve cybersecurity culture. Consequently, improved cybersecurity culture and learned cybersecurity practices could also be applied in the home context (Hanna, 2020).

From the points above, the authors' efforts in searching for the most adequate framework for cybersecurity culture are evident. Moreover,

almost all the analyzed research involved empirical validation of the specific cybersecurity culture model. In this sense, connecting theoretical knowledge with practical solutions for the development and improvement of cyber security culture will undoubtedly contribute to this field. Generally speaking, experts in this field always remind us that collaboration and constant communication between researchers, practitioners, academia, and industry are of great importance for improving the body of knowledge on cybersecurity culture and security culture in general (Uchendu et al., 2021; Cooper, 2016).

Conclusion

It is obvious that the application of the RD model in the study of cyber security culture has not yet caught on, as is the case with security culture in general. However, systematization of the existing cybersecurity culture models revealed that the most common elements that predominantly determine cybersecurity culture are organizational and individual (which include behavioral in some cases). From a practical point of view, this should remind us of the role of humans and human factors in the cybersphere as an emerging part of our everyday lives. So, the current research could be relevant both for organizations and individuals by emphasizing that developing and improving the cybersecurity culture is the best strategy to ensure cyber security in such a fragile digital world, which faces us with threats at work, at home, at school, etc.

Addressing the research question, the potency of the RD model in cybersecurity culture research is recognized in the presence of three

broader aspects (psychological, organizational, and behavioral) of security culture in all selected models, with varying elements. Moreover, this diversity of elements of cybersecurity culture actually speaks about the complexity of the concept and can help researchers consider what levels and elements should be taken into account in order to study and understand cybersecurity culture in depth.

Relevance of the RD model is also evident in the application of SCT in a broader cybersecurity context. This lies in the fact that SCT is the well-established, validated, and mostly used theory on behavior change, while on the other hand, understanding the human factor, i.e., human behavior and its constant interaction with the environment, is becoming increasingly important for cybersecurity culture research and practice. Thus, offering adequate theoretical framing of the cybersecurity concept, which is a precondition for its assessment and practical improvement, is what makes the RD model and, broader, SCT, worth the attention of researchers in the study of cybersecurity culture.

Moreover, engaging theoretical frameworks from psychology, sociology, and organizational behavior in cybersecurity culture research is in favor of interdisciplinarity in the research and is valuable for “both sides”. Namely, in such a way, these theories could be tested in a specific context and thus improve their value, while at the same time providing insight into broader aspects of cybersecurity cuThe relevanceat are not covered by “more traditional” theories in this field (e.g., information systems theories).

Besides useful insights from this research, it should be said that the main limitation lies in the fact that all selected models are partially consistent with the RD model (based on the author’s reorganization of

the elements according to three key aspects). The main reason for this could be seen in the state-of-the-art: the concept of cybersecurity culture is relatively recent and does not have a long history of study. Although cybersecurity is a growing field, it is not surprising that no model directly based on the RD model has been found, as researchers are still exploring the field and searching for the most adequate theoretical underpinning and conceptual framework for cybersecurity culture. So, all findings and conclusions from this paper could serve as an impetus for the future wider application of the RD model in cybersecurity culture research. One of the recommendations to the researchers refers to the testing of existing models of cybersecurity culture based on the RD model in order to empirically verify and revise them and thus come to new knowledge, which would be integrated back into the theory. Of course, more replicative studies of cybersecurity culture, based both on SCT and other theories, from a broader perspective will contribute to finding the most suitable model of cybersecurity culture.

References

- Andronache, A. (2021). Increasing security awareness through lenses of cybersecurity culture. *Journal of Information Systems and Operations Management*, 15(1), 7–22.
- Alasamri, H., Chrisp, M. T., & Bowles, G. (2012). A framework for enhancing and improving the safety culture on Saudi construction sites. In S. D. Smith (Ed.), *Proceedings of the 28th Annual ARCOM Conference* (pp. 475–485). Association of Researchers in Construction Management. https://www.arcom.ac.uk/docs/proceedings/ar2012-0475-0485_Alasamri_Chrip_Bowles.pdf
- Bandura, A. (1986). *Social foundations of thought and action: A social-cognitive view*. Prentice Hall.
- Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, 44(9), 1175–1184. <https://doi.org/10.1037/0003-066X.44.9.1175>
- Bandura, A. (1999). Social cognitive theory: An agentic perspective. *Asian Journal of Social Psychology*, 2(1), 21–41. <https://doi.org/10.1111/1467-839X.00024>
- Bandura, A., & Wood, R. E. (1989). Effect of perceived controllability and performance standards on self-regulation of complex decision-making. *Journal of Personality and Social Psychology*, 56(5), 805–814. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=a5989e-20385c9972919084a9ed357b103625aeb9>
- Buluc, R., Lungu, C., & Deac, I. (2018). Perceptions on security culture. *Redefining Community in Intercultural Context*, 7(1), 149–156. <https://www.cceol.com/search/article-detail?id=874195>
- Carpenter, D., Young, D. K., Barrett, P., & McLeod, A. J. (2019). Refining technology threat avoidance theory. *Communications of the Association for Information Systems*, 44, 380–407. <https://doi.org/10.17705/1CAIS.04422>
- Cavelty, M. D. (2010a). Cyber-security. In P. J. Burgess (Ed.), *The Routledge Handbook of New Security Studies* (pp. 155–162). Routledge. <https://doi.org/10.4324/9780203859483>

Cavelty, M. D. (2010b). Cyber-threats. In M. D. Cavelty & V. Mauer (Eds.), *The Routledge Handbook of Security Studies* (pp. 180–189). Routledge.

Choudhry, R. M., Fang, D., & Mohamed, S. (2007a). Developing a model of construction safety culture. *Journal of Management in Engineering*, 23(4), 207–212. [https://doi.org/10.1061/\(ASCE\)0742-597X\(2007\)23:4\(207\)](https://doi.org/10.1061/(ASCE)0742-597X(2007)23:4(207))

Choudhry, R. M., Fang, D., & Mohamed, S. (2007b). The nature of safety culture: A survey of the state-of-the-art. *Safety Science*, 45(10), 993–1012. <https://doi.org/10.1016/j.ssci.2006.09.003>

Cooper, M. D. (1994). Implementing the behaviour-based approach to safety: A practical guide. *The Health and Safety Practitioner*, 12(11), 18–23. http://behavioural-safety.com/articles/Implementing_Behavior-Based_Safety_a_practical_guide.pdf

Cooper, M. D. (1997a). Evidence from safety culture that risk perception is culturally determined. *The International Journal of Project and Business Risk Management*, 1(2), 185–202.

Cooper, M. D. (1997b). *Improving safety culture: A practical guide*. J. Wiley.

Cooper, M. D. (2000). Towards a model of safety culture. *Safety Science*, 36(2), 111–136. https://www.behavioral-safety.com/articles/Towards_a_model_of_safety_culture.pdf

Cooper, M. D. (2002). Understanding and quantifying safety culture: A reciprocal model for success. *Professional Safety*, 47(6), 30–36.

Cooper, M. D. (2016). *Navigating the safety culture construct: A review of the evidence*. B-Safe Management Solutions Inc. https://www.behavioral-safety.com/articles/safety_culture_review.pdf

Cooper, M. D. (2018). The safety culture construct: Theory and practice. In C. Gilbert, B. Journé, H. Laroche, & C. Bieder (Eds.), *Safety cultures, safety models: Taking stock and moving forward* (pp. 47–61). Springer. https://doi.org/10.1007/978-3-319-95129-4_5

Craigen, D., Diakun-Thibault, N., & Purse, R. (2014). Defining cybersecurity. *Technology Innovation Management Review*, 4(10), 13–21. https://www.timreview.ca/sites/default/files/article_PDF/Craigen_et_al_TIMReview_October2014.pdf

Da Veiga, A. (2016). A cybersecurity culture research philosophy and approach to develop a valid and reliable measuring instrument. In *2016 SAI Computing Conference (SAI)* (pp. 1006–1015). IEEE. <https://doi.org/10.1109/SAI.2016.7556102>

Fernández-Muñiz, B., Montes-Peón, J. M., & Vázquez-Ordás, C. J. (2007). Safety culture: Analysis of the causal relationships between its key dimensions. *Journal of Safety Research*, 38(6), 627–641. <https://doi.org/10.1016/j.jsr.2007.09.001>

Freaney, C. (2011). *Safety culture and safety behaviors among firefighters* [Doctoral dissertation, University of Tennessee]. TRACE. https://trace.tennessee.edu/utk_graddiss/969

Gcaza, N., & von Solms, R. (2017). Cybersecurity culture: An ill-defined problem. In M. Bishop, L. Fletcher, N. Miloslavskaya, & M. Theocharidou (Eds.), *Information security education for a global digital society: WISE 2017* (pp. 98–109). Springer. https://doi.org/10.1007/978-3-319-58553-6_9

Gcaza, N., von Solms, R., Grobler, M. M., & Van Vuuren, J. J. (2017). A general morphological analysis: Delineating a cybersecurity culture. *Information and Computer Security*, 25(3), 259–278. <https://doi.org/10.1108/ICS-12-2015-00>

Georgiadou, A., Mouzakitis, S., & Askounis, D. (2021). Assessing MITRE ATT&CK risk using a cybersecurity culture framework. *Sensors*, 21(9), 1–14. <https://doi.org/10.3390/s21093267>

Georgiadou, A., Mouzakitis, S., & Askounis, D. (2021). Designing a cybersecurity culture assessment survey targeting critical infrastructures during the COVID-19 crisis. *International Journal of Network Security and Its Applications (IJNSA)*, 13(1), 33–50. <https://doi.org/10.5121/ijnsa.2021.13103>

Georgiadou, A., Mouzakitis, S., Bounas, K., & Askounis, D. (2022). A cyberse-
curity culture framework for assessing organization readiness. *Journal of
Computer Information Systems*, 62(3), 452–462. <https://doi.org/10.1080/08874417.2020.1845583>

Georgiadou, A., Michalitsi-Psarrou, A., & Askounis, D. (2022). Evaluating the
cybersecurity culture of the EPES sector: Applying a cybersecurity culture
framework to assess the EPES sector's resilience and readiness. In *Proceed-
ings of the 17th International Conference on Availability, Reliability and
Security (ARES'22)* (pp. 1–10). Association for Computing Machinery. <https://doi.org/10.1145/3538969.3543813>

Georgiadou, A., Mouzakitis, S., & Askounis, D. (2022). Working from home
during the COVID-19 crisis: A cybersecurity culture assessment survey.
Security Journal, 35(2), 486–505. <https://doi.org/10.1057/s41284-021-00286-2>

Gioulekas, F., Stamatiadis, E., Tzikas, A., Gounaris, K., Georgiadou, A., Michal-
itsi-Psarrou, A., Doukas, G., Kontoulis, M., Nikoloudakis, Y., Marin, S., Cabec-
inha, R., & Ntanos, C. (2022). A cybersecurity culture survey targeting
healthcare critical infrastructures. *Healthcare*, 10(2), 1–19. <https://doi.org/10.3390/healthcare10020327>

Hanna, M. M. (2020). *Exploring cybersecurity awareness and training
strategies to protect information systems and data* [Doctoral dissertation,
Walden University]. ProQuest Dissertations and Theses Global.

Huang, K., & Pearlson, K. (2019). For what technology can't fix: Building a
model of organizational cybersecurity culture. In T. X. Bui (Ed.), *Proceedings
of the 52nd Hawaii International Conference on System Sciences* (pp.
6398–6407). University of Hawaii at Manoa. <http://hdl.handle.net/10125/60074>

International Telecommunications Union (ITU). (2008). *Series X: Data
networks, open system communications and security: Telecommunication
security. Overview of cybersecurity*. ITU.

Ismail, F., Hashim, A. E., Ismail, R., & Majid, M. Z. A. (2009). The operationalisation of safety culture for Malaysian construction organizations. *International Journal of Business and Management*, 4(9), 226–237. <https://pdfs.semanticscholar.org/ae77/4ae366497a05d1c6efe187f1686263f7cb63.pdf>

Kabanda, G., & Chingoriwo, T. (2021). A cybersecurity culture framework for grassroots levels in Zimbabwe. *Oriental Journal of Computer Science and Technology*, 14(1–3), 17–34. <https://doi.org/10.13005/ojcs14.010203.03>

Karyda, M. (2017). Fostering information security culture in organizations: A research agenda. In *The 11th Mediterranean Conference on Information Systems (MCIS) Proceedings* (pp. 1–11). Association for Information Systems. <https://aisel.aisnet.org/mcis2017/28>

Kaur, J., & Ramkumar, K. R. (2022). The recent trends in cybersecurity: A review. *Journal of King Saud University-Computer and Information Sciences*, 34(8), 5766–5781. <https://doi.org/10.1016/j.jksuci.2021.01.018>

Kwak, Y., Lee, S., Damiano, A., & Vishwanath, A. (2020). Why do users not report spear phishing emails? *Telematics and Informatics*, 48(3), 101343. <https://doi.org/10.1016/j.tele.2020.101343>

Lefranc, G., Guarnieri, F., Rallo, J. M., Garbolino, E., & Textoris, R. (2012). Does the management of regulatory compliance and occupational risk have an impact on safety culture? In *11th International Probabilistic Safety Assessment and Management Conference and Annual European Safety and Reliability Conference (PSAM11 and ESREL 2012)* (pp. 6514–6523). IAPSAM and ESRA. <https://hal-mines-paristech.archives-ouvertes.fr/hal-00734322>

Maalem Lahcen, R. A., Caulkins, B., Mohapatra, R., & Kumar, M. (2020). Review and insight on the behavioral aspects of cybersecurity. *Cybersecurity*, 3(1), 1–18. <https://doi.org/10.1186/s42400-020-00050-w>

Mattila, M. (2013). Different views on defining safety, security, and social responsibility. *Interdisciplinary Studies Journal*, 3(1), 7–20.

Mwim, E. N., & Mtsweni, J. (2022). Systematic review of factors that influence the cybersecurity culture. In N. Clarke & S. Furnell (Eds.), *International symposium on human aspects of information security and assurance* (pp. 147-172). Springer. https://doi.org/10.1007/978-3-031-12172-2_12

Ogden, S. E. (2021). *Cybersecurity: Creating a cybersecurity culture* [Doctoral dissertation, California State University]. Electronic Theses, Projects, and Dissertations. <https://scholarworks.lib.csusb.edu/etd/1284>

Pessemier, W. L., & England, R. E. (2012). Safety culture in the US fire service: An empirical definition. *International Journal of Emergency Services*, 1(1), 10–28. <https://doi.org/10.1108/20470891211239290>

Reegård, K., Blackett, C., & Katta, V. (2019). The concept of cybersecurity culture. In M. Beer & E. Zio (Eds.), *29th European safety and reliability conference* (pp. 4036-4043). Research Publishing. https://doi.org/10.3850/978-981-11-2724-3_0761-cd

Reid, R., & Van Niekerk, J. (2014). From information security to cybersecurity cultures. In H. S. Venter, M. Laack, M. Coetzee, & M. M. Eloff (Eds.), *2014 Information security for South Africa* (pp. 1-7). IEEE. <https://doi.org/10.1109/ISSA.2014.6950480>

Rohan, R., Funilkul, S., Pal, D., & Chutimaskul, W. (2021). Understanding human factors in cybersecurity: A systematic literature review. In S. Paul & J. K. Verma (Eds.), *2021 International Conference on Computational Performance Evaluation* (pp. 133-140). IEEE. <https://doi.org/10.1109/ComPE53109.2021.9752358>

Sas, M., Hardyns, W., Van Nunen, K., Reniers, G., & Ponnet, K. (2021). Measuring the security culture in organizations: A systematic overview of existing tools. *Security Journal*, 34(2), 340–357. <https://doi.org/10.1057/s41284-020-00228-4>

Schatz, D., Bashroush, R., & Wall, J. (2017). Towards a more representative definition of cybersecurity. *Journal of Digital Forensics, Security and Law*, 12(2), 53–74. <https://doi.org/10.15394/jdfsl.2017.1476>

Stacey, P., Taylor, R., Olowosule, O., & Spanaki, K. (2021). Emotional reactions and coping responses of employees to a cyber-attack: A case study. *International Journal of Information Management*, 58(3), 102298. <https://doi.org/10.1016/j.ijinfomgt.2020.102298>

Stanarević, S. (2012). *Koncept bezbednosne kulture i pretpostavke njegovog razvoja* [Doctoral dissertation, University of Belgrade]. <https://fb.bg.ac.rs/download/RepozitorijumDisertacija/2012-11-09%20Stanarevic%20Svetlana/Disertacija.pdf>

Stoneburner, G. (2001). *Computer security: Underlying technical models for information technology security*. National Institute of Standards and Technology (NIST). <https://www.nist.gov/publications/underlying-technical-models-information-technology-security>

The Crown Prosecution Service (CPS). (2019). *Cybercrime – prosecution guidance*. <https://www.cps.gov.uk/legal-guidance/cybercrime-prosecution-guidance>

Tziarras, Z. (2014). The security culture of a global and multileveled cybersecurity. In E. Carayannis, D. Campbell, & M. Efthymiopoulos (Eds.), *Cyber-development, cyber-democracy and cyber-defense* (pp. 319–335). Springer. https://doi.org/10.1007/978-1-4939-1028-1_13

Uchendu, B., Nurse, J. R., Bada, M., & Furnell, S. (2021). Developing a cybersecurity culture: Current practices and future needs. *Computers & Security*, 109, 102387. <https://doi.org/10.1016/j.cose.2021.102387>

Von Solms, R., & Van Niekerk, J. (2013). From information security to cybersecurity. *Computers & Security*, 38(7), 97–102. <https://doi.org/10.1016/j.cose.2013.04.004>

White, J. K. (2017). *Impact of protection motivation theory and general deterrence theory on the behavioral intention to implement and misuse active cyber defense* (Publication No. 10622990) [Doctoral dissertation, Capella University]. ProQuest Dissertations and Theses Global

Unsettling Future of AI-Powered Police-Human Interaction: A Shadow of a Doubt over Moral Sensitivity

Srdan T. Korać

Policing in the Early Twenty-First Century: Can New Technologies Fix the Same Old Problems?

Advances in technology have always influenced policing through the reliance of the police on various innovative technical devices and scientific methods. It has been only recently that a revolutionary shift might be emerging, “a change from police forces which *use* technology, towards forces which are *used by* technology” (McGuire, 2021, 21). As police departments across the globe are looking at deployment of AI as novel tool and a force multiplier for the twenty-first century policing (UNICRI and INTERPOL, 2019), the hot debate on this prominent

Srdan T. Korać, Senior Research Fellow, Institute for Political Studies, Belgrade ORCID: 0000-0003-0722-6419, email: srdjan.korac@ips.ac.rs;

DOI: <https://doi.org/10.5281/zenodo.14204751>

issue has been ranging between two opposite stances: one that highlights clear benefits for police work; the other that underlines the potentials for irresponsible utilisation of this emerging technology. The most optimal utilisation of AI for police purposes has come in the form of mass surveillance of public areas, facial and object recognition, gunshot detection, data/intelligence gathering, analytics, crime prediction, decision making while on patrol (see Kitsos, 2020).

The technologization of the police as a public agency, so-called “smartification”, has been legitimised in terms of improving the safety and well-being of both police staff and citizens. In his account of the today’s perceptions of the police-technology relationship, McGuire denoted the utilitarian view as dominant, which is advocated by the police/government and expressed in the policy outlook that more efficient policing requires more technology, hence more investment in new technologies (2021, 22). On the part of the corporate sector involved in the development of AI for public purposes, the advantages of AI assistant software are being advertised for the police as “a smart way to digitally manage their resources” through automation of administrative procedures, while citizens are now benefiting from a far easier search for desired information, reporting crimes, and sharing concerns about security in their community (ICS.AI, 2023). In the United States, testing of AI-assisted community policing in several towns has already been praised for the improved efficiency of allocated car patrols in responding to gun violence in the urban areas forecasted by AI as the riskiest for serious crime to occur (Davenport, 2021). On the part of the public, Hobson and collaborators (2023, 180–183) demonstrated that citizens, although sceptical about the positive effects of algorithmic policing and favouring human decision-maker, are still more likely to support the utilisation of AI decision making in police

missions when presented with successful examples or experience benefits from them.

Policing as a key governmental mode of control of socially harmful behaviour in modern democracy, and the police as a public agency with a considerable level of discretion in executing legitimate force, are expected to rest on a transparent decision-making process always open to scrutiny (Steverson, 2008, 34–42; Sklansky, 2007, 106–113). Transparency also refers to a clear and comprehensible rationale for decision-making, mirrored in the principle of democratic accountability (Sklansky, 2007, 90–91). Yet the opaque manner in which AI has so far been and is to be deployed in future policing can have unintended consequences due to the black box nature of algorithmic decision-making. That is why the United Nations and Interpol emphasised that the safe and ethical use of AI and robotics technology in law enforcement requires continuous observation and careful examination through various pilot projects (UNICRI and INTERPOL, 2019, 23–24). There has been a rising concern that an AI-driven police force could compromise the delicate and ever-challenging balance between privacy and security in modern democracy, which was already shaken in the last two decades by various controversial legislative and policy measures adopted and enforced in the state response to global terrorism. Summing up reasonable concerns and fears of the use of AI in policing, Elizabeth E. Joh (2018, 1144; 2022, 271–273) pointed out that the essential quandary is “how best to balance its potential benefits while raising and assessing difficult questions of fairness, accountability, transparency, and ethics as the technology is applied to a core government responsibility”.

Being constituted for the peace and security of the bourgeois order, the liberal state has viewed crime through the logic of permanent war conducted by the police against lawless, unruly, disorderly, or antisocial commoners (Neocleous, 2014, 128–133). Supported by the liberal vision of political order, that was mirrored in the imperialist projection of liberal peace onto “Other” (Korać and Stekić, 2018, 99–101), the *mission civilisatrice* of the police was and still is to prevent some people from falling back to savagery through disciplining and the fabrication of social order (Neocleous, 2014, 134). Walter Benjamin (1979, 142) observed a long time ago the all-pervasive, omnipresent, and diffusive character of policing as a normative force: the police power is projecting itself freely, without restrictions; it goes all the way across social, legal, and moral boundaries. Not surprisingly, the chronic unethical performance of the police has been a hot topic both in the media and among scholars and practitioners worldwide (see Kutnjak Ivković, 2014). Dozens of U.S.-based civil action groups have emerged with the aim of exposing the cases of police brutality across the nation via videos and eyewitness accounts.¹ As the most successful example of a grass-roots response to systemic police violence, the movement Black Lives Matter was formed in 2013 to fight unjust police killings of Black Americans by organising street protests and public campaigns (see Hillstrom, 2018). The civil reaction against the systemic police racially biased misconduct culminated in 2014 in relation to the deaths of Eric Garner and Michael Brown and again in 2020 following the deaths of Breonna Taylor and George Floyd, with

¹ Cop Watch, Cop Block, Recording Police, Filming Cops, International Day Against Police Brutality, People United Against Police Brutality, Americans Against Corrupt Cops (AACC), just to mention the most prominent.

massive discontent erupting in a months-long civil unrest across the nation (King, 2020). Black Lives Matter has illuminated and continues to shed light on a series of cases revealing a disturbing pattern of white police staff's use of excessive force, particularly in interaction with African Americans.

Public trust in the police, a government agency with considerable executive powers and discretion that may and does cause legally and morally inappropriate harm to the public, has been on a downward trend over the last decade, at least in the United States (Washburn, 2023) and the United Kingdom (Mynenko and Ditcham, 2022). The research conducted by Hamm and his associates (2021) has shown that the ongoing crisis of public trust in policing—whether caused by agency-level or individual morally wrong decisions—is directly connected to citizens' perceptions of the extent of their vulnerability to police-caused harm. These authors (2021, 1767) suggested that the legitimacy of police-community relations can be enhanced by reducing or eliminating pre-emptive doctrine in police work, e.g., pretextual and mixed-motive traffic stops, that include more intrusive manners of contact with random citizens, i.e., citizens who are not targeted because of their assumed or actual illegal behaviour. The chronic problems with trust in the police are now possibly becoming more serious due to potential controversies related to the forthcoming utilisation of AI systems in policing. Joh (2022, 270) warned that “AI systems are already a part of ordinary police work” and that several applications have emerged as “controversial because they raise the specter of vastly increased police power with new risks and few checks”. The United Nations and Interpol recommended police agencies worldwide to carefully examine the fast-approaching challenges to AI policing with a view to criminal exploitation of AI, technical problems in

using AI to support police operations and deployment of robots in patrolling and surveillance, and a general lack of expertise in AI among police staff (UNICRI and INTERPOL, 2019, 24–25).

Touching upon the problematic trustworthiness of some AI tools (e.g., facial recognition and DNA phenotyping), Paulsen (2021, 7–8) identified an unethical professional practice of so-called “Digital Dirty Harryism”, which he defined as “a conscious utilisation of a dirty means to achieve some good”. Despite their sophistication and usefulness, Paulsen is sceptical about whether AI-powered police applications are ready for the real world of law enforcement and crime prevention, which is usually overwhelmed with numerous legal and moral issues. The limitations of those tools are accepted by police staff as “reasonable most of the time”, which brings inherent relativism to their deployment without proper testing or understanding of the system (Paulsen, 2021, 8). A study conducted by the National Institute of Standards and Technology (NIST) in 2019 has determined that most facial recognition software in use displays demographic biases (race, age, and sex) in such a manner that “Asian, African American and Native groups are 10 to 100 times more likely to be wrongly identified compared to Caucasians”, as well as women compared to men and older adults compared to middle-aged adults (NIST, 2019; Grother et al., 2019). By the end of 2022, in the United States alone, three cases had been registered of Black men falsely arrested for crimes they did not commit because they were mistakenly identified due to a mismatch caused by flawed facial recognition technology (Hill, 2021). In all three cases, the victimised citizens were not only arrested but jailed and paid costs for defending themselves in court. In addition to insufficient understanding of how AI algorithms make decisions and produce their insights, recent experiences of the AI transformation

of the wider public sector suggest that success factors for effective adoption as well as ethical and safe implementation of AI policing solutions—that is, building, training, and deploying AI systems—are connected to conscious and unconscious biases that already exist in the police force (Dineva, 2022). Hartmann (2018, 161-162) highlighted the phenomenon of “grey zone creativity” in the police that emerges when police staff favour the use of brand-new technologies over standard equipment and operating procedures if they are easy to utilise to achieve police missions optimally.

The practical application of algorithmic forms of policing can be represented by semi-autonomous and autonomous devices, such as drones, patrol vehicles, and robots, aimed at enhancing the workability of the frontline officers (first responders), who play a crucial role in protecting communities. A drone, patrol car, or robot is used as a platform for an AI system, which is the core equipment, a sort of “brain”, with additional accessories such as HD 4K cameras, night sensors, thermal and panoramic imaging, data transmitters, long-range loudspeakers, and communication apps (Thomas, 2022). For instance, equipped with a facial/object recognition system, police drones allow real-time tracking of individuals and vehicles through non-intrusive observation (CPT, 2016, 34). Drones are now being developed towards greater operational autonomy so that they can respond quickly to an emergency call and provide the command/dispatch centre with preliminary information on an illegal action in progress before frontline officers reach the crime scene. Surveillance drones can operate without being detected if engaged in the pursuit of suspects and can inspect open or enclosed, potentially hostile areas to prevent unnecessary injuries or the loss of lives of police officers. Although most police drones are so far deployed for surveillance and intelligence gathering

missions, there has been an emerging trend towards their gradual weaponization. For instance, North Dakota legalised the “less than lethal” armed police drones in 2015: drones can use only tear gas, pepper spray, beanbags, rubber bullets, and Tasers (Wagner, 2015). In the near to mid future prospect, autonomous patrol cars will provide the police with the capability of remote traffic control, identifying road traffic violations and stopping law-breaking cars, responding to traffic accidents, observing and intervening in traffic jams, acquiring footage of fatal road traffic for the coroners and courts, as well as participating in fugitive searches and rescue operations (Gregg, 2019, 27–30, 49–60).

Yet the most promising AI-driven device in terms of expected effectiveness and efficiency is a police robot. Policing in the United States has been leaning on the utilisation of robots since the early 2000s to assist police officers in surveillance, intelligence gathering, bomb disposal, and hostage situations (Sharkey, 2016). Police robots available now on the market have the capability of manoeuvring in tight spaces and under water, climbing stairs, crawling over different types of terrain, lifting objects, clearing obstacles, and diffusing explosives (Thomas, 2022). How fierce is the debate about robotic policing illustrates the recent example of the failed introduction of police robots in San Francisco. City authorities gave city police the powers to use potentially lethal, remote-controlled robots in emergency situations, only to withdraw their decision several weeks later (Har, 2022). According to the initial decision, the robots, whose lethal use was intended to be authorised only by a narrow group of police managers, should have been equipped with explosive charges “to contact, incapacitate, or disorient violent, armed, or dangerous suspect”, if the alternative force or de-escalation tactics failed to end the mission (Har, 2022). A lethal police robot was first used in July 2016, when it

was dispatched by Dallas police during a protest against police violence to eliminate a sniper holed-up in a building who had killed five officers and wounded seven (Sinder and Simon, 2016). The Remotec Androx Mark V A-1 robot, with a half kilo of explosive mounted on it, was sent only after a fruitless 45-minute gun battle and a failed two-hour negotiation with the sniper (Sinder and Simon, 2016). The robot was not autonomous but remotely operated, and its lethality was improvised as it was repurposed against its original design. The police officers who used the Remotec robot were later cleared of criminal charges for the murder of the sniper (Kenning, 2018). Early warning of the possible serious negative consequences that an increased use of armed drones in policing may have for human and civil rights alerted the public several years before the Dallas incident. In 2011, a mobile home in Tennessee was mistakenly burned to ashes during the course of a day-long police action when a deployed robot caused a fire as a side-effect of the excessive launching of tear gas grenades aiming at the trailer's interior (Hennigan and Bennett, 2016). At the end of the mission, the police realised that the fugitive they had been hunting for second-degree murder had not been holed up in the trailer.

Further advances in the design of AI-driven devices utilised in policing, just like many times before when new technology was introduced into police work, open the question of reaching equilibrium between the cost-benefit approach in policy implementation and societal values. In policing, the neoliberal utilitarian logic of "value for money", that is, making the most of the public budget dedicated to the police, is combined with the sacred goal of saving police staff's and innocent citizens' lives. As much as this logic seems to be correct in abstract terms of democratic values and standards of good governance, it becomes prone to moral corrosion as soon as it is materialised in

decisions regarding the course of concrete action. For instance, the police may prefer deploying armed robots excessively on the grounds of avoiding or minimising risks for the police force and protecting citizens in the most efficient way possible, that is, by allegedly saving the public money and private property. McGuire (2021) gave a somewhat apocalyptically toned prophecy that automation means the end of policing as we know it in modern democracy. He argued that the combination of three major factors might be critical for the anticipated revolutionary transformation of the police and policing (2021, 30–32). First, the indifference of the ruling elite towards the probable ethical quandaries of the utilisation of armed robots, which is rooted in utilitarian “enchantment” with quantitatively based evaluation of the performance (cost-effectiveness and efficiency). Politicians are prone to support and legitimise the massive use of police robots if it can be presented to the public as a more elegant cost-benefit policy option. Second, McGuire reminds us that we are oftentimes more inclined to accept intrusions into the privacy of advanced technologies if we perceive that technological novelties provide obvious benefits, even smaller ones (2021, 32). Third, McGuire underlines that in everyday situations we are likely to favour communication with artificial agents over “a uniformed human officers with poor PR skills” (2021, 32).

Plausible negative social and ethical consequences of existing or emerging professional practices related to AI-based policing may not lead to a Robocalypse. Nevertheless, they may have serious and long-term effects on the humane foundation of the police–citizen relationship, bearing in mind that marginalised communities are already being overpoliced. As AI systems can provide more intensive and intrusive policing of marginalised communities, the deployment of 24-hour-operating smart devices could lead to a more oppressive

and, hence, more inhumane society. The well-established practice of remote killing by military drones in U.S./NATO-led interventions across the globe has revealed depersonalisation as a new side-effect: the enemy soldier is now a mere object: distant, as it is placed beyond the horizon, invisible to the drone operator, and, thus, devoid of moral value (Korać, 2018). AI-driven policing's technological purification of police procedures and practices seems to bear analogy to the utilisation of military drones, as there is a risk of transforming policing into an unempathetic set of automated processes that relieve human officers from constant coping with the psychological burden of moral concerns. In the circumstances of chronic public distrust in the police, caused by widespread unethical performance by police staff, the issue of ensuring trustworthiness in AI systems that are being or will be deployed in policing seems to be central. Trustworthiness in an AI system refers to the basic assumption that it will operate in the manner programmed by its designers and software engineers. In other words, we expect the device to function in predictable ways in any given situation. Yet social situations, such as most police missions, are complex as they imply uncertainty inherent to interaction dynamics between human beings. Regarding the police AI tools—drones, robots, and autonomous cars that are or will be utilised to interact with humans—the quality of trustworthiness is becoming more important if police management and government are to restore the once-eroded humane aspect of policing.

Perfect trustworthiness in police AI tools can hardly be set and pursued as a realistic goal. As community-oriented policing has been perceived by the public as a mode of police work critical to social well-being, the design of AI tools has to embed all necessary requirements for the truthful replication of traditional, face-to-face

interaction. My analysis will be focused primarily on the issues arising from the use of police drones and robots, as they are expected to have the capacity to interact with humans in the near or mid-future. Engineering properly operating computer software for sophisticated moral reasoning has so far shown many difficulties when it comes to translating the contextualisability of leading ethical theories into algorithms (Klincewicz 2017). Similarly, the problem of how to act appropriately in situations of moral uncertainty is complicated to translate into the context of computer programming (Bhargava and Wan Kim, 2017). We humans choose our course of action oftentimes in the face of uncertainty. Does one situation require an ethically informed decision? If it does, and all other relevant non-moral information is known, it may still remain to reflect on what exactly ethical norms require of us in the situation. The use of drones and robots for remote interaction between police officers and targeted individuals has to acknowledge the significance of social emotions as a component of this interaction, particularly in police missions with a high level of risk and intensity and, thus, greater uncertainty about the outcome.

Why are social emotions essential for moral sensitivity as a prerequisite for sound ethical decision-making? Curry (2016, 29) summarised the long history of the conceptualisation of morality in the notion of morality as “a collection of biological and cultural solutions to the problems of cooperation and conflict recurrent in human social life”. During the millennia-long process of evolution by natural selection, the brain has crafted psychological mechanisms upon which prosocial behaviour is based (Joyce, 2006). In parallel, humans have undergone group selection that has made social exchange and reciprocity much more preferred modes of behaviour (Tomasello, 2016, 137–143). Empathy and love, gradually developed social emotions (high-order

emotions), are at the root of prosocial behaviour and of genuine (long-term) altruistic motivation (Hoffman, 2007). Social emotions constitute a human being's capacity to act morally, which is needed for sustaining a livable society. They also help us recognise moral issues when they arise in a social situation, providing us with the ability of moral sensitivity. Backed by self-consciousness, social emotions in part uphold our acceptance of responsibility for how we behave towards others. In contemporary psychology, empathy is usually referred to as "*knowing* another person's feelings", "*feeling* what that person feels", and "*responding compassionately* to another's distress" (Goleman, 2006, 58). Not denying the importance of empathic accuracy (proper cognitive assessment of another person's feelings), Hoffman (2007, 30) defined empathy rather in terms of affective empathy, that is, empathic response as "prosocial moral action [that] usually involves helping someone in discomfort, pain, danger, or some other type of distress". Empathy is the primary driver of prosocial behaviour and a precondition for moral sensitivity. It is, thus, important to investigate how the utilisation of AI does or may shape the police-citizen relationship in the context of identifying with the other in her/his situation, followed by an affective empathic response to her/his needs,

The main research question is how the design of AI systems incorporated in police drones and robots affects and may affect the moral sensitivity of decision-makers and, consequently, the whole process of ethical decision-making. The starting assumption is that a sound ethical decision-making process, an essential component of the police daily routine, requires adequate awareness of the moral implications of intense social situations in which police officers interact with citizens, either in person or remotely via drones and robots. The following investigation is to identify the critical points in the initial

phases of the ethical decision-making process regarding only the part of police work related to direct interaction with the populace. I will delve into plausible negative effects for citizens subjected to police procedures that may stem from the use of AI-powered devices by examining controversies that are likely to emerge in a police officer's or manager's deliberation about the course of action. I will take into consideration the current, near-, and mid-future technological advances discussed previously in this section, which means that the analysis will be narrowed to the utilisation of remotely controlled and semi-autonomous AI-powered devices.

Designing AI-Driven Police Devices to Recognise Social Emotions: An Algorithmic Riddle?

The way AI is designed, or we wish it to be designed in the future, may settle existing social problems or generate new ones. Popular culture, especially in the form of science fiction literature and films, has been igniting as much our imagination as our fear with a view to the direction in which civilisation is evolving—either utopian or dystopian. One of the most favourite topics is the troubled relationship between humans and intelligent machines acquiring self-consciousness. The brutish habitus of the leading character in the legendary film franchise *RoboCop* (1987–2014), though not a robot but a heavily armoured and armed cyborg, comes to mind as the best-known portrayal of the dystopian future of police work. In near-future Detroit, descended

into chaos as its streets are flooded with violent crime, RoboCop is a project of greedy public–corporate partnership designed to efficiently fight criminal gangs, resulting in overwhelming public support for the protagonist’s deadly performance (Verhoeven, 1987). In the second sequel, *RoboCop 2* (Kershner, 1990), the plot is partly centred around problems in programming the cyborg so that he can perform efficiently police tasks but in a more human-friendly manner; yet it turns out it is impossible to achieve this equilibrium. Although plotted rather in a naive way in terms of technical explanations and ethical dilemmas, these action films at least indicate the issue of how humans will cope with the emotionless “nature” of AI-driven devices in daily interaction. The problem of disbalance between unempathetic decision making and “supernatural” powers will actualise abruptly if Artificial General Intelligence (AGI) is achieved. While Artificial Narrow Intelligence (ANI)² supplies solutions for a limited scope of problems, the ultimate goal is to build AGI, a self-teaching system that would imitate human intelligence and allow for adaptation based on environmental information in the absence of explicit rules. Only AGI could enable an effective and fast adaptation of a device to never-ending changes in social surroundings on the basis of its programme.

The reality of present-day policing is populated by boring robots: benevolent and inelegant, or even clumsy, as they come shaped as mechanic dogs or R2-D2-esque traffic stop robots (Holley, 2019). In 2017, a pace towards police androids was set in Dubai, where the authorities introduced their own Robocop but in the benign role of a “mobile information kiosk” equipped with a “touch-screen communication

2 Note: Artificial Narrow Intelligence is the type of artificial intelligence to which I refer throughout my paper when I use the acronym “AI”.

panel on its chest” (Hambling, 2018, 221). The Dubai Robocop (the REEM robot) is capable of carrying out a conversation with citizens asking for information, paying traffic fines, or reporting a crime (Hambling, 2018, 222). Rich and Sindner (2009) claimed that we prefer to interact with “agents which are overtly ‘human-like’”, which, in addition, means that we would more easily accept intelligent machines capable of perceiving, reasoning, and executing. In the circumstances of the rapid development of AI and related cutting-edge technologies, the Dubai Robocop seems to remain just an advertising trick to prepare the public for the advent of fully autonomous robotic policing (Hambling, 2018, 222). A fully autonomous AI-driven device would be able to perceive the environment (through diverse sensory capacities), identify and extract data from that environment (e.g., data relevant for its goals), apply situational thinking to collected and processed data to make a decision on its own how to react (with the help of machine learning), and then perform the chosen course of action independently of human control or intervention (Beetz et al., 2016, 329; Negnevitsky, 2005, 1-17).

Exploring AI systems for remote interaction between police officers and targeted individuals in their operational side related to moral sensitivity, I am now shifting my attention to how an AI-driven device perceives the environment in order to identify and extract data relevant for recognising moral issues in a police mission. The quality of data used in decision-making could be compromised by low-quality end products resulting from omissions and errors in their collection, processing, and analysis, or it may be undermined due to embedded conceptual biases. At the conceptual level, AI designers and software engineers have already acknowledged that human beings are involved in two parallel, interdependent decision-making processes: “one, slow,

deliberate and mostly rational, the other fast, impulsive, and able to match the present situation to prior experience...” (Cristianini, 2017, 51). The second mental process, which is biased by emotions, enables quick conclusions, making human intelligence so effective compared to any artificial intelligence (at least, so far). Trustworthy interaction between human beings and police AI-driven devices requires solving the problem of the recognition of emotions, particularly social emotions, as they play a fundamental role in developing moral awareness.

Affective computing is an interdisciplinary field spanning computer science, psychology, and cognitive science dedicated to the design and development of systems and devices that can recognise, interpret, process, and simulate human emotions (Tao and Tieniu, 2005). Picard (1997) laid the foundation of affective computing and developed an approach to emotional components in AI-human interaction that is more responsive to the user’s context. Picard (1997, 165–192) envisioned and conceptualised the capabilities of computers to recognise human emotions and express their own emotions. In her later paper, Picard (2010) proposed a model for computer measurement of observable functions of human’s emotional states, acknowledging emotion recognition as a dynamic pattern recognition problem. In response to identified problems in existing affective computing methods and results, Jeon (2017, 17) suggested that further improvements should include: designing better detection of a user’s emotional states and adaptive responding to those states; the fusion of data from multiple sensors; developing AI system that can empathise with humans; and designing AI system with emotions and capability for them when making decisions. Present-day AI systems are capable of recognising human emotions and predicting mental states based on facial recognition, voice and text analysis. For example, Imotions,

an IT company specialised in human behaviour research software founded in 2005, produces software that enables combining the data from multiple biosensors—collected from eye tracking, facial expressions, electrodermal activity, electroencephalography, electrocardiography, and electromyography—to “get better understanding of human thoughts, feelings, and behaviors in any environment” (Imotions, 2023). In order to successfully recognise human emotions, an AI system uses audio, video, and other sensors, measures physiological responses, computes the relevant data to estimate the expressed emotion, computes an appropriate emotion in return, then generates it using the hardware, and finally measures the success of this interaction for further learning and improvement. The most influential face emotion recognition model, in academia and in practice, is Ekman’s Facial Action Coding System (FACS), conceptualised to taxonomize human facial movements based on their facial appearance (PEG, 2023). AI systems, such as computers and robots, have been using training of deep neural networks to learn to detect facial emotions in real time, as they deliver information about an actual human’s inner state (Pentland, 2007). The researchers of the Media Lab at the Massachusetts Institute of Technology successfully created machine learning models supporting robots to, based on the analysis of facial expressions, correctly recognise and interpret the current emotional state of the person they are communicating with (Matheson, 2018). With the help of a new methodology, the robot can see, measure, and understand basic facial expressions and translate them into human emotions. This is the missing link in the relationship between robots and humans that could not be solved technically for a long time because human emotions are a very complex and inconsistent aspect of personality, sometimes difficult to recognise even in direct, face-to-face communication between human beings alone.

There has always been a disparity between the design of technical devices and their social application, that is, between the vision of inventors and the needs of end users. For instance, when it comes to AI, there are hazards inherent to algorithmic decision making associated with core meaning and open texture risks. Weng and collaborators (2009) stressed that the vagueness in interpretations of the core meaning of any term—allowed by the open texture characteristic of human language based on abstract thinking—may lead to misinterpretation of the meaning of sentences, i.e., the inability of an AI-driven device to grasp the abstract meanings and vague terms used in communication with humans (276–278). Having argued that the sharing of commonsense knowledge has to constitute the communicative framework of AI-human interaction, Ein-Dor (1999) cautioned that this sort of knowledge is almost impossible to translate into computer programmes primarily because it is culturally and contextually specific. That is why successful utilisation of any police AI system has to take into account the sheer fact that humans, with whom such a system is to interact, are woven into a large social network, which in itself bears plenty of different cultural and symbolic meanings.

If AI-driven devices, such as robots, are to be used in a natural environment, that is, outside the constrained domain of the laboratory, they must display the capability of responding to the nuances in verbal and nonverbal communication as well (Meacham and Studley, 2017, 100–101). Proposing the framework for mutual shaping and co-production between society and robotic technologies as an alternative perspective for evaluating social robots, Šabanović (2010, 445) underlines the need for “a more open definition of the context of robot design, in which uncertainty, situational awareness, adaptability, and social responsibility play an important role”. Westerman and collaborators

(2020, 397) emphasised that the AI-human interaction is based on the anticipation of human agents that it will replicate the human-human interaction, although it begins with “a greater sense of uncertainty and lower anticipations of social presence and liking”. In other words, “the primary cue of interaction triggers an expectation of that partner being human, and then the cue of who that partner is (human or machine) leads to comparisons to that expectation” (Westerman et al., 2020, 397). By conducting a non-verbal, non-symbolic robotic experiment in which an observer (AI system) had the task of visualising the future plans of the other agent (AI system) without any prior symbolic information or assumptions, i.e., based only on an image depicting the initial scene of the other agent, Chen and collaborators (2021) found that their AI system was able to achieve it with 98.5% success. This team of researchers holds that such visual behaviour modelling may constitute a cognitive capacity needed for an AI system to understand the diversity of social behaviour and predict the actions of other agents without language or background knowledge. Still, Chen and collaborators (2021) were not sure if it was or would be possible to design the cognitive capacity of an AI system to comprehend and predict behavioural patterns in a multi-agent environment.

Hieda and Nagai (2022, 26) challenge the proposition that AI-driven devices, such as robots, can address social emotions as higher-level emotions because the replication of complex emotional behaviour is unlikely so far. Two researchers hold that the optimal method to incorporate social emotions into AI software would be to manually write down all situations and to define the reaction in each situation (2022, 27). This is hardly possible, according to Hieda and Nagai (2022, 27), because it would require “collecting a large amount of human physiological data and first-person view images to identify the

agent's behavior and their situation at the time that they experience different emotions". The essential obstacle to programming artificial social emotions, Hieda and Nagai (2022, 28) have found, is the fact that social emotions are constituted gradually within the realm of social interactions and cemented with the help of shared beliefs, symbols, and values. As they are internalised in early childhood through primary socialisation and later developed in relationships with other people, social emotions are impossible to replicate by translating them into machine learning. Mintrom and collaborators (2022, 128) pointed out that "the competency of robots will be dependent on both the capabilities they have, the effectiveness of robot operators in programming them and their abilities to cope with the unpredictable nature of public space". Unlike a human brain, which relies on a mental model of reality and past experiences to predict probable outcomes when choosing a course of action in response to a new challenge or social situation, an AI system always has to process enormous sets of training data and relevant data extracted from the environment to derive an optimal way of acting and then perform it (Medlock, 2020).

Possessing feelings in the context of sociality is a two-way path: the capability of feeling has to be reciprocal if it is to be evaluated positively by the standards of ethical behaviour. In discussing the moral status of artificial beings, Labossiere (2017, 296-299) argues that even highly advanced AI-driven devices will never be capable of feeling and suffering, i.e., possessing emotions, a constraint that considerably limits the relationship with human beings in terms of "mutual" trust. The police AI system would be indifferent to the suffering of victims of crime or abused suspects, while humans would be prone to fake feelings to manipulate the robotic police officer in order to get its sympathy and favourable treatment. Leaning on the assumption

that a person can fake having an emotion while engaged in a flawless performance, Labossiere claims that the only way to determine fake feelings would be a test that could reveal whether there is a presence or absence of feelings (2017, 297). As correct recognition of emotional states and credible interpersonal communication are based on empathy, one of the essential programming problems that remains to be solved in social robotics is how to design robots with artificial empathy based on natural (human) empathy. Asada (2015) proposed a conceptual constructive model of Affective Developmental Robotics (ADR) as an approach that should develop more authentic artificial empathy. In his model, Asada advocates the design of artificial empathy that can rely on imitation in the form of emotional contagion, that is, through the process by which the feelings, facial expressions, and behaviour of one person are transferred to another person or group of people. Asada (2015, 31) considers emotional contagion to be the initial step towards cognitive empathic connection, which further develops into sympathy (compassion) and higher forms of emotion linked to subcortical areas of the brain.

Can AI-powered Policing Be Morally Sensitive?

Trust in AI systems, based on human perceptions of the safety and competence of those systems, is likely to erode if greater autonomy does not bring expected levels of performance quality (Mintrom et al., 2022, 127). Although currently powered only by narrow, or so-called weak AI (i.e., ANI), police algorithmic devices have to meet criteria of novelty and predictability at the same time. Novelty is a necessary

precondition for creative responses to new challenges arising daily from the police work, while predictability fits the requirements of safety and security in interaction with human beings. The good departing point of my investigation is a recent remark made by Szocik and Abylkasymova (2022, 543) that, being an old structural ingredient of policing, biases in algorithmic decision-making are not likely to be “more hazardous” than existing ones. The moral dimension of a decision on the course of action in police work, as much as in the professional conduct in any public office, may be interfered with, distorted, or blocked by biased reasoning. Werhane and collaborators (2013, 69) remind us that the capacity for ethical action is “universally vulnerable to factors unrelated to the actor’s consciously and deeply held value systems”. Those external undermining factors are commonly displayed in management or peer pressures stemming from morally insensitive professional mindsets and practices inherent in local/national police organisational culture, or in general prejudices internalised from society at large (e.g., racism, sexism, homophobia, and transphobia). As external, situational factors do not inhibit the automatic triggering of moral frameworks and are mostly of second-order significance to cognitive biases, I will touch on their role only partially. The ethical decision-making process has five steps (see Josephson, 2002; Lewis and Gilman, 2005, 141–160; Martinez, 2009, 112–132; Hartman and DesJardines, 2020, 34–45), but the first two steps are critical for the result of the whole process: 1) becoming aware of a situation as a dilemma that calls for an ethical response; 2) gathering all facts relevant to reasoning about the ultimate course of action. Two stages are somewhat interdependent because a police officer, as a moral agent, has to observe a situation, that is, obtain preliminary information, in order to determine if the situation involves moral concerns to be taken into consideration in later stages of the ethical decision-making

process. Two parallel avenues for the utilisation of police AI-empowered devices will be embraced: one that assumes devices with weak AI and the other that includes devices with strong AI. The first avenue refers to “remote” recognition and interpretation of social emotions in the case when a human decision-maker learns about the emotional state of a suspect (or other agents) through the “epistemological veil” of an AI-driven device: on the basis of data collected and partly processed by the AI assistant programme. The second avenue anticipates a device capable of reading and interpreting the emotional state of a suspect autonomously, which then presents its own evaluation to a human decision-maker in the form of recommendations.

*Gathering Preliminary
Information and The
Recognition of the
Emotional State*

When used in patrol missions under full or partial human control, drones or robots perform the tasks of observing and gathering information/intelligence in the reported area, especially when high-risk criminal activities are still in progress (e.g., gun shooting, large-scale violent behaviour in public spaces). Existing and cutting-edge technologies can provide enormous volumes of various types of instantly accessible data (online via notebook in a patrol car), but the problem for a human decision-maker in the police is how to select credible information gathered in the highly dynamic and almost unpredictable circumstances of an ongoing police action. It is crucial to have all

information relevant at hand to properly determine if moral concerns are involved in a mission.

In the near and mid future, a set of high-tech sensors mounted on drones and robots will provide a real-time and mobile capacity for on-the-spot retinal, voice, and facial recognition as well as the extraction of fingerprints and DNA samples of suspects and matching them with biometric data from police databases (Williams and Williams, 2008, 168). AI-driven devices allow situational awareness through HD 4K video surveillance of the crime scene and surrounding area and collect valuable preliminary evidence, transmitting all data to the command centre for further processing and decision-making. In addition, video imagery of a suspect running from the crime scene, produced with a drone camera and processed by image enhancement software, leads to quick identification (Williams and Williams, 2008, 169). The critical point in this stage of ethical decision-making is how to properly combine a large volume of in-real-time streamed data from multiple biosensors to arrive at the correct reading and interpretation of the emotional states of the individuals subjected to police action. Although an AI-driven drone or robot, equipped with enhanced audio, video, and other sensors, will be capable of measuring physiological responses and processing the relevant data to estimate the expressed facial expressions and translate them into human emotions, it is safe to question the reliability of remote as well as artificial recognition of the emotional state of perpetrators of criminal actions. In the case of remote recognition of emotional states, the adaptive survival response to a sudden potentially threatening event occurs in the form of visual and auditory selective human attention, meaning that the attention becomes narrow and externally focused on the perceived threat. The police officer's brain "switches to the intuitive

mode of processing, in which decision making is faster, emotionally based, automatic, and subconscious” (Best et al., 2019, 566), but with “derangement of cognitive function and rational control” (Brisinda et al., 2019, 309). Along with fatigue due to overwork and sleep deprivation, sensory and information overload heavily reduce cognitive capacities (Werhane et al., 2013, 86–87). An overburdened police officer may handle the abundance of available data provided by an AI-driven device by either misinterpreting it as insignificant for ethical decision-making or simply omitting it as irrelevant to moral concerns.

The utilisation of technology for live facial recognition has so far unveiled some deficiencies in accuracy that may endanger civil liberties and human rights, which become more troublesome in the light of expected enhancements in the capacities of AI-driven devices. Present-day problems with accuracy are mostly related to the creative tactics of the populace in response to the oppressive character of massive CCTV-based systems of video surveillance that have the option of facial recognition. Obstruction embraces a wide scope of methods and tools: wearing plastic face masks (simple or with illusive shapes), so-called “computer vision dazzle” makeup, and small face light projectors; using special camouflage patterns on hats and clothes that only confuses the camera or even makes a whole person invisible (Hayward, 2021, 185; Szocik and Abylkasymova, 2022, 534). The future practice of scanning suspects on the spot with a drone or robot camera to conduct an instant biometric identity check might still result in misidentification and the erroneous arrest of the innocent. The additional problem for the innocent might be caused by the rapid exchange of newly acquired wrong biometric identity data for an allegedly suspect, a feature of the AI-powered police IT system that connects police departments and agencies in one network. Once

rapidly shared throughout the network, information about the wrong identity of a suspect becomes difficult to correct and enforces inaccuracies. Such errors may have a tragic outcome in high-risk situations when drone or robot swarms are deployed to kill suspected terrorists or serial killers. As some authors warned, future use of drone swarms for actions that include lethal force in a crowded public space might cause fatalities due to the operation of facial recognition systems inadequate for scanning large numbers of individuals (Brundage et al., 2018, 21).

Despite fast-paced advancements in the technology of audio, video, and other sensors, there will remain a possibility of misreading or inaccurate reading of the emotional state of targeted individuals caused by the dynamic circumstances in which a drone or robot are operating over the crime scene or designated area. In order to stay safe from the adversarial impact of criminal actors, police drones usually operate one hundred metres above the ground on average, in which case the sensor function may come under the influence of physical obstacles or damage from counteractions (e.g., gun fire, electronic jamming, or hacking). For instance, continuous and quick body motions or covering up the face or head, as well as eye-tracking sensors aimed at an individual currently affected by drugs, are likely to give incomplete or wrong data. Even when the face is clearly visible via camera, like in direct interaction between a suspect and an AI-driven device, facial expressions of emotions can be faked, just like in human-human face-to-face communication. An important difference is the lower degree of successful interpretation of fake emotions through video imagery than in personal interaction with an experienced police officer. As human emotions are a fluctuating psychological feature of personality, reading them correctly even in direct human-human interaction

can require sophistication, let alone AI-human communication. In the case of utilisation of strong AI, a police drone or a robot capable of reading social emotions could still be misled in identifying a suspect's emotional state by, for instance, the gap between physiological data and first-person view images of mentally healthy persons and the facial expressions of a mentally disturbed suspect.

McGuire (2014, 82–85) argued that innovative technological solutions do not necessarily improve professional practices due to a poor understanding of emerging technologies and their incorporation into operational tactics. In his account, police staff prefer using new technology more as a symbolic and PR-motivated act of devotion to progress in general; in reality, innovations are rather interpreted as to confirming existing practices and cognitive biases (2014, 84–85). Selection of information relevant to alert moral sensitivity might be hindered by a police officer's inclination to seek and remember only information that fits her/his expectations, beliefs, attitudes, and values and interpret it in such a way only to confirm her/his professional and/or personal viewpoint—so called confirmation bias. Garner (2008, 11) maintained that our social beliefs shape our perception of reality so that we “readily deny real influences that act upon us in favour of explanations that may better match, not reality, but our social beliefs”. Along with the confirmation bias, the process of selecting massive amounts of information provided by AI systems may also be impeded by two additional cognitive biases: the overconfidence phenomenon, when a police officer's subjective confidence in her/his judgements is reliably greater than the objective accuracy of those judgements; and the base-rate fallacy bias, when a police officer gives more weight to individuating information about an event instead of its general occurrence rate (Garner, 2008, 11–12). Szocik and Abylkasymova (2022,

533–534) remind us that individual cognitive biases are oftentimes supplemented by behavioural patterns shaped by the dominant organisational culture in a police department or agency, which all together can distort machine learning algorithms within specialised police AI programmes by deceiving unaware software engineers to incorporate hidden biases. Thereby, programmed models might be far from reality as “biased data will lead to AI making wrong, biased decisions” (Szocik and Abylkasymova, 2022, 534).

Having distrust in algorithmic decision-making and the robotisation of policing, McGuire (2021, 29) argued that human intervention must remain because the plain experience of police officers cannot be replaced in interpreting gathered data in the context of recognising of the emotional state of suspects and other actors subjected to police action. A good example is a proper understanding of criminal subculture and its slang as a prerequisite of successful fighting and preventing crime within the strategic model of community policing. In a ghetto, as an urban area defined in terms of social isolation and disempowerment, marginalised social groups cultivate their identity in a genuine counterculture, within which slang is as much a means of communication as a conveyor of social meanings. In the case of the deployment of autonomous police robots empowered by strong AI in a ghetto, reading and interpreting the emotional state of suspects through the conversation could be seriously limited due to the probable incapability of the AI programme to understand the complexity and vagueness of social meanings encapsulated in slang vocabulary. Emerging technologies, though presenting a more accurate and detailed factual depiction of crime scenes, cannot replace traditional methods of police work in communities, including personal interaction based on verbal skills and empathic connectedness (McGuire,

2014, 85). Socialisation with the local populace remains the proper path to a better understanding of daily problems in a neighbourhood and the subcultural standpoint of its inhabitants, which is oftentimes very critical of the mainstream culture of liberal order.

Moral sensitivity and emotional distance

The properly built capability of moral sensitivity is central for a police manager or officer to initiate any ethical decision-making process at all, that is, to recognise and ascribe importance to moral dimensions when they arise in the performance of police duty. Yet day-to-day police work requires continuous and quick but effective decision-making, oftentimes in circumstances of extreme pressure and uncertainty. Advances in neuroscience and psychophysiology regarding the emotion-cognition interaction have demonstrated the interdependence and cohesion of two functions in regulating the thought and behaviour of a person facing a high-stress situation (see Brisinda et al., 2019, 308–312). A police officer's capacity to recognise moral issues when they arise, particularly in high-stress situations, is based on synchronicity between empathic accuracy, a correct cognitive assessment of another person's feelings, and empathic response as a prosocial action of helping someone at risk. Once gathered by an AI-empowered system, preliminary information and data on a situation—which always constitutes the police-citizen relationship, no matter if it involves a crime scene or ongoing illegal activity—ought to be included in reasoning on whether that situation has moral concerns or not. The question is: Do available preliminary information and

data encourage the responsible police officer to sympathise with the other in her/his situation “remotely”, mediated through an AI-driven device and its interface, along with recognising the moral issue in the situation and responding to it by choosing and enacting the morally right course of action? Before turning back to discuss this question, let us see what would happen when, or if at all, fully autonomous police robots stepped into the domain of police work.

The second avenue of this two-track account of the utilisation of police AI-empowered devices deals with robots capable of full autonomy in reading and interpreting the emotional state of suspects and then evaluating possible options for moral action. In this case, when a police officer is out of the loop, the AI system must have a well-engineered capacity for credible interpersonal communication and artificial empathy in order to correctly recognise the emotional states of individuals with whom it interacts. On the contrary, the police AI system will be indifferent to moral concerns involved in a situation related to an assigned police mission. For instance, in cases of disruptive and escalating massive street violence (e.g., protests, sports-related hooliganism), autonomous decision-making about the use of force against citizens by police drones and robots without the capacity for artificial empathy may, in turn, provoke a more violent response and spiral of violence. The impossibility of correct recognition of social emotions in endless numbers of individuals composing an unruly crowd, immersed in a fast-changing group mood, may lead to excessive police force instead of handling the conflicting situation towards pacification and resolution. Another example from police work could be the unpredictable operation of robots in dealing with persons suffering from mental illness. Emphasising the general need for adequate police response in the form of mental health assessment on

the spot, Rudofossi (2019, 636–637) pointed out that mentally ill citizens are additionally victimised as they are unable to interact normally with police, e.g., they cannot be expected to respond positively to any police officer's command. It is hard to imagine the positive outcome of the interaction between a person suffering from mental illness and a police device, empowered by either weak or strong AI, due to numerous uncertainties in terms of the person's unpredictable negative perceptions of and reactions to a drone or robot as an inanimate agent. Even if a robot is capable of recognising the emotional state of a mentally ill person, the person may be irritated, distressed, or frightened because of the inhuman appearance of the intelligent machine, which, for example, might trigger paranoid thoughts.

The police–citizen relationship is becoming increasingly mediated through AI-driven devices, creating a new type of social distance that seems to undercut empathic connectedness. A failed empathic response may originate from emotional distancing as well, which is a learned skill intertwined with police personality. According to Best and collaborators (2019, 567), the ability of emotional distancing (or detachment) refers to the trained response of a police officer who finds herself in a sudden and unexpected life-threatening situation or when she must use deadly force. Gallo (2019, 383–384) reminded us that police officers, in the course of professional socialisation, internalise an overemphasised image of the world as a jungle full of danger. Gallo claims that, contrary to the fact that most police–citizen interactions are resolved without force, police officers perceive the likelihood of armed confrontation to a far greater extent. For Gallo (2019, 382), a serious problem for the ethical dimension of police work occurs if this exaggerated perception eventually grows into “pseudo-threats to survival”, so much so that some officers, particularly

those with personality disorders, easily find themselves involved in abusive practices targeting allegedly “threatening” parts of the population. The utilisation of AI-driven devices furthermore increases emotional distancing and, consequently, impairs the empathic responsiveness of police managers and officers. Unlike in the case of emotional distancing as a coping mechanism, the dehumanisation of citizens subjected to police action comes out in parallel as a result of technologically induced emotional detachment.

Within the visual regime of drones, and of robots in the mid-future, based on the god’s-eye perspective, the screen is the visual interface between the reality of the crime scene and the drone operator that shapes and filters the way in which it is possible to see the conduct of a police mission. A police officer, i.e., drone operator or team leader, is experiencing the feeling of a palpable, on-the-spot presence; that is, the physical distance seems to disappear thanks to an all-seeing view of the designated area of the ongoing police mission. Although each activity is visible in its physical entirety from the sky, the video imagery processed and visualised by AI, generates specific mental images in the police officer’s brain that replicate a hybrid type of reality of police action, resembling hyperreality. AI-empowered devices facilitate the perception of being involved in the reality of police action to the extent that the drone operator, or the team leader, feels almost physically present and engaged as if on the crime scene. Yet video streaming of the ongoing police action from above lacks the totality of real sounds from the street as it is only followed by radio communication between the dispatcher/command and police officers on mission: unlike the officers on the spot, the command is missing that important piece of puzzle. More importantly for sound ethical decision making, the command lacks a factual basis for exercising the

ability of cognitive empathy. It is interacting “face-to-face” with suspects or other citizens subjected to the police action only through streamed video communication, which can as much deceive as it can reveal about the motivation and intent of targeted illegal behaviour. Without correct attribution of intent, it is impossible to correctly read, interpret, and anticipate the behavioural patterns of an individual in the dynamic context of an ongoing police mission: Does the individual pose a security threat as a criminal actor just pretending to be a law-abiding local resident or is she/he really a harmless citizen? As a necessary step towards moral sensitivity, attribution of intention requires correct recognition and interpretation of the emotional states of human agents, that is, the ability of cognitive empathy. The successful performance of this ability may be largely undermined if the police-populace interaction is mediated by an AI system because drones or robots deliver a situational image of reality that rather resembles that of advanced simulation video games.

By constructing in her/his mind a hybrid mental image of an ongoing police mission, assembled from many factual/visual fragments of AI-mediated representation of events that are taking place in physical reality, the police officer is experiencing ambivalent feelings of immersiveness and, at the same time, of disconnectedness. In a virtual environment that tends to absorb the reality of police action by using its one part to stand for the whole, it seems psychologically quite demanding to correctly read and contextualise the behaviour of the other with regard to her/his situation and to recognise if the situation bears any moral concerns at all. In the case of the utilisation of police autonomous robots, the elusive boundary between the physical reality of police actions and its factual/visual representation on desk monitors at the command centre or on laptops in patrol cars

will start to fade. In the case of fully autonomous AI-driven devices, the police staff might be downgraded to mere technicians whose responsibility would be to control whether the device is operating according to the assigned tasks and general guidelines on AI-human interaction. Affective empathic responses would become marginalised due to greater emotional distance between the police manager and officers assigned to a mission, and citizens subjected to police action. This outcome would be more probable if the responsibility for the unethical actions of AI-driven devices could be dispersed (the problem of many hands) or redirected by managers or officers in order to avoid getting punished.

Such practices may also be legitimised by the idea of moral neutrality, which is a bureaucratic stance that denies ethical dimensions to decisions made in the performance of public duty (Lewis and Gilman, 2005, 111; Martinez, 2009, 58–60). The idea of moral neutrality often-times motivates public servants to adopt a strategy of moral disengagement so as to cope with a difficult situation in their daily routine that discloses an ethical controversy. When professional practices demean or discourage the ability of moral sensitivity over an extended period of time, the phenomenon of moral disengagement is easier to emerge and flourish (Bandura, 2015, 48–69). Although, in most cases, moral failure is the result of disengagement in ethical deliberation (Werhane et al., 2013, 65–68), it may also stem from the false perception of the moral intensity level in some situations. A police officer will neglect the moral dimensions of a situation if she/he perceives it as of insignificant importance—the case of low moral intensity. Attributing low moral intensity to some situations in police action, in the end, leads to dehumanisation as the common denominator of disengagement practices. Emotional distancing, followed by

empathic disconnectedness, may strip police officers of compassion and may smooth the path for adopting abusive treatment of citizens with whom they interact. While it is, as Bandura (2015, 84) claimed, “difficult to inflict suffering on humanized persons without experiencing distress and selfcondemnation”, the mistreatment becomes easier when the persons are “diminished to subhuman beings”.

The intermediate character of AI-empowered devices, resulting in greater emotional distance and a higher level of empathic unresponsiveness, may help reproduce traditional as well as constitute new dehumanising practices in policing. Morally disengaged police officers might become prone to opt for the routine use of AI-driven devices in their interactions with the populace, even when it is not necessary or safe. AI systems seem to facilitate the justification of moral indifference by leaning on the bureaucratic utilitarian logic associated with the neoliberal principle of “value for money”, so wholeheartedly embraced in public governance in the age of late capitalism. An overreliance on AI-empowered devices may be attractive to morally disengaged police managers and officers because of the illusion it provides that the public duty is being performed through scientifically proven methods and cutting-edge technologies. In other words, the utilisation of AI gives the illusion of a police work “purified” from all “dirty” inconsistencies typical for human decision-making that is distorted by various cognitive biases. Yet, in the long run, AI-mediated interaction, in which citizens are objectified by their reduction to bits of digital information displayed on screens, might lead to the denial of the intrinsic moral value of those subjected to police actions.

Conclusion

George Orwell, Aldous Huxley, Yevgeny Zamyatin, Ray Bradbury, and many other authors of the dystopian genre in fictional literature shared their concerns with the role of technology as a means to control the population. The early twenty-first century might witness the birth of a new, worrisome, advanced technological tool with great potential for manipulation and control: algorithmocracy. In the pursuit of new technologies capable of fixing the same old problems in the effectiveness of public governance, “smartification” is the latest fashionable neoliberal strategy aimed at reconciling the safety of police officers and citizens with the utilitarian outlook of highly productive and efficient policing entangled in the cost/risk-benefit analysis. In addition, emerging technologies are also instrumentalised within the neoliberal ethos of social ordering with the goal of eliminating the traditional limitations of the human factor with a view to the deficiencies of psychological, cognitive, and physical abilities. The psychophysical limitations of police officers’ performance are expected to be abridged with the help of AI systems in the near and mid future.

I have not attempted to provide a comprehensive list of ethical concerns regarding the ways in which the police have embraced or plan to utilise AI-empowered devices, but to investigate how the design of AI systems incorporated in police drones and robots affects and may affect the ability of police managers and officers to exercise moral sensitivity. Having examined the current and future utilisation of remotely controlled and semi-autonomous AI-powered devices in the part of police work related to face-to-face interaction with citizens, I have identified several critical points with the potential of undermining

the ethical decision-making process or even inhibiting its initiation. In the phase of observing the crime scene and gathering preliminary information about a situation in which the police action against illegal behaviour is placed, the most critical moment for the police officer is the proper selection, i.e., extraction of credible information from an enormous volume of gathered, streamed, and matched data from various AI-assisted sensors, particularly regarding the reliability of both remote and artificial recognition of the emotional state of the individuals subjected to police action. The reliability of a read and interpreted emotional state is mostly contested by diverse obstructive practices of the policed population aimed at confusing AI-assisted sensors as well as by reduced police officer's cognitive capacities due to sensory and information overload (along with fatigue). In the phase of determining if the situation involves moral concerns to be taken into consideration and the need for an ethical response, the most critical points can be narrowed to one essential issue: A social distance as a side-effect of the police-citizen interaction mediated through high-tech but still AI-simulated hyperreality. Social distance gradually downgrades the level of empathic responsiveness and, in the long run, constitutes the practice of moral disengagement, imprinting it into the daily professional routine of the police.

In the circumstances of decades-long public distrust in the police, coupled with the government's request for highly efficient police work within the limits of reduced public money spending, smartification through AI arrives as a pragmatic manoeuvre in the form of re-branding policing by associating it with the fancy image of emerging technologies that are employed to eradicate human errors. Yet pragmatism, under the neoliberal logic of utilitarian calculus, is by definition adverse to ethical deliberation due to its particularistic nature

expressed in valuing an action primarily by its practicality and success. From the pragmatist outlook, a police officer is always right when she/he does what “gets the job done”, no matter how complex the moral dimensions of a situation may be. From the perspective of sound ethical decision-making, it is not only important what is done but whether it is done in a morally right or wrong way. Without the capacity for credible remote interaction, followed by artificial empathy, the promise of AI-augmented policing might easily slip into an unempathetic form of bureaucratised algorithmocracy.

The paper presents findings of a study developed as a part of the 2023 Research Plan of the Institute for Political Studies, Belgrade, and financed by the Serbian Ministry of Science, Technological Development and Innovation.

References

- Asada, M. (2015). Towards artificial empathy: How can artificial empathy follow the developmental pathway of natural empathy? *International Journal of Social Robotics*, 7(1), 19–33. <https://doi.org/10.1007/s12369-014-0253-z>
- Beetz, M., Chatila, R., Hertzberg, J., & Pecora, F. (2016). AI reasoning methods for robotics. In B. Siciliano & O. Khatib (Eds.), *Springer handbook of robotics* (pp. 329–356). Springer. <https://doi.org/10.1007/978-3-319-32552-1>
- Benjamin, W. (1979). *One-way street and other writings* (E. Jephcott & K. Shorter, Trans.). NLB.
- Best, S., Kirschman, E., & Artwohl, A. (2019). Critical incident reactions and early interventions. In J. Kitaeff (Ed.), *Handbook of Police Psychology* (pp. 564–580). Routledge.
- Bhargava, V., & Wan Kim, T. (2017). Autonomous vehicles and moral uncertainty. In P. Lin, R. Jenkins, & K. Abney (Eds.), *Robot ethics 2.0: From autonomous cars to artificial intelligence* (pp. 5–19). Oxford University Press. <http://dx.doi.org/10.1093/oso/9780190652951.003.0001>
- Brisinda, D., Fenici, R., & Sorbo, A. R. (2019). Methods for real-time assessment of operational stress during realistic police tactical training. In J. Kitaeff (Ed.), *Handbook of Police Psychology* (pp. 307–332). Routledge.
- Chen, B., Vondrick, C., & Lipson, H. (2021). Visual behavior modelling for robotic theory of mind. *Scientific Reports*, 11(424). <https://doi.org/10.1038/s41598-020-77918-x>
- CPT [Center on Privacy and Technology]. (2016). *The perpetual line-up: Unregulated police face recognition in America*. Center on Privacy and Technology, Georgetown Law. <https://www.perpetuallineup.org/report>
- Cristianini, N. (2017). Machines that learn: The mechanics of artificial minds. In D. Heaven (Ed.), *Machines that think* (pp. 25–67). New Scientist.

Curry, O. S. (2016). Morality as cooperation: A problem-centred approach. In T. K. Shackelford & R. D. Hansen (Eds.), *The evolution of morality* (pp. 27–52). Springer. <https://doi.org/10.1007/978-3-319-19671-8>

Davenport, T. (2021, February 3). The future of work now: AI-driven policing in Wilmington, NC. *Forbes*. <https://www.forbes.com/sites/tomdavenport/2021/02/03/the-future-of-work-now-ai-driven-policing-in-wilmington-nc/?sh=7538b82435af>

Dineva, E. (2022, March 21). The four success factors to effective adoption of AI in policing. *TechUK*. <https://www.techuk.org/resource/the-four-success-factors-to-effective-adoption-of-ai-in-policing.html>

Ein-Dor, P. (1999). Artificial intelligence: A short history and the next 40 years. In K. E. Kendall (Ed.), *Emerging information technologies: Improving decisions, cooperation and infrastructure* (pp. 117–140). SAGE Publications.

Gallo, F. J. (2019). Police use of force. In J. Kitaeff (Ed.), *Handbook of Police Psychology* (pp. 377–401). Routledge.

Garner, R. (2008). Psychology for police leadership. In J. Ruiz & D. Hummer (Eds.), *Handbook of police administration* (pp. 3–16). CRC Press.

Goleman, D. (2006). *Social intelligence: The new science of human relationships*. Bantam Books.

Gregg, A. (2019). *Autonomous police vehicles: The impact on law enforcement* [Master's thesis, Center for Homeland Defense and Security, Naval Postgraduate School]. Homeland Security Digital Library. <https://www.hsdl.org/c/view?docid=825205>

Grother, P., Ngan, M., & Hanaoka, K. (2019). *Face recognition vendor test (FRVT) part 3: Demographic effects*. National Institute of Standards and Technology, U.S. Department of Commerce. <https://doi.org/10.6028/NIST.IR.8280>

Hambling, D. (2018). *We: Robot—The robots that already rule our world*. Aurum Press.

- Hamm, J. A., Searle, R., Carr, J. D., & Rivers, L. (2021). Public vulnerability to the police: A quantitative inquiry. *Criminal Justice and Behavior*, 48(12), 1749–1769. <https://doi.org/10.1177/00938548211008489>
- Har, J. (2022, November 30). San Francisco will allow police to deploy robots that kill. *AP News*. <https://apnews.com/article/police-san-francisco-government-and-politics-d26121d7f7afb070102932e6a0754aa5>
- Hartman, L. P., & DesJardins, J. (2020). *Business ethics: Decision-making for personal integrity and social responsibility* (5th ed.). McGraw-Hill Education.
- Hartmann, M. R. K. (2018). Grey zone creativity. In N. R. Fyfe, H. I. Gundhus, & K. V. Rønn (Eds.), *Moral issues in intelligence-led policing* (pp. 161–181). Routledge.
- Hayward, K. (2021). Five smart city futures: A criminological analysis of urban intelligence. In M. Schuilenburg & R. Peeters (Eds.), *The algorithmic society: Technology, power, and knowledge* (pp. 172–192). Routledge.
- Hennigan, W. J., & Bennett, B. (2016, July 8). Dallas police used a robot to kill a gunman, a new tactic that raises ethical questions. *Los Angeles Times*. <https://www.latimes.com/nation/la-na-dallas-robot-20160708-snap-story.html>
- Hieida, C., & Nagai, T. (2022). Survey and perspective on social emotions in robotics. *Advanced Robotics*, 36(1–2), 17–32. <https://doi.org/10.1080/01691864.2021.2018103>
- Hill, K. (2021, January 6). Another arrest, and jail time, due to a bad facial recognition match. *The New York Times*. <https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html>
- Hillstrom, L. C. (2018). *Black Lives Matter: From a moment to a movement*. Greenwood.
- Hobson, Z., Yesberg, J. A., Bradford, B., & Jackson, J. (2023). Artificial fairness? Trust in algorithmic police decision-making. *Journal of Experimental Criminology*, 19(1), 165–189. <https://doi.org/10.1007/s11292-022-09496-8>

Hoffman, M. L. (2007). *Empathy and moral development*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511805851>

Holley, P. (2019, May 14). One solution for keeping traffic stops from turning violent: A robot that separates police officers from drivers. *The Washington Post*. <https://www.washingtonpost.com/technology/2019/05/14/one-solution-keeping-traffic-stops-turning-violent-robot-that-separates-police-officers-drivers/>

ICS.AI. (2023). *AI transformation for police*. <https://www.ics.ai/ai-transformation-for-police>

Imotions. (2023). *Imotions Core—The world's leading biosensor platform*. <https://go.imotions.com/software-brochure>

Jeon, M. (2017). Emotions and affect in human factors and human-computer interaction: Taxonomy, theories, approaches, and methods. In M. Jeon (Ed.), *Emotions and affect in human factors and human-computer interaction* (pp. 3-26). Academic Press. <https://doi.org/10.1016/B978-0-12-801851-4.00001-X>

Joh, E. (2018). Artificial intelligence and policing: First questions. *Seattle University Law Review*, 41(4), 1139–1144. <https://ssrn.com/abstract=3168779>

Joh, E. (2022). Ethical AI in American policing. *Notre Dame Journal on Emerging Technologies*, 3(2), 261–287. <https://ssrn.com/abstract=4096953>

Josephson, M. S. (2002). *Making ethical decisions*. Josephson Institute of Ethics.

Joyce, R. (2006). *The evolution of morality*. MIT Press.

Kenning, C. (2018, January 31). No charges for Dallas officers who killed sniper with robot bomb. *Reuters*. <https://www.reuters.com/article/us-texas-crime-idUSKBN1FK35W>

King, S. (2020). Anti-Black violence, African American memory, and the Trayvon generation. *The American Historian*. <https://www.oah.org/tah/issues/2020/history-for-black-lives/anti-black-violence-african-american-memory-and-the-trayvon-generation/>

Kitsos, P. (2020). The limits of government surveillance: Law enforcement in the age of artificial intelligence. In G. Giannakopoulos, E. Galiotou, & N. Vasillas (Eds.), *Workshops of the 11th EETN Conference on Artificial Intelligence 2020* (pp. 164–168). CEUR Workshops Proceedings, 2844. <https://ceur-ws.org/Vol-2844/ethics11.pdf>

Klincewicz, M. (2017). Challenges to engineering moral reasoners. In P. Lin, R. Jenkins, & K. Abney (Eds.), *Robot ethics 2.0: From autonomous cars to artificial intelligence* (pp. 244–257). Oxford University Press.

Szocik, K., & Abylkasymova, R. (2021). Ethical issues in police robots: The case of crowd control robots in a pandemic. *Journal of Applied Security Research*. <https://doi.org/10.1080/19361610.2021.1923365>

Korać, S. T., & Stekić, N. (2018). Etičke oprečnosti teze o demokratskom miru [Ethical contradictions of the democratic peace thesis]. *Srpska politička misao*, 54(4), 95–117.

Korać, S. T. (2018). Depersonalisation of killing: Towards a 21st-century use of force “beyond good and evil?”. *Philosophy and Society*, 29(1), 49–64.

Kutnjak Ivković, S. (2014). Police misconduct. In M. D. Reisig & R. J. Kane (Eds.), *The Oxford handbook of police and policing* (pp. 302–336). Oxford University Press.

Labossiere, M. (2017). Testing the moral status of artificial beings. In P. Lin, R. Jenkins, & K. Abney (Eds.), *Robot ethics 2.0: From autonomous cars to artificial intelligence* (pp. 293–306). Oxford University Press.

Lewis, C. W., & Gilman, S. C. (2005). *The ethics challenge in public service*. Jossey-Bass.

Martinez, M. (2009). *Public administration ethics for the 21st century*. ABC Clio.

Matheson, R. (2018, July 24). Helping computers perceive human emotions. *MIT News*. <http://news.mit.edu/2018/helping-computers-perceive-human-emotions-0724>

McGuire, E. R. (2014). Police organizations and the iron cage of rationality. In M. D. Reisig & R. J. Kane (Eds.), *The Oxford handbook of police and policing* (pp. 68–98). Oxford University Press.

McGuire, M. R. (2021). The laughing policebot: Automation and the end of policing. *Policing and Society*, 31(1), 20–36. <https://doi.org/10.1080/10439463.2020.1780184>

Meacham, D., & Studley, M. (2017). Could a robot care? It's all in the movement. In P. Lin, R. Jenkins, & K. Abney (Eds.), *Robot ethics 2.0: From autonomous cars to artificial intelligence* (pp. 97–112). Oxford University Press.

Medlock, B. (2017, March 14). The body is the missing link for truly intelligent machines. *Aeon*. <https://aeon.co/ideas/the-body-is-the-missing-link-for-truly-intelligent-machines>

Mintrom, M., Sumartojo, S., Kulić, D., Tian, L., Carreno-Medrano, P., & Allen, A. (2022). Robots in public spaces: Implications for policy design. *Policy Design and Practice*, 5(2), 123–139. <https://doi.org/10.1080/25741292.2021.1991244>

Mynenko, K. D., & Ditcham, K. (2022, February 11). Public confidence in the police: A new low for the service. *The Royal United Services Institute for Defence and Security Studies (RUSI)*. <https://rusi.org/explore-our-research/publications/commentary/public-confidence-police-new-low-service>

Negnevitsky, M. (2005). *Artificial intelligence: A guide to intelligent systems*. Pearson Education.

Neocleous, M. (2014). *War power, police power*. Edinburgh University Press.

NIST [National Institute of Standards and Technology]. (2019). NIST study evaluates effects of race, age, sex on face recognition software. *The Journal of Research of the National Institute of Standards and Technology*. <https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software>

Oppy, G., & Dowe, D. (2021, October 4). The Turing test. *Stanford Encyclopedia of Philosophy* (Winter ed.). <https://plato.stanford.edu/entries/turing-test/>

Paulsen, J. E. (2021). AI, trustworthiness, and the digital Dirty Harry problem. *Nordic Journal of Studies in Policing*, 8(2), 1–19.

[PEG] Paul Ekman Group. (2023). *Facial Action Coding System*. <https://www.paulekman.com/facial-action-coding-system/>

Pentland, A. (2007). Social signal processing [Exploratory DSP]. *IEEE Signal Processing Magazine*, 24(4), 108–111.

Picard, R. (1997). *Affective computing*. MIT Press.

Picard, R. (2010). Affective computing: From laughter to IEEE. *IEEE Transactions on Affective Computing*, 1(1), 11–17.

Rudofossi, D. (2019). Mentally ill—Crying “Help!” Police and public health prevention, intervention, and postvention. In J. Kitaeff (Ed.), *Handbook of police psychology* (pp. 615–637). Routledge.

Sharkey, N. (2016, July 12). Are we prepared for more killer police robots? *The Guardian*. <https://www.theguardian.com/commentisfree/2016/jul/12/killer-police-robots-legal-consequences-dallas>

Sinder, S., & Simon, M. (2016, July 12). How robot, explosives took out Dallas sniper in unprecedented way. *CNN*. <https://edition.cnn.com/2016/07/12/us/dallas-police-robot-c4-explosives/index.html>

Sklansky, D. A. (2007). *Democracy and the police*. Stanford University Press.

Steverson, L. A. (2008). *Policing in America: A reference handbook*. ABC-CLIO.

Szocik, K., & Abylkasymova, R. (2022). Ethical issues in police robots: The case of crowd control robots in a pandemic. *Journal of Applied Security Research*, 17(4), 530–545. <https://doi.org/10.1080/19361610.2021.1923365>

Šabanović, S. (2010). Robots in society, society in robots: Mutual shaping of society and technology as a framework for social robot design. *International Journal of Social Robotics*, 2(4), 439–450. <https://doi.org/10.1007/s12369-010-0066-7>

Tao, J., & Tieniu, T. (2005). Affective computing: A review. In J. Tao, T. Tieniu, & R. W. Picard (Eds.), *Affective computing and intelligent interaction—First international conference, ACII 2005, Proceedings* (pp. 981–995). Springer. https://doi.org/10.1007/11573548_125

Thomas, M. (2022, August 25). Are police robots the future of law enforcement? *Built In*. <https://builtin.com/robotics/police-robot-law-enforcement>

Tomasello, M. (2016). *A natural history of human morality*. Harvard University Press.

UNICRI & INTERPOL. (2019). *Artificial intelligence and robotics for law enforcement*. United Nations Interregional Crime and Justice Research Institute and International Criminal Police Organization. https://unicri.it/sites/default/files/2019-10/ARTIFICIAL_INTELLIGENCE_ROBOTICS_LAW%20ENFORCEMENT_WEB_0.pdf

Wagner, L. (2015, August 27). North Dakota legalizes armed police drones. *NPR*. <https://www.npr.org/sections/thetwo-way/2015/08/27/435301160/north-dakota-legalizes-armed-police-drones>

Washburn, E. (2023, February 3). America less confident in police than ever before: A look at the numbers. *Forbes*. <https://www.forbes.com/sites/emilywashburn/2023/02/03/america-less-confident-in-police-than-ever-before-a-look-at-the-numbers/?sh=268e92746afb>

Weng, Y.-H., Chen, C.-H., & Sun, C.-T. (2009). Toward the human-robot co-existence society: On safety intelligence for next generation robots. *International Journal of Social Robotics*, 1(4), 267–282. <https://doi.org/10.1007/s12369-009-0022-6>

Werhane, P. H., Hartman, L. P., Archer, C., Englehardt, E. E., & Pritchard, M. S. (2013). *Obstacles to ethical decision-making: Mental models, Milgram, and the problem of obedience*. Cambridge University Press.

Westerman, D., Edwards, A. P., Edwards, C., Zhenyang, L., & Spence, P. R. (2020). I-It, I-Thou, I-Robot: The perceived humanness of AI in human-machine communication. *Communication Studies*, 71(3), 393–408. <https://doi.org/10.1080/10510974.2020.1719311>

Williams, V. S., & Williams, B. O. (2008). Technology applications: Tools for law enforcement. In J. Ruiz & D. Hummer (Eds.), *Handbook of police administration* (pp. 165–173). CRC Press.

Strategies for Cybersecurity Innovation and Crisis Communication in Digital Democracy Era

Mokhtarrudin
Ahmad
Kavitha
Balakrishnan

Convergence of Democracy, Security, and Technology

The dawn of the Digital Democracy Era has ushered in unprecedented challenges and opportunities at the nexus of democracy, security, and technology (Mergel, 2021). A significant change in the way society's function and are governed has been signalled by the dawn of a new era marked by the integration of digital technologies into these core aspects, which has presented previously unheard-of opportunities as well as challenges. Elections and political discourse are impacted by

Mokhtarrudin Ahmad, Associate Professor, Faculty of Applied Communication, Multimedia University, Kuala Lumpur ORCID: 0000-0002-3696-3015, email: mokhtarrudin@mmu.edu.my

Kavitha Balakrishnan, Assistant Professor, Faculty of Applied Communication, Multimedia University, Kuala Lumpur ORCID: 0000-0003-4446-2656, email: kavitha.balakrishnan@mmu.edu.my

DOI: <https://doi.org/10.5281/zenodo.14204895>

the pervasive use of digital platforms, such as social media, for political activity (Bronstein, Aharony and Bar-Ilán, 2018). Opportunities could include improved communication and easier access to information, while obstacles could include things like disinformation and cybersecurity risks (de Cock Buning, 2018; Schünemann, W. J. (2022). Political communication has evolved because of the growing popularity of social media platforms like Facebook and Twitter, which allow public figures and political leaders to communicate in real time. It has, however, also sparked concerns about the dissemination of false information and the impact of digital media on political viewpoints.

The intertwining of these three dynamic pillars has given rise to pressing concerns that demand careful consideration and strategic responses. Democracy, security, and technology are closely related, and this has led to important and pressing problems that call for careful consideration and well-thought-out solutions (Whyte, Thrall, and Mazanec, 2020). These “dynamic pillars” suggest that these components are dynamic, continuously evolving, and that sustained attention is needed to their relationships (Howard, 2020; Carayannis, and Campbell, 2021). Within the realm of cybersecurity, this may allude to the requirement for flexible approaches to tackle dynamic cyber-threats (Singh et al., 2022) within the framework of democratic procedures. To create and execute cybersecurity policies that are effective, political leaders and technological specialists must work together (Lehto and Limnell, 2021). In particular, the use of digital voting systems presents new security risks, and safeguarding election integrity necessitates calculated solutions, such as strong cybersecurity defences and public education initiatives.

Democracy, once confined to traditional political landscapes, now navigates the complex terrain of the digital realm, where the very foundations of societal governance face both enhancement and disruption. Conventional notions of democracy, which were hitherto restricted to real political settings, are currently adjusting to the intricacies of the digital age. The underlying ideas that guide societies are both improved and disrupted by this adaptation (Helbing, 2021; Helbing and Helbing, 2021). This could be a reference to the creation of virtual communities for e-governance and civic participation projects. While these developments improve accessibility, they also present new security and privacy problems. Democratic involvement can be increased by online platforms that let people communicate their thoughts on policies or participate in decision-making processes. Nevertheless, concerns about digital manipulation and the safety of personal information emerge.

Simultaneously, the exponential growth of technology, particularly in the realms of cybersecurity, presents a dual-edged sword, promising innovation while posing threats that transcend traditional borders. Beyond traditional physical borders, there are concerns associated with the rapid improvements in technology, particularly in the domain of cybersecurity. This could be termed as a “dual-edged sword” because there are advantages and disadvantages to these technical advancements. Within the realm of cybersecurity, this may pertain to ongoing technological advancements in areas such as artificial intelligence for threat identification and action, as well as the global reach of cyberthreats that have the potential to affect any country in the world. The total security posture is improved by the implementation of cutting-edge cybersecurity technologies, such as machine learning algorithms for cyberattack detection and prevention. Although

digital systems are interconnected, a cyberattack in a specific area of the world can have adverse consequences throughout the globe in general (Lawrence, Blackett and Cradock-Henry, 2020; Palleti, et al, 2021)

It is imperative to understand the convergence of democracy, security, and technology for exploring the strategies for cybersecurity innovation and crisis communication. To fully comprehend these strategies, it is necessary to acknowledge the convergence of democracy, security, and technology, emphasizing their mutual reliance. Researchers contend that in a time of growing cyberthreats, principles of democracy should be reflected in cybersecurity strategies that safeguard individuals' privacy and rights (Fuster and Jasmontaite, 2020; Savaş and Karataş, 2022). In the same way, crisis communication is obliged to fit the digital era given the speed at which information can be disseminated online (Sadri, Ukkusuri and Ahmed, 2021). Understanding the emergence of new technologies in the cybersecurity space requires taking a comprehensive approach that combines democratic values with technology solutions. Using encryption technology, for example, to safeguard user data is consistent with democratic ideals of privacy. Businesses could implement crisis communication techniques following a cyberattack, which include providing users with transparent information about the occurrence, outlining the response measures, and soliciting public feedback on potential future cybersecurity policies (Knight and Nurse, 2020; Kaschner and Kaschner, 2021).

The interdependence of these pillars shapes the social environment in which our societies develop and function. It also has a fundamental impact on how communities' function and advance. Considering how interconnected each of these connections is, modifications to one

frequently have an impact on the others. Technology advancements, depending on how they are applied, have the potential to strengthen or weaken democratic processes. Similarly, technological advancements can be impacted by security measures, which might influence their accessibility and ethical implications (van de Poel, 2020). Sustaining trust requires careful consideration of user privacy and data protection, and technological advancements need to be in line with security measures to provide a secure online environment. The mutual dependence of these fundamentals is demonstrated by the way that biometric authentication mechanisms in smartphones not only improve personal device security but also furnish developers with considerations over privacy and monitoring.

While upholding fundamental principles of participation, representation, and accountability, its efficacy is now intricately tied to the digital tools that seamlessly integrate different technologies to facilitate communication, interaction, and decision-making. Digital tools facilitate easier access to information and engagement platforms, which can improve democratic involvement (Suherlan, 2023). Concerns over the digital gap, however, highlight how inclusive technological improvements must be necessary to achieve true representation (Lythreathis, Singh, and El-Kassar, 2022). For example, social media platforms are now fundamental to interaction and communication. Yet, problems like the dissemination of misleading and false information highlight how crucial it is to use these digital tools responsibly while adhering to ethical standards. Assume a large corporation is dealing with a crisis, perhaps a product recall brought on by safety concerns. To ensure openness and obtain important feedback, the company ought to interact with its stakeholders, including customers and employees. Email newsletters, social media, and internal collaboration

tools are examples of digital platforms that could assist and engage users. By giving the staff a virtual forum, these tools enable them to actively engage in conversations, voice their concerns, and offer solutions for resolving the situation. Nonetheless, concerns over the integrity and transparency of information sharing must be highlighted here.

Security, which was once limited to physical borders, is now extended to the digital domain, where data integrity and critical infrastructure resilience are of utmost importance. Specialists emphasize the necessity of all-encompassing security measures that transcend conventional borders (Chowdhury and Gkioulos, 2021). Furthermore, in an increasingly globalized world, integrating digital security measures is essential to protecting national, organizational, or individual interests. To combat digital threats, such as assaults on vital infrastructure like communication networks and power grids, cybersecurity as an industry is continually evolving. To defend against online attacks, governments and corporations are investing in technologies such as intrusion detection systems. The Stuxnet virus, which was designed to target Iran's nuclear facilities, is a prime example of how security concerns are shifting to the digital sphere, where sophisticated cyberattacks can have tangible consequences for critical infrastructure (Mohee, 2022).

Technology, as a catalyst for societal advancement, introduces new possibilities and challenges, necessitating a recalibration of our understanding of governance, security, and civic engagement. Blockchain and other emerging technologies have the potential to completely transform authorities by increasing accountability and transparency. Nevertheless, reservations about the moral ramifications of technology like facial recognition highlight the necessity of

using them with caution. Regulations are frequently outpaced by innovation. Industry leadership and policymakers must work together to create regulations that strike a balance between ethical concerns and technological advancement. The emergence of smart cities presents opportunities for effective governance by utilizing Internet of Things (IoT) technologies for management. Therefore, maintaining cybersecurity and data privacy in these networked systems becomes a crucial concern.

Role of Effective Communication

At the heart of this convergence lies the pivotal role of effective communication. Communication serves as the linchpin that binds democracy, security, and technology together. The ability to convey information, articulate policies, and engage citizens has become synonymous with the health of democratic institutions. Likewise, in the realm of cybersecurity, the effectiveness of communication determines how societies respond to threats, recover from incidents, and collectively fortify against future challenges. It is in this juncture that we turn our attention to the critical interplay between cybersecurity innovation and crisis communication.

At the heart of this convergence lies the pivotal role of effective communication. It implies that effective communication is the fundamental component that unites democracy, security, and technology, affecting how they collaborate and function. Building trust and promoting understanding are two more aspects of effective

communication in addition to information dissemination (Kalogiannidis, 2021). This emphasizes the importance of communication in handling complicated cybersecurity challenges in the business or professional setting (Nkongolo, 2023). Establishing a productive work culture, guaranteeing policy clarity, and bringing individuals into conformity with organizational goals all depend on effective communication. This becomes more important in times of crisis when it acts as a calming influence.

Effective communication serves as the linchpin that plays a critical and stabilizing role during times of crisis. Effective communication is essential to all institutions, not only because it keeps individuals informed and involved, but also because it is essential to security because it preserves public confidence and reduces uncertainty in times of crisis. In the workplace, good communication is crucial for coordinating, resolving issues, and building collaboration. In the event of a cybersecurity incident, communication plays a crucial role in organizing reactions and providing workers with reassurance regarding the organization's security procedures and measures.

The ability to convey information, articulate policies, and engage citizens has become integral to the health and functionality of institutions. To preserve the public's trust, communication must be transparent (Henderson et al., 2020). Others, on the other hand, can express worries about possible information manipulation and highlight the significance of ethical communication practices (van de Poel, 2020). Employees who receive effective communication adhere to the goals and values of the company (Yue, Men and Ferguson, 2021; Balakrishnan, Angusamy and Rosli, 2023). To preserve confidence and ensure the welfare of employees during a crisis, open communication

regarding policies, safety precautions, and recovery strategies is essential (Henderson et al., 2020).

Likewise, in the realm of cybersecurity, the effectiveness of communication determines how societies respond to threats, recover from incidents, and collectively fortify against future challenges. Cybersecurity experts stress that communication is a key component of incident response, helping to minimize damage and facilitate recovery. Hence, there is a need for public-private collaboration in communicating cybersecurity best practices to fortify societies against evolving cyber threats. Effective communication during a cybersecurity incident involves promptly informing stakeholders, outlining response plans, and offering guidance to employees. Proactive communication also includes training employees on cybersecurity practices to prevent future incidents.

Likewise, in the realm of cybersecurity, communication performance dictates how societies react to risks, bounce back from difficulties, and unite to face fresh challenges. Experts in cybersecurity emphasize that effective communication is essential to crisis response since it reduces damage and accelerates recovery. Therefore, to strengthen against constantly evolving cyber dangers, public-private collaboration is required in the dissemination of cybersecurity best practices. In the event of a cybersecurity crisis, it is imperative to communicate effectively by providing employees with direction, describing response strategies, and quickly alerting stakeholders. Employee cybersecurity training is another aspect of proactive communication that helps prevent future problems.

It is in this juncture that we turn our attention to the critical interplay between cybersecurity innovation and crisis communication. It

suggests that managing the challenges presented by digital threats and crises requires an understanding of how these two components interact. According to Petrenko (2022), the successful adoption and exploitation of cybersecurity technologies requires excellent communication strategies in addition to innovation in the field. Adaptive crisis communication plans are necessary due to the evolving nature of cyber threats and the requirement for continuous communication. The integration of innovation and communication in addressing cybersecurity concerns entails not only applying sophisticated security measures but also successfully educating members about these developments. It also entails improving crisis communication strategies in the context of the knowledge gained from past mistakes.

In short, communication is a catalyst for understanding, trust-building, and collaboration, in addition to being an instrument for information dissemination. It is critical to acknowledge that communication is both a means and an objective in the process of examining strategies for cybersecurity innovation and crisis communication. It serves as a channel for the introduction, adoption, and dissemination of cybersecurity technologies among many facets of society. In addition, it serves the purpose, as proficient crisis communication is essential to both the defence of institutions and the mitigation of cybersecurity risks.

Chapter Overview

This chapter provides an in-depth analysis of “Strategies for Cybersecurity Innovation and Crisis Communication in the Digital Democracy

Era.” Our exploration will take place in individual chapters as we explore the complicated structure that unites democracy, security, and technology. Each section will add to a more nuanced understanding of the potential and challenges inherent with this common goal. By providing communication specialists, decision-makers, and stakeholders with a fundamental knowledge of the critical role communication plays in navigating the challenges of the Digital Democracy Era, this chapter sets the stage for practical and informative exploration.

In the next chapter, we will examine the theoretical underpinnings of the Diffusion of Technology Theory, gaining insight into how this theory influences the uptake and diffusion of cybersecurity technologies. Next, we will explore the domain of crisis communication theory, analysing the fundamental ideas that support efficient communication in the face of cybersecurity emergencies and threats. By combining these theories, options for building a resilient digital democracy will become clearer, highlighting the necessity for an anticipatory approach in a time of rapid technological advancement and uncertainty.

Diffusion of Innovations

Theory: Fostering

Cybersecurity Innovation

Effective crisis communication is integral to navigating the evolving landscape of cybersecurity innovation within the Digital Democracy Era. This chapter explores the application of Diffusion of Innovations Theory from a crisis communication standpoint, shedding light on its role in fostering cybersecurity innovation. Diffusion of Innovations

Theory, initially focused on the adoption of new technologies, is examined here through the lens of crisis communication.

E.M. Rogers developed the Diffusion of Innovation (DOI) Theory in 1962, providing a fundamental framework for comprehending how innovations are adopted by social systems (Rogers, 1962). The theory, rooted in communication studies, explains how concepts, behaviors, or products gain traction and spread within specific groups of people. Adoption, a crucial component of diffusion, denotes people acting differently than before, such as by buying and employing a new product or adopting an unconventional behavior (Rogers, 1962).

The DOI identifies four major variables influencing the adoption of innovative behavior: the social system, adopters' characteristics, communication channels, and innovation qualities (Mo et al., 2021; Rogers, 2003). It provides a framework for investigating the acceptance of novel concepts or behaviors and clarifies the transmission of innovations within social systems. Four dimensions are identified by the theory: time, social systems, communication channels, and innovation itself. Innovation is defined as any novel concept, item, process, or system with the attributes of relative advantage, compatibility, complexity, trialability, and observability. The last factor, social systems, represents the context in which diffusion takes place. Innovations are not adopted simultaneously by every member of a social system; rather, the diffusion process takes time. Researchers have identified different adopter groups, each distinguished by qualities influencing their inclination to accept innovations. Innovators, due to their daring disposition, are the first to accept novel concepts without seeking strong convincing. Early Adopters, considered opinion leaders, are

open to change and frequently need educational materials such as instructional guides (Jabro, 2011).

Despite not being leaders, the Early Majority adopts innovations ahead of the typical individual in search of proof of their efficacy (Taylor and Perry, 2005). The Late Majority is resistant to change and only accepts innovations with the consent of the majority; hence, data on successful adoptions is required. Laggards, due to their ties to tradition, are the most resistant and need persuasion to change by numbers, fear appeals, or pressure from other adopter groups (Jabro, 2011).

Stages of Innovation Adoption in Cybersecurity

Planning for effective crisis communication in the field of cybersecurity requires an awareness of the stages of innovation adoption. The DOI Theory sheds light on the dynamics of innovation adoption by emphasizing relative benefit, compatibility, complexity, trialability, and observability (Rogers, 2003). When promoting innovations, it is important to understand the characteristics of target groups because different techniques are used to appeal to different categories of adopters. Cybersecurity crises require a deliberate approach to crisis management, ranging from hacking efforts to phishing assaults. Communicators must negotiate the intricacies of cyber emergencies at different phases while working with IT and security departments. Reputation management becomes crucial before, during, and after a cybersecurity incident.

To plan a thorough and coordinated reaction to lessen the effects of cyberattacks and data breaches, communicators must work closely with IT and security departments. Continuous monitoring and gathering information are necessary for the proactive detection and mitigation of cybersecurity vulnerabilities. When communicators collaborate with technical teams, they help identify vulnerabilities and malicious activity early on. This kind of proactive approach is critical when cybersecurity concerns first arise. It is imperative to have a comprehensive crisis management plan that includes well-defined standards and processes. Communicators are involved in the development and execution of plans that tackle the technical and communication facets of the crisis, working closely with IT and security departments. Planning a crisis communication strategy that works is essential to handling public relations amid cyberattacks.

Communicators make sure that cyber issues are handled quickly and effectively, limiting reputational harm by being transparent, including stakeholders, and disseminating accurate information on time. **Strategies for Engaging Stakeholders:** Communicators facilitate stakeholder engagement by means of consistent message, proactive outreach, timely resolution of problems, regular updates, and the provision of resources and support. Stakeholders comprise employees, consumers, partners, and regulators. During a cyber crisis, these communication strategies help restore confidence and trust. The following sections examine how the DOI Theory is used in crisis communication and how innovations spread to response techniques in new media.

Early Adopters and Leaders in Cybersecurity

Social media has changed the communication landscape, particularly during times of crisis. Information distribution and audience targeting techniques have changed because of social media platforms, such as corporate websites, Facebook, YouTube, and Twitter (Jabro, 2011). Social media enables real-time, focused communication with specialized and broad audiences, in contrast to conventional notification systems during crises. The socioeconomic position of the people, the availability of technology for universal notification, and emergency responders' investigation of alternative information dissemination channels were found to be the three main obstacles to the spread of social media in crisis response. This emphasizes how crucial it is to comprehend the traits of the intended audience as well as the larger societal structure when mitigating crisis communication plans.

Experiences such as the global spread of COVID-19 demonstrate how innovation diffusion takes on a unique and complex nature during emergencies. The virus led to unprecedented global responses, such as lockdowns and orders to stay at home, which had a significant effect on businesses especially involved in hospitality, tourism, and event management (Dillette and Ponting, 2021; Baum and Hai, 2020; Jamal and Budke, 2020). Using the Diffusion of Invention (DOI) theory, crises such as the COVID-19 pandemic might provide valuable insights into the invention process during times of crisis. DOI offers a structured technique for comprehending emerging innovations in industries within the framework of crisis management (Dillette and Ponting, 2021). Information systems, strategic management,

healthcare, hospitality, tourism education, and business are just a few of the sectors where research has utilized DOI (Currie and Spyridonidis, 2019; Goh and Sigala, 2020; Vargo et al., 2020).

Researchers have also used DOI to investigate innovations in institutional contexts and service-centered ecosystems, recognizing that all involved parties are crucial during the adoption process (Vargo et al., 2020). However, the DOI theory, which is widely recognized for its applicability in a variety of areas, offers a strong foundation for comprehending the challenges associated with adopting innovative behavior (Mo et al., 2021). Crisis communication experts can contend that the idea offers a useful framework for comprehending how innovations—particularly in cybersecurity—spread among various societal segments during a time of crisis. It supports pattern analysis, potential challenges identification, and strategy development for effective communication. This adaptation emphasizes how applicable the Diffusion of Innovations Theory is to cybersecurity crisis management. For instance, in the event of a data breach, it is essential to understand how employees accept new security procedures or technology to develop communication plans that improve comprehension and compliance.

Communication Channels for Promoting Cybersecurity Best Practices

Social media usage varies throughout organizations dealing with emergencies. About half of the companies going through crises

included the Internet in their response plans, according to five-point-in-time research on the integration of social media, or the Internet, in crisis communication (Taylor and Perry, 2005). This emphasizes the changing nature of Internet adoption during crises and its ability to impact media relations. It is based on Rogers' diffusion of innovation research. People increasingly choose to express their dissatisfaction on social media during times of crisis (Ly-Le, 2018). The study established that social media, which is thought to have relative advantages and observability, is likely to be adopted in crisis communication by applying the Diffusion of Innovation (DOI) theory.

However, the dimensions of complexity and trialability issues surfaced as possible roadblocks, causing companies to be reluctant to fully adopt social media for crisis communication. This result supports the DOI hypothesis by indicating that innovations should be carefully considered before being adopted. As a result, it is recognized that crisis communication is changing, and that social media is becoming increasingly vital. It highlighted the need for businesses to adjust to emerging technologies, including social media, to keep up with trends and stakeholders' needs,

This entails knowing how different sections of society are informed about and receptive to innovations—especially in cybersecurity—during times of crisis. From the standpoint of crisis communication, researchers could stress how crucial it is to modify strategies for communication considering the unique characteristics and interests of various social segments. This concept helps in developing communications that resonate with each group, such as how communication to customers and employees may differ from one another. In times of crisis, such as a widespread cyberattack, it is critical to share

information to various stakeholders, such as the general public, governmental organizations, and industry partners.

The Diffusion of Innovations Theory aids in communication strategy development to guarantee effective understanding and adoption of cybersecurity solutions. Experts in crisis communication believe that the theory's fundamental principles can direct the delivery of vital cybersecurity information, guaranteeing that it reaches a broad range of audiences during an emergency. Experts in crisis communication may point out that developing communication methods that appeal to various audience segments can be greatly aided by the theory's principles, which include recognizing early adopters and comprehending communication channels.

This guarantees that cybersecurity information is received and adopted in addition to being given on time. Experts may advise applying the Diffusion of Innovations Theory's concepts to communication plans in a workplace where cybersecurity is an issue, considering employees' differing degrees of technology literacy. For example, early adopters could be used as a means of influencing their colleagues and as agents of change. Communicators have a crucial role in selecting and exploiting efficient channels of communication to advance cybersecurity best practices when managing cybersecurity disasters. This calls for strategic communication at several points along the way, with an emphasis on openness, including stakeholders, and promptly disseminating correct information.

Establishing and upholding open lines of communication is essential to developing stakeholder trust. Communicators make ensuring that communication is organized and consistent while giving internal and external audiences accurate details on time. Communicators place

a high priority on the fast and accurate delivery of information. They also prioritize regular updates. Frequent updates that are in line with incident response protocols help to control stakeholder expectations and guarantee a coordinated crisis resolution effort. It is also critical to maintain consistency in messages throughout all communication channels. To present a cohesive front to stakeholders, communicators work with IT and security teams to integrate communication with technical actions. In addition, by utilizing both conventional and social media platforms, communicators actively include stakeholders in their work. The organization is better equipped to respond to issues quickly and uphold its strong cybersecurity position because of this proactive outreach.

Moreover, organizations need to use post-crisis analysis as a key tool to learn from previous cybersecurity disasters. To improve continuously and increase resilience, communicators must strengthen communication lines, update security measures, and improve incident response methods. By highlighting the importance of communication in overcoming resistance, an investigation into the ways in which communication tactics impact the diffusion process involves an examination at the communication channels, and narratives that are used to promote the adoption of cybersecurity technologies. The rollout of new security procedures must consider effective communication techniques that facilitate easy adoption and comprehension of cutting-edge cybersecurity measures.

Examining the nuances of communication at every level while taking into consideration variables like the perception of danger, information-seeking habits, and the impact of social networks is essential. These techniques for adapting crisis communications engage

and effectively attract audiences at different stages of adoption. To conclude, the importance of crisis communication in applying the diffusion of innovations theory to cybersecurity innovation is emphasized in this chapter. The chapter offers helpful tips and strategies for encouraging the adoption of cutting-edge cybersecurity solutions through the perspective of crisis communication, supporting the growth of a robust digital democracy.

Crisis Communication

Theory: Navigating Cybersecurity Crises

Exploring Crisis Communication Theory

Due to the proliferation and widespread adoption of new technologies, individuals are significantly impacted by technology in their daily lives. This has increased the risk to the lives of the general population and society. Organisations have a crucial role in maintaining the well-being of the public, particularly their stakeholders. Nevertheless, under specific circumstances, cybersecurity crises can pose a substantial menace to both organisations and their stakeholders. In Malaysia, the number of recorded cyber threat cases in 2022 was 4,741, which is a 16% increase compared to the previous year. Moreover, cybercrime has evolved into a more intricate and elusive form, with numerous attacks specifically exploiting weaknesses in commonly utilised software and systems.

Asia Pacific has emerged as the primary hub for cybercrime in Asia. In the first quarter of 2023, organisations in the region encountered a weekly average of 1,835 attacks, which is much higher than the global average of 1,248 attacks per week. Asia Pacific maintained its position as the region with the highest cybercrime activity for the second consecutive year, representing 31% of global attacks. In Malaysia, there were 269,533 documented cases of phishing targeting Small Medium Enterprises in 2022, representing a 56% increase compared to 2019 (Fazlida, et al., 2022). This condition has made a substantial contribution to the risk faced by the business.

With the ongoing process of digitalization, Asia will become a growing target for hackers. One of the issues that triggered the problem is human mistake (Algami, Almesalm and Syed, 2019; Nobles, 2018; Maglaras, et al., 2016) and social engineering (Wang, Sun and Zhu, 2020; Aldawood and Skinner, 2019). Mistakes made by humans are unavoidable. Employees may inadvertently transmit sensitive information to unintended recipients or establish vulnerable passwords that can be easily deciphered by malicious individuals. Social engineering is significantly more malicious. Social engineering refers to the manipulation of individuals by external attackers to get private information and passwords, typically through the use of malware, ransomware, and phishing campaigns. In 2022, around US\$10 million was lost in Singapore as a result of phishing schemes targeting OCBC Bank. An effective approach is providing employees with comprehensive training in social engineering and fundamental cybersecurity methodologies. Organisations should establish a comprehensive security programme for relevant stakeholders, which should be supplemented by teaching stakeholders how to effectively address vulnerabilities.

Although there is a collective effort within the organisation to reduce human error, additional efforts are required to enhance cybersecurity knowledge throughout the entire region. In many parts of Asia, there is a shortage of qualified cybersecurity specialists, which leaves businesses ill-prepared to defend against sophisticated attacks (Furnell, Fischer and Finch, 2017). ISC2's analysis indicates that the number of cybersecurity experts in the Asia Pacific region grew by 15.6% in 2022. Nevertheless, the region is currently facing a shortage of over 2.2 million skilled workers in the industry.

The utilisation of numerous technologies and services in cybersecurity, known as fragmented cybersecurity, may first appear advantageous due to the presence of various layers of protection. According to a survey conducted by Oracle and KPMG, a significant majority of organisations worldwide, namely 78%, employ over 50 cybersecurity products in their operational processes. Although organisations have a wide range of options and innovative technologies to help them find solutions, these can pose a greater risk by complicating the identification and prioritisation of more critical threats for IT teams.

Therefore, in order to address these concerns, organisations should prioritise the management of cybersecurity as a method of controlling crisis communication. Crisis communication encompasses the exchange of information (Marsen, 2020), in different modes and structures, between an organisation and its target audience, concerning a crisis (Heath and O'Hair, 2020; Fearn-Banks, 2017). During a crisis, organisations are required to carry out crisis communication by gathering, analysing, and distributing crisis-related information (Haupt and Azevedo, 2021). The purpose is to decrease the public's

uncertainty and limit any harm to the organization's reputation (Ulmer, Sellnow, and Seeger, 2022; Coombs, 2012).

Scholars have devoted significant research in the past two decades to determining the most successful messaging techniques for crisis communication (Claeys and Opgenhaffen, 2021; Kim and Lee, 2018). Crisis managers have found resonance in the current study on crisis communication, which specifically examines tactics for crisis-response messages. These managers are responsible for handling the management of an organization's reputation after a crisis, with the goal of repairing its image and achieving positive outcomes from external audiences (Ravell, 2023; Heide and Simonsson, 2021).

Nevertheless, scholars have recently expressed apprehension regarding the excessive focus on the "external dimensions" of crisis communication research, while neglecting one of the most crucial target audiences for any organisation: its employees (Heide and Simonsson, 2021; Mazzei and Butera, 2021; Abdul Hamid et al., 2020; Strandberg and Vigsø, 2016). There has been a lack of scholarly attention on crisis communication for internal audiences (Mazzei et al., 2022; Frandsen and Johansen, 2017; Johansen, Aggerholm, and Frandsen, 2012).

Employees are crucial in enhancing an organization's competitive advantage due to their possession of extensive knowledge about the organisation and their networks of contacts with colleagues, supervisors, and customers (Alfawaire and Atan, 2021). Organisational success greatly benefits from the presence of such networks (Ribeiro-Navarrete et al., 2021). During a crisis, the important links between employees and their organisation might cause employees to have specific emotional and cognitive responses, such as insecurity, stress, emotions of betrayal, anxiety, and wrath (Mazzei et al., 2012).

Therefore, the distinct relationships between an organisation and its employees can impact how employees understand and assess the crisis and its communication. This, in turn, can result in employees forming specific expectations and responses towards their organisation that are specific to internal stakeholders (Shin and Hur, 2020; Koehler and Raithel, 2018; Men, 2014; Mazzei and Ravazzani, 2011).

Principles of Effective Crisis Communication

Nearly every organisation, regardless of its size or type whether it be for-profit, public, or non-profit, is bound to face a crisis at some juncture. Certain crises are significant and easily identifiable, such as health hazards, security threats, and environmental perils. While certain issues may not appear to be “crises” at first glance, they can still have a significant influence on an organization’s reputation, brand, or perception within its community (Carmeli and Schaubroeck, 2008). Regardless of its manifestation, one indisputable fact remains: the consequences of a crisis on the organisation are exacerbated if organisations lack preparedness (Knight and Nurse, 2020). An essential component of crisis planning is the development of a communication strategy.

It is essential for every organisation to acquire fundamental principles for effectively managing crisis communication, including strategies for collaborating with media outlets, consumers, response agencies, and, most crucially, its own staff. The effectiveness of crisis communication relies heavily on many essential principles that assist an

organisation or corporation in effectively handling a crisis while maintaining transparency and trust:

- 1 | **Punctuality (Being First):** Timely communication is of utmost importance. By taking the initiative to address the problem promptly, an organisation may create the ideal environment, effectively control the narrative, and demonstrate proactive engagement.
- 2 | **Accuracy and Truthfulness (Be Correct):** Organisations should distribute information that is exact and true, even if it is limited, when it is of paramount importance. Participating in speculation or spreading incorrect information could worsen the problem and damage trust. Timely updates are essential to ensure that new information is promptly communicated.
- 3 | **Credibility (Establishing Credibility):** Credibility is bolstered via the practice of being transparent and honest. It is imperative to avoid employing ambiguous, elusive, or misleading statements in order to maintain trust among stakeholders.
- 4 | It is crucial for organisations to demonstrate authentic concern and compassion for those impacted by the crisis in order to exhibit empathy. It imbues the reaction with a personal element and showcases a level of commitment that surpasses mere management of the situation.
- 5 | Encourage proactive communication by not just describing the issue, but also specifying the steps being taken to address it. Every company must prioritise taking immediate action and clearly communicating future plans to reassure stakeholders that steps are being taken to address the issue.

- 6 | Exhibit reverence: Manifest reverence towards persons affected by the matter and those involved in its settlement. Organisations should acknowledge the impact on individuals, communities, or stakeholders and demonstrate sensitivity in their communication.

Implementing these principles in crisis communication helps maintain transparency, manage the narrative, and foster trust during challenging situations. It is imperative to not only address the present issue but also safeguard reputation and connections in the future.

***Transparency, Accuracy,
and Public Trust in
Cybersecurity Crises***

Communicating risk involves conveying information. Therefore, a key part of trustworthy and successful risk communication is comprehending the factors that influence the perceived reliability of information. In a few studies (Shipunova et al., 2022; Nurse et al., 2011), scholars investigate inquiry by conducting a thorough examination of the trust literature. The work encompassed both offline and online domains, examining many important studies (Rose et al., 2022; Bunker 2020; Kelton et al., 2008; Moturu and Liu, 2011). These articles evaluated confidence in digital information, web content, and the emerging field of trust in social media information. The poll yielded a comprehensive compilation of more than 35 elements that influence an individual's decision to trust an information object. The extract from

a comprehensive list and classify the criteria determining trustworthiness are as Table 1.

Table 1: Perceived Information Trust Worthiness Factors

Source	Deception, ethics, identity, incentive, location of source, objectivity, popularity, positive intentions, recommendation, related resources, reputation, motivation, similarity to receiver beliefs
Information	Accuracy, believability, competence of information, consistency, corroboration, objectivity, predictability, presentation and format, provenance, recommendation, relevance, specificity, timeliness, topic, verifiability
End-user	Bonding, context and criticality, beliefs, disposition to trust, expertise, motivation, propensity, risk propensity, trust in general technology, familiarity, limited resources/choice

These characteristics are categorised based on their association with (i) the source of information (its origin or sender), (ii) the information itself, or (iii) the recipient of the message (end-user). Although straightforward, this analysis can be highly advantageous for system designers to identify the specific stages at which a user's confidence in information is established, or to pinpoint any flaws in the design that could unintentionally undermine trustworthiness. Based on this element listing, it is evident that the trustworthiness of an information message can be influenced by factors such as the source's identification, the relevancy and timeliness of the information, and the competence of the end-user.

These traits can also greatly affect the trustworthiness of a security-risk communication, such as a warning prompt, and its reception by an individual, in addition to their impact on information in general. A risk communication that is factual, explicit, correctly delivered, and familiar is more likely to gain trust and prompt action compared to a message that is not. Several of the variables listed in Table I are

currently employed to determine trustworthiness on the Internet. When considering Web browsers, the factor of identification plays a crucial role in the selection process. This involves associating names like Chrome and Internet Explorer with their respective sponsoring corporations, namely Google Inc. and Microsoft Corp.

Once the source has been determined, individuals typically take into account other factors such as the reputation, expertise, and popularity of the browser or company. Regarding security, it is worth noting that many browsers offer timely and detailed warning messages regarding websites that are infested with malware, prior to allowing users to access them. This highlights the need of timely information. These elements, especially those pertaining to the end-user and their cognitive state, will become more significant as system designers aim to establish a sense of trust and enable effective risk communication, resulting in the development of more secure and reliable systems.

Role of Communication Specialists and Cybersecurity Experts in Crisis Communication

Composition of Crisis Communication Team

The constitution of a crisis communication team may differ based on the magnitude and character of the organisation, as well as the particular demands of the crisis being addressed. Nevertheless, a

standard crisis communication team often comprises the subsequent essential members:

Within the realm of a cyber security crisis, every member of the crisis communication team assumes a crucial and interconnected role in effectively managing and reducing the consequences of the disaster. Each function has a distinct contribution to make in effectively managing a cyber security crisis:

During a cyber security crisis, the Team Leader is in charge of coordinating efforts to control the breach and ensuring that the crisis communication team responds immediately. The decisions taken by the Team Leader will encompass both communication strategy and technical considerations. This includes determining the appropriate timing and method for disclosing specific information about the cyber assault, while ensuring that security is not compromised.

The spokesperson is responsible for communicating precise and transparent information regarding the cyber security event to the media and stakeholders. Due to the intricate nature of cyber threats, it is imperative for the Spokesperson to collaborate closely with the Technical Expert in order to effectively convey complicated information without unintentionally revealing sensitive facts.

The function of the Information Coordinator becomes crucial in gathering and distributing precise information regarding the characteristics and consequences of the cyber-attack. Ensuring the security and confidentiality of data is of utmost importance, particularly when handling sensitive information pertaining to the breach.

The Legal Counsel plays a vital role in a cyber security crisis by evaluating the legal ramifications, ensuring compliance, and identifying potential repercussions associated with data breaches. Their primary responsibility will be to scrutinise statements, press releases, and other communications to guarantee the organization's adherence to pertinent data protection and privacy legislation.

The PR Specialist is responsible for overseeing the organization's reputation following a cyber security issue, by addressing concerns and exhibiting transparency. It is crucial to create messaging that effectively communicate the organization's dedication to cybersecurity and safeguarding client interests in order to restore trust.

The Technical Expert offers expertise on the technical components of the cyber assault, assisting the communication team in comprehending and conveying the extent and seriousness of the situation. They facilitate the translation of complex technical information into easily understandable language for successful communication with both internal and external stakeholders.

The duty of the Corporate Communication Coordinator is to handle communication with employees during a cyber security crisis. They are responsible for keeping employees informed about the breach and any potential effects it may have on their positions. They have a role in distributing information regarding measures employees can do to safeguard themselves and the organisation.

The Stakeholder Engagement Specialist is responsible for effectively communicating with external stakeholders regarding the cyber security incident and its implications for

their interests. Within the realm of a cyber assault, effectively managing relationships with customers, partners, and regulatory agencies becomes essential in order to showcase the organization's dedication to fixing the issue.

To summarise, successful communication in the midst of a cyber security crisis necessitates a synchronised endeavour from all members of the team. The Subject Matter Specialist's technical expertise and the Leader's ability to make informed decisions are essential for effectively managing the intricacies of a cyber security issue.

Conclusion

Efficient communication plays a crucial role in successfully managing a cybersecurity crisis, exerting its impact well in advance, during the turmoil, and even after the event. It is not just a tool, but a crucial element that influences an organization's ability to withstand challenges, maintain trust, and manage relationships with stakeholders. The effectiveness of communication is crucial in anticipating, managing, and recovering from crises. Prior to a crisis, it functions as a protective barrier, strengthening an organization's defences; throughout the chaos, it acts as a source of direction, guiding stakeholders through periods of uncertainty; and after the disaster, it serves as a navigational tool, guiding the process of recovery and rebuilding trust. Recognising the need of efficient communication at all stages is essential for enduring challenges and ultimately coming out stronger.

Engaging in good crisis communication necessitates employing a methodical strategy that is essential for successfully managing or anticipating unexpected difficulties. At the core of this method is the creation of a Crisis Communication Plan, a thorough plan that establishes the foundation for efficient and impactful reactions.

Firstly, the plan requires a careful assessment of potential risks, envisioning the worst possible situations, and undertaking thorough “what if” evaluations. The preparatory phase is crucial as it allows for the selection of actions that comprehensively fulfil the needs of stakeholders. Implementing proactive contingency planning guarantees that in the event of a disaster, carefully thought-out responses are easily accessible, thereby reducing the damage and strengthening the organization’s ability to recover. Furthermore, it is crucial to utilise a Crisis Communication Team when confronted with substantial hazards or intricate communication requirements. This team, consisting of senior officials, plays a crucial role in assessing the nature and extent of the problem. By working together with appropriate parties, they develop and coordinate communication methods, guaranteeing a unified and efficient response.

Selecting a prospective Spokesperson is a crucial element in pre-crisis communication. The ideal applicant should have extensive and in-depth understanding of the situation, preferably a senior executive within the organisation. This choice improves credibility and guarantees a skilled and self-assured presentation of difficult information, fostering trust among all involved. Developing a thorough compilation of factual information acts as a fundamental basis for precise and uniform communication. This summary statement effectively balances stakeholders’ information rights with the company’s privacy

concerns, providing concise and crucial elements that facilitate dependable communication in times of crisis.

The foundation of a communication strategy is to develop a Key Message. This meticulously composed message serves as a fundamental guideline for all future communication endeavours. By focusing the communication on this central point, it guarantees a unified storyline, providing the appropriate tone and context to efficiently communicate essential information.

It is crucial to anticipate such queries and include answers inside the communication package. Anticipating and preparing for expected questions offers a more thorough and informative communication package, hence improving the effectiveness of the crisis response. Essentially, these meticulous procedures highlight the significance of a thorough Crisis Communication Plan. This strategy not only enables prompt and efficient crisis responses but also highlights the need of organised and deliberate communication in reducing the impact of crises on stakeholders and the organisation as a whole.

Determine the Communication Channels

In the intricate web of crisis communication, selecting the most effective channels to disseminate information is critical for organizations. A meticulous approach in choosing communication channels involves assessing various factors.

Firstly, determining the content that necessitates personal delivery—either live or via media channels—is pivotal. For instance, sensitive or urgent information might warrant a live address by a company official, ensuring a personal and immediate connection with stakeholders. Conversely, considering media channels demands coordination strategies—this entails planning the timing, content, and spokesperson for media interactions.

Suppose a company faces an issue due to data leak, opting for a live announcement by the CEO ensures a personal touch, emphasizing the gravity of the situation, while media coordination enables widespread coverage and addresses diverse audiences. Moreover, the necessity for a press conference or a crisis hotline needs careful consideration based on the crisis's nature and urgency. These channels offer platforms for detailed information dissemination or immediate access for stakeholders seeking assistance or clarification during crises.

Leveraging digital platforms like websites, emails, and social media emerges as an efficient means to swiftly communicate vital information. For instance, disseminating updates, safety guidelines, or emergency procedures via social media ensure broad reach and timely communication to stakeholders. Guiding the switchboard on handling inquiries ensures uniformity and accuracy in information dissemination, particularly when addressing stakeholder queries. Additionally, assessing the viability of traditional communication tools like phones and faxes and exploring alternative channels such as meetings or advertisements becomes pivotal in ensuring comprehensive outreach. In essence, the selection of communication channels in crisis situations hinges upon the nature of information, urgency, and the audience's accessibility. A strategic mix of personal, media, digital, and

traditional channels ensures efficient and effective dissemination of information, mitigating the impact of crises on stakeholders and bolstering organizational credibility.

Coordinate Internal and External Communication

Coordinating internal and external communication during a crisis requires a strategic and ethical approach that prioritizes transparency and empathy towards stakeholders. A well-crafted communication plan ensures simultaneous release of information to the media, employees, and stakeholders. However, if simultaneous communication isn't feasible, prioritizing direct communication to essential stakeholders from the organization remains imperative. This approach builds trust and prevents misinformation or confusion among critical stakeholders.

Furthermore, withholding information or revealing it gradually can adversely impact the organization's credibility. In crisis situations, it's crucial to divulge pertinent information comprehensively from the outset, especially adverse news. This upfront disclosure, while not necessitating revealing every detail, fosters transparency and authenticity, dispelling any perception of deceit or concealment.

Moreover, maintaining honesty in communication, even about limitations in discussing specific details, is pivotal. If certain information cannot be disclosed due to sensitivity or incomplete data, acknowledging this limitation during communication ensures transparency

and cultivates trust among stakeholders. Empathy serves as a cornerstone in crisis communication. Acknowledging the emotional impact on stakeholders and delivering messages with humility and sensitivity is essential. Aiming to view the situation from the audience's perspective and using personal, relatable language helps in connecting with stakeholders and alleviating their concerns. Balancing the message by highlighting positive aspects without downplaying the negative elements is crucial in inspiring confidence and resilience.

Amidst the crisis, maintaining composure is challenging yet vital. Demonstrating calmness while delivering critical messages exudes control, confidence, and competence. This demeanor reassures stakeholders and fosters a sense of stability, even in tumultuous times. In summary, the principles governing crisis communication necessitate a delicate balance of transparency, honesty, empathy, and composure. Adhering to these principles not only ensures effective communication but also establishes trust, credibility, and resilience among stakeholders, mitigating the impact of crises on organizational reputation and fostering stronger stakeholder relationships.

Post-crisis Stage

In the aftermath of a crisis, the focus shifts from immediate response to post-crisis operations and recovery. This phase is crucial for restoring normalcy, rebuilding trust, and learning from the crisis to fortify the organization against future challenges. The performance criteria for the Crisis Management (CM) team at this juncture

are multifaceted, aiming to encapsulate, evaluate, and learn from the crisis response.

Accounting for and demobilizing resources is the initial step, ensuring a streamlined post-crisis transition. Initiating post-incident recovery is paramount, focusing on stabilizing operations, assessing damages, and strategizing the return to normality. Simultaneously, evaluating and documenting the effectiveness of operations forms the foundation for introspection and improvement.

Debriefing all relevant individuals becomes crucial, providing a platform to discuss experiences, lessons learned, and areas for enhancement. This debriefing process not only serves to close the crisis chapter but also initiates the critical phase of learning and improvement.

The restoration phase heralds the return to business as usual, emphasizing the repair of confidence in the organization and its offerings. Efforts are directed towards swift normalization, the resumption of standard operations, and declaring the end of the crisis officially.

Documenting and reviewing the crisis and its response is a meticulous process aimed at ensuring a comprehensive understanding and learning. Timely recording, feedback analysis from stakeholders and witnesses, identifying the crisis's root cause, and disseminating findings to relevant stakeholders constitute this phase.

Learning from the crisis entails evaluating its effects, appointing an independent crisis evaluation team, and building a comprehensive document log. This stage focuses on integrating lessons learned into the organization's crisis management system, ensuring a more robust approach in the future.

The upgrade phase marks the final steps in crisis management, implementing effective resolutions and evaluating their efficacy. Simultaneously, delivering all information promised to stakeholders becomes imperative, ensuring transparency, and sustaining trust.

Planning for post-crisis rebuilding involves forward-thinking, assessing potential damage, crisis prevention strategies, and communications planning. A crucial aspect is the consideration of communication strategies aimed at restoring confidence in the organization and its products.

In the post-crisis phase, while the organization resumes normal operations, continued communication remains pivotal. Following through on promises made during the crisis and providing updates on recovery progress, corrective measures, and investigation outcomes sustains stakeholder trust and transparency. The post-crisis phase, albeit signalling a return to regular operations, demands meticulous attention to communication, learning, and improvement. It stands as a critical juncture where the organization not only rebuilds but also fortifies itself against future uncertainties.

*Embracing a
Proactive Approach
to Cybersecurity
Communication*

Amidst a time characterised by swift advancements in digital dangers, it is essential for any organisation to prioritise the establishment

of strong and resilient cybersecurity measures. Effectively conveying proactive cybersecurity measures to stakeholders is crucial in promoting a cohesive and watchful approach to protecting organisational assets. This essay presents a detailed analysis of effective communication methods that may be employed to enhance cybersecurity within a business organisation.

The process of becoming proactive in cybersecurity starts with a thorough and careful evaluation of potential risks. The process revolves around the identification and assessment of organisational assets, including hardware, software, and data. It is crucial to convey the underlying reasoning for this evaluation to the individuals involved. Demonstrate how the act of prioritising assets and evaluating risks allows for the implementation of customised security measures. Highlight the importance of making well-informed decisions in order to properly safeguard crucial assets.

Stakeholders should comprehend that this policy embodies the organization's strategy towards cybersecurity. Articulate its constituents explicitly: standards for safeguarding data, directives for employee behaviour, and procedures for responding to incidents. Demonstrating the thoroughness of this policy guarantees that stakeholders acknowledge the organization's dedication to a safe workplace and align them with security goals.

The human element frequently emerges as the most vulnerable aspect in cybersecurity. It is essential to involve stakeholders in order to ensure their understanding of the importance of frequent employee training. Explain the importance of providing staff with knowledge and training on recognising and dealing with potential threats.

Highlighting the crucial contribution of employees in strengthening defences will generate backing for regular training programmes.

Highlight the need of implementing multi-factor authentication to stakeholders, along with the significance of software updates. Explain how this authentication technique strengthens access controls. Highlight the need of consistently updating software and systems with the most recent security updates. Highlight these procedures as proactive actions to discourage cyber threats and provide strong defence against weaknesses. Network monitoring, data backup, and security audits involve the use of instruments such as intrusion detection systems to provide effective surveillance of the network. Explain how the use of real-time monitoring assists in swiftly identifying and reducing potential hazards. Emphasise the significance of consistent data backups, both stored at the same location and at a different one. The goal of security audits is to consistently find vulnerabilities and strengthen cybersecurity defences. Stakeholders must comprehend the importance of ongoing development in upholding a resilient security posture.

Emphasise the importance of utilising cybersecurity professionals or teams to get an advantage. Illustrate how these professionals leverage their specialised expertise and experience to strengthen defensive measures. Highlight their responsibility in finding vulnerabilities and offering recommendations on industry standards. Demonstrate how an outsider viewpoint contributes objectivity in assessing and improving the organization's cybersecurity position.

Implementing proactive communication techniques enhances the effectiveness of cybersecurity efforts. Consistently providing updates on precautionary measures and effectively communicating Standard

Operating Procedures (SOPs) to stakeholders increases their awareness and readiness. Creating a Quick Response Team (QRT) demonstrates the organization's preparedness to address urgent cybersecurity situations promptly. Establish a clear hierarchy within the Communication Department to guarantee the consistent and accurate distribution of information, so guaranteeing that stakeholders receive timely updates.

To summarise, proactive cybersecurity necessitates a thorough and interconnected approach that is complemented by efficient communication tactics. Emphasising the significance and execution of risk assessments, policies, staff training, technological measures, expert engagement, and communication strategies to stakeholders promotes a shared dedication to upholding a secure organisational environment. In the end, proactive cybersecurity communication enhances resilience and unifies stakeholders in protecting against ever-changing cyber threats.

*Advancing the Cause
of a Resilient Digital
Democracy, Security,
and Technology*

In the midst of this digital transformation, the equilibrium between technological progress, economic development, and societal welfare becomes progressively crucial. To effectively adjust to these changes, it is necessary to engage in careful strategizing, demonstrate a

strong commitment to inclusivity, and actively tackle the difficulties that emerge in the always changing digital environment.

The internet has emerged as the fundamental basis of democracy, offering unparalleled access to information without any limitations. Social media platforms have magnified voices, facilitating the dissemination and consumption of information by individuals. The widespread access to information has given citizens the ability to actively participate in societal matters to an unprecedented extent.

The digital economy has transformed lifestyles by utilising technology to enable the mobility of productive economic sectors. Amidst the COVID-19 pandemic, the digital economy played a vital role in facilitating inclusive economic growth. Indonesia for example had a remarkable 11% increase in their Gross Domestic Product (GDP) in 2023, which serves as a testament to the robustness and flexibility of the digital economy during difficult circumstances.

Although the digital transition offers significant prospects, it also presents obstacles for local firms. The proliferation of online applications has resulted in the decline of several local businesses, despite the ease they offer. Governments, such as the one in Indonesia, have taken action by imposing regulations like the prohibition of TikTok Shop in order to save local enterprises. It is necessary to develop long-term strategies that incorporate security measures in order to revive and support these firms, ensuring their continued growth in the digital age.

SMEs have become prominent participants in the digital environment. Female entrepreneurs, specifically, have utilised technology to advertise their firms on the internet. Nevertheless, this increasing

interconnectedness also exposes individuals to cybersecurity vulnerabilities, which could potentially result in fraudulent activities or misuse of data. Ensuring a harmonious equilibrium between the empowerment of SMEs through digital means and the preservation of security is of utmost importance for their long-term growth.

Data innovation has been integrated with enterprises, leading to the emergence of four fundamental services: business process automation, information provision, customer interaction, and productivity solutions. The challenges in the realm of information technology encompass problem-solving, nurturing innovation, and optimising efficiency. In addition, the enactment of data protection legislation, exemplified by Malaysia's Personal Data Protection Act, demonstrates a worldwide endeavour to protect consumer data and govern organisations as data users.

References

- Abdul Hamid, A. S., Ahmad Anuar, A. Z., & Megat Ibrahim, N. Z. (n.d.). Managerial perspective on internal crisis communication (ICC). *European Proceedings of Social and Behavioural Sciences*.
- Aldawood, H., & Skinner, G. (2019). Reviewing cyber security social engineering training and awareness programs—Pitfalls and ongoing issues. *Future Internet*, 11(3), 73. <https://doi.org/10.3390/fi11030073>
- Alfawaire, F., & Atan, T. (2021). The effect of strategic human resource and knowledge management on sustainable competitive advantages at Jordanian universities: The mediating role of organizational innovation. *Sustainability*, 13(15), 8445. <https://doi.org/10.3390/su13158445>
- Algarni, M., Almesalm, S., & Syed, M. (2019). Towards enhanced comprehension of human errors in cybersecurity attacks. In *Advances in human error, reliability, resilience, and performance: Proceedings of the AHFE 2018 International Conference* (pp. 163–175). Springer. https://doi.org/10.1007/978-3-319-94334-3_16
- Balakrishnan, K., Angusamy, A., & Rosli, M. D. R. (2023). Two-way asymmetrical communication mediating internal communication and employee engagement. *Environment and Social Psychology*, 9(2), 2074. <https://doi.org/10.54517/esp.v9i2.2074>
- Baum, T., & Hai, N. T. T. (2020). Hospitality, tourism, human rights, and the impact of COVID-19. *International Journal of Contemporary Hospitality Management*, 32(7), 2397–2407. <https://doi.org/10.1108/IJCHM-03-2020-0242>
- Bronstein, J., Aharony, N., & Bar-Ilan, J. (2018). Politicians' use of Facebook during elections: Emotionally based discourse, personalization, and vividness. *Aslib Journal of Information Management*, 70(5), 551–572. <https://doi.org/10.1108/AJIM-03-2018-0067>

- Bunker, D. (2020). Who do you trust? The digital destruction of shared situational awareness and the COVID-19 infodemic. *International Journal of Information Management*, 55, 102201. <https://doi.org/10.1016/j.ijinfo-mgt.2020.102201>
- Carayannis, E. G., & Campbell, D. F. (2021). Democracy of climate and climate for democracy: The evolution of quadruple and quintuple helix innovation systems. *Journal of the Knowledge Economy*, 12(4), 2050–2082. <https://doi.org/10.1007/s13132-021-00778-x>
- Carmeli, A., & Schaubroeck, J. (2008). Organisational crisis-preparedness: The importance of learning from failures. *Long Range Planning*, 41(2), 177–196. <https://doi.org/10.1016/j.lrp.2008.01.001>
- Chowdhury, N., & Gkioulos, V. (2021). Cyber security training for critical infrastructure protection: A literature review. *Computer Science Review*, 40, 100361. <https://doi.org/10.1016/j.cosrev.2021.100361>
- Claeys, A. S., & Opgenhaffen, M. (2021). Changing perspectives: Managerial and legal considerations regarding crisis communication. *Public Relations Review*, 47(4), 102080. <https://doi.org/10.1016/j.pubrev.2021.102080>
- Currie, G., & Spyridonidis, D. (2019). Sharing leadership for diffusion of innovation in professionalized settings. *Human Relations*, 72(7), 1209–1233. <https://doi.org/10.1177/0018726718796175>
- de Cock Buning, M. (2018). *A multi-dimensional approach to disinformation: Report of the independent high-level group on fake news and online disinformation*. Publications Office of the European Union.
- Dillette, A., & Ponting, S. S. A. (2021). Diffusing innovation in times of disasters: Considerations for event management professionals. *Journal of Convention and Event Tourism*, 22(3), 197–220. <https://doi.org/10.1080/15470148.2020.1860847>
- Evans, M., Maglaras, L. A., He, Y., & Janicke, H. (2016). Human behaviour as an aspect of cybersecurity assurance. *Security and Communication Networks*, 9(17), 4667–4679. <https://doi.org/10.1002/sec.1657>

- Furnell, S., Fischer, P., & Finch, A. (2017). Can't get the staff? The growing need for cyber-security skills. *Computer Fraud and Security*, 2017(2), 5–10. [https://doi.org/10.1016/S1361-3723\(17\)30013-1](https://doi.org/10.1016/S1361-3723(17)30013-1)
- Fuster, G. G., & Jasmontaite, L. (2020). Cybersecurity regulation in the European Union: The digital, the critical, and fundamental rights. In M. Christen, B. Gordijn, & M. Loi (Eds.), *The ethics of cybersecurity* (pp. 97–115). Springer. https://doi.org/10.1007/978-3-030-29053-5_5
- Goh, E., & Sigala, M. (2020). Integrating information and communication technologies (ICT) into classroom instruction: Teaching tips for hospitality educators from a diffusion of innovation approach. *Journal of Teaching in Travel and Tourism*, 20(2), 156–165. <https://doi.org/10.1080/15313220.2020.1740636>
- Haupt, B., & Azevedo, L. (2021). Crisis communication planning and nonprofit organizations. *Disaster Prevention and Management: An International Journal*, 30(2), 163–178. <https://doi.org/10.1108/DPM-08-2020-0257>
- Heath, R. L., & O'Hair, H. D. (2020). The significance of crisis and risk communication. In R. L. Heath & H. D. O'Hair (Eds.), *Handbook of risk and crisis communication* (pp. 5–30). Routledge. <https://doi.org/10.4324/9780203891629>
- Heide, M., & Simonsson, C. (2021). What was that all about? On internal crisis communication and communicative coworkership during a pandemic. *Journal of Communication Management*, 25(3), 256–275. <https://doi.org/10.1108/JCOM-01-2021-0010>
- Helbing, D. (2021). *Next civilization: Digital democracy and socio-ecological finance—How to avoid dystopia and upgrade society by digital means*. Springer Nature. <https://doi.org/10.1007/978-3-030-62330-2>
- Helbing, D. (2021). Digital democracy: How to make it work? In *Next civilization: Digital democracy and socio-ecological finance—How to avoid dystopia and upgrade society by digital means* (pp. 249–268). https://doi.org/10.1007/978-3-030-62330-2_12

Henderson, J., Ward, P. R., Tonkin, E., Meyer, S. B., Pillen, H., McCullum, D., Tonson, B., Webb, T., Coveney, J., & Wilson, A. (2020). Developing and maintaining public trust during and post-COVID-19: Can we apply a model developed for responding to food scares? *Frontiers in Public Health*, 8, Article 369. <https://doi.org/10.3389/fpubh.2020.00369>

Howard, P. N. (2020). *Lie machines: How to save democracy from troll armies, deceitful robots, junk news operations, and political operatives*. Yale University Press. <https://doi.org/10.12987/9780300252415>

Jabro, A. D. (2011). The use of social media in crisis response strategy: A diffusion of innovations analysis. *Journal of Communications Media Studies*, 3(1), 95–109.

Jamal, T., & Budke, C. (2020). Tourism in a world with pandemics: Local-global responsibility and action. *Journal of Tourism Futures*, 6(2), 181–188. <https://doi.org/10.1108/JTF-02-2020-0014>

Johansen, W., Aggerholm, H. K., & Frandsen, F. (2012). Entering new territory: A study of internal crisis management and crisis communication in organizations. *Public Relations Review*, 38(2), 270–279. <https://doi.org/10.1016/j.pubrev.2011.11.008>

Kalogiannidis, S. (2021). Business trust and communication as predictors of business growth. *WSEAS Transactions on Business and Economics*, 18, 1512–1522. <https://doi.org/10.37394/23207.2021.18.138>

Kaschner, H., & Kaschner, H. (2021). Cyber crisis response. In *Cyber Crisis Management: The Practical Handbook on Crisis Management and Crisis Communication* (pp. 19–98). Springer. https://doi.org/10.1007/978-3-658-35489-3_3

Kelton, K., Fleischmann, K. R., & Wallace, W. A. (2008). Trust in digital information. *Journal of the American Society for Information Science and Technology*, 59(3), 363–374.

- Knight, R., & Nurse, J. R. (2020). A framework for effective corporate communication after cyber security incidents. *Computers & Security*, 99, 102036. <https://doi.org/10.1016/j.cose.2020.102036>
- Koehler, I., & Raithel, S. (2018). Internal, external, and media stakeholders' evaluations during transgressions. *Corporate Communications: An International Journal*, 23(4), 512–527. <https://doi.org/10.1108/CCIJ-11-2017-0112>
- Lawrence, J., Blackett, P., & Cradock-Henry, N. A. (2020). Cascading climate change impacts and implications. *Climate Risk Management*, 29, 100234. <https://doi.org/10.1016/j.crm.2020.100234>
- Lehto, M., & Limnell, J. (2021). Strategic leadership in cyber security, case Finland. *Information Security Journal: A Global Perspective*, 30(3), 139–148. <https://doi.org/10.1080/19393555.2020.1813851>
- Ly-Le, T. M. (2018). Social media use in crisis communication in Vietnam: An organizational viewpoint. In *Proceedings of the Fifth International Conference on Business, Management and Accounting (IBSM 2018)* (pp. 245–258). SSRN. <https://ssrn.com/abstract=3392766>
- Lythreathis, S., Singh, S. K., & El-Kassar, A. N. (2022). The digital divide: A review and future research agenda. *Technological Forecasting and Social Change*, 175, 121359. <https://doi.org/10.1016/j.techfore.2021.121359>
- Marsen, S. (2020). Navigating crisis: The role of communication in organizational crisis. *International Journal of Business Communication*, 57(2), 163–175. <https://doi.org/10.1177/2329488419882981>
- Mazzei, A., & Butera, A. (2021). Internal crisis communication. In *Current trends and issues in internal communication: Theory and practice* (pp. 165–181).
- Mazzei, A., & Ravazzani, S. (2011). Manager-employee communication during a crisis: The missing link. *Corporate Communications: An International Journal*, 16(3), 243–254. <https://doi.org/10.1108/13563281111156899>

Mazzei, A., Ravazzani, S., Fisichella, C., Butera, A., & Quarantino, L. (2022). Internal crisis communication strategies: Contingency factors determining an accommodative approach. *Public Relations Review*, 48(4), 102212. <https://doi.org/10.1016/j.pubrev.2022.102212>

Men, L. R. (2014). Strategic internal communication: Transformational leadership, communication channels, and employee satisfaction. *Management Communication Quarterly*, 28(2), 264–284. <https://doi.org/10.1177/0893318914524536>

Mergel, I. (2021). Study on the impact of digital transformation on democracy and good governance. *Council of Europe*. https://doi.org/10.1007/978-3-030-53697-8_19

Mo, P. K. -H., Luo, S., Wang, S., Zhao, J., Zhang, G., Li, L., Li, L., Xie, L., & Lau, J. T. F. (2021). Intention to receive the COVID-19 vaccination in China: Application of the diffusion of innovations theory and the moderating role of openness to experience. *Vaccines*, 9(2), 129. <https://doi.org/10.3390/vaccines9020129>

Mohee, A. (2022). A realistic analysis of the Stuxnet cyber-attack. *APSAnet Preprints*. <https://doi.org/10.33774/apsa-2022-qs797>

Moturu, S. T., & Liu, H. (2011). Quantifying the trustworthiness of social media content. *Distributed and Parallel Databases*, 29, 239–260. <https://doi.org/10.1007/s10619-011-7088-2>

Nkongolo, M. (2023). Navigating the complex nexus: Cybersecurity in political landscapes. *arXiv Preprint*, arXiv:2308.08005. <https://arxiv.org/abs/2308.08005>

Nobles, C. (2018). Botching human factors in cybersecurity in business organizations. *HOLISTICA: Journal of Business and Public Administration*, 9(3), 71–88.

Palleti, V. R., Adepu, S., Mishra, V. K., & Mathur, A. (2021). Cascading effects of cyber-attacks on interconnected critical infrastructure. *Cybersecurity*, 4(1), 1–19. <https://doi.org/10.1186/s42400-021-00071-z>

- Petrenko, S. (2022). *Cyber security innovation for the digital economy: A case study of the Russian Federation* (e-book ed.). CRC Press. <https://doi.org/10.1201/9781003337782>
- Ravazzani, S., & Mazzei, A. (2022). Social media and the role of internal communication for crisis prevention and management. In Y. Jin & L. L. Austin (Eds.), *Social media and crisis communication* (pp. 215–226). Routledge. <https://doi.org/10.4324/9781003043409>
- Ravell, H. (2023). Rowling, Potterheads, and why “The best way to manage a crisis is to prevent one”: Twitter communication analysis. *Communication Research and Practice*, 1–18.
- Ribeiro-Navarrete, S., Botella-Carrubi, D., Palacios-Marqués, D., & Ore-ro-Blat, M. (2021). The effect of digitalization on business performance: An applied study of KIBS. *Journal of Business Research*, 126, 319–326. <https://doi.org/10.1016/j.jbusres.2021.01.037>
- Rogers, E. M. (1962). *Diffusion of Innovations*. (1st ed.). Free Press.
- Rogers, E. M. (2003). *Diffusion of Innovations*. (5th ed.). Free Press.
- Ross Arguedas, A., Badrinathan, S., Mont’Alverne, C., Toff, B., Fletcher, R., & Nielsen, R. (2022, April 4). *Snap judgments: How audiences who lack trust in news navigate information on digital platforms*. Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/snap-judgments-how-audiences-who-lack-trust-news-navigate-information-digital-platforms>
- Sadri, A. M., Ukkusuri, S. V., & Ahmed, M. A. (2021). Review of social influence in crisis communications and evacuation decision-making. *Transportation Research Interdisciplinary Perspectives*, 9, 100325. <https://doi.org/10.1016/j.trip.2021.100325>
- Savaş, S., & Karataş, S. (2022). Cyber governance studies in ensuring cybersecurity: An overview of cybersecurity governance. *International Cybersecurity Law Review*, 3(1), 7–34. <https://doi.org/10.1365/s43439-021-00045-4>

- Schünemann, W. J. (2022). A threat to democracies? An overview of theoretical approaches and empirical measurements for studying the effects of disinformation. In *Cyber Security Politics* (pp. 32–47). <https://doi.org/10.4324/9781003110224-4>
- Shin, I., & Hur, W. M. (2020). How are service employees' perceptions of corporate social responsibility related to their performance? Prosocial motivation and emotional labor as underlying mechanisms. *Corporate Social Responsibility and Environmental Management*, 27(6), 2867–2878.
- Shipunova, O., Berezovskaya, I., Pozdeeva, E., Evseeva, L., & Barlybayeva, S. (2022, April). Digital Trust Indicators in Human–Computer Interaction. In *World Conference on Information Systems and Technologies* (pp. 245–254). Springer International Publishing.
- Singh, B., Choudhary, R., Sharma, A., & Sharma, P. (2022). Artificial intelligence as a human mind tool saving from cybercrimes and cyber threats. *International Journal of Medical Toxicology and Legal Medicine*, 25(3–4), 37–46. <https://doi.org/10.5958/0974-4614.2022.00047.X>
- Strandberg, J. M., & Vigso, O. (2016). Internal crisis communication: An employee perspective on narrative, culture, and sense-making. *Corporate Communication: An International Journal*, 21(1), 89–102.
- Suherlan, S. (2023). Digital technology transformation in enhancing public participation in democratic processes. *Technology and Society Perspectives (TACIT)*, 1(1), 10–17. <https://doi.org/10.61100/tacit.v1i1.34>
- Taylor, M., & Perry, D. C. (2005). Diffusion of traditional and new media tactics in crisis communication. *Public Relations Review*. <https://doi.org/10.1016/j.pubrev.2005.02.018>
- Ulmer, R. R., Sellnow, T. L., & Seeger, M. W. (2022). *Effective crisis communication: Moving from crisis to opportunity*. Sage Publications.
- van de Poel, I. (2020). Core values and value conflicts in cybersecurity: Beyond privacy versus security. In *The Ethics of Cybersecurity* (pp. 45–71). https://doi.org/10.1007/978-3-030-29053-5_3

Vargo, S. L., Akaka, M. A., & Wieland, H. (2020). Rethinking the process of diffusion in innovation: A service–ecosystems and institutional perspective. *Journal of Business Research*, 116, 526–534. <https://doi.org/10.1016/j.jbusres.2020.01.038>

Wang, Z., Sun, L., & Zhu, H. (2020). Defining social engineering in cybersecurity. *IEEE Access*, 8, 85094–85115.

Whyte, C., Thrall, A. T., & Mazanec, B. M. (2020). *Information warfare in the age of cyber conflict*. Routledge.

Yue, C. A., Men, L. R., & Ferguson, M. A. (2021). Examining the effects of internal communication and emotional culture on employees' organizational identification. *International Journal of Business Communication*, 58(2), 169–195. <https://doi.org/10.1177/2329488420914066>

Part IV

Art and New Media

The Mechanical Turkness: Tactical Media Art and the Critique of Corporate AI

Dejan Grba

Introduction

Since its largely obscure beginnings in the 1970s, AI art has expanded, gained visibility, and attained sociocultural relevance since the second half of the 2010s (Audry, 2021). It has benefited from the accelerating affordance of multilayered subsymbolic machine learning (ML) architectures such as Deep Learning, AI's increasing sociopolitical impact, and the art market's integration with the crypto economy. Contemporary AI art includes diverse creative pathways informed by functions, applications, and consequences of modern ML systems (Cetinić and She, 2022; Grba, 2022a, 3-17).

The rapid industrialization of AI drives the massive concentration of wealth, power, and (often controversial) sociopolitical influence, which defines a space for the critical discourse that traces and questions the constitution of social reality through the technical logic of

Dejan Grba, Associate Professor, Interdisciplinary Graduate Center, Digital Art Program, University of the Arts in Belgrade, Belgrade ORCID: 0000-0002-5154-9699, email: dejan.grba@gmail.com

DOI: <https://doi.org/10.5281/zenodo.14205010>

AI. Artists play a significant part in that engagement. They deconstruct, reassess, reappropriate, and sometimes democratize AI technologies to uncover or underline AI development's epistemic and existential issues and question the economic and sociopolitical consequences of AI capitalism. Their production continues the flux of heterogeneous tactical media practices that have energized art and culture since the late 20th century with hybrid forms of academic or critical interventions into the neoliberal condition's technological, political, economic, and cultural layers. With the increasing accessibility of technologies that can be modified and repurposed by the actors who operate outside of the established hierarchies of power and knowledge, tactical media art has emerged as a response to a shift in postindustrial society towards the information economy in which efficiency, operationalism, and instrumental rationality become core values, and market transactions the predominant social good.

Although it often maps top-down power relations, tactical media art embodies a sense of bottom-up resistance in a manner and style associated with cultural dissent and opposition. It challenges the dominant semiotic regime through signs, messages, and narratives that foster critical thinking and offer new ways of seeing, understanding, and (in some scenarios) interacting with the systems of power. Tactical media works are not sweeping revolutionary events but micro-political acts of disruption, intervention, and education for which absolute victory or fundamental transformation is neither a desirable nor truly attainable objective. They usually generate fleeting, ephemeral, and pliable statements or initiatives that must be continually reconfigured in response to their changing targets. Their effects are often not immediate but cumulative and relational because they provide

insights, knowledge, and tools that may become transformative in the hands of the audience (Raley, 2009).

In different ways and contexts, tactical media art contends with business or state strategies based on quantization, statistical reductionism, datafication, behavioral prediction, and the manipulation of decision-making (Grba, 2020, 71-73). AI technologies increase these strategies in scope, extent, and sophistication, often amplifying their undesirable consequences. Hence, a notable portion of the tactical art's topical range involves biases, prejudices, inequities, and political agendas in the AI architectures and relates to the AI industry's exploitative economic practices. Artists criticize the aggregation of transnational labor for training and testing ML models, the systematic concealment of the extent, value, and implications of human creative endeavor in AI development, and the immobilization of the AI workforce through manipulative interfaces, complacency, lack of protection, and precarity. Their expressive methodologies do not always follow overtly activist agendas but offer subtle, sometimes covert or intentionally ambivalent, critical narratives. They occasionally combine provocation with humor by recontextualizing AI systems' objectives or operative modes for ironic revelatory effects. Successful works usually let the audience actively identify the corporate AI's interests, animosities, and injustices. They emphasize presence, engagement, and response but sometimes constrain them to raise the audience's awareness of limited actionability or immediate political impact. The tactical AI art's critical value is proportional to the artists' understanding of the political and moral nuances dispersed across the abstract, technically convoluted, and functionally opaque realities of AI technologies and their applications. This makes tactical AI

art conducive to understanding how AI simultaneously reflects and influences sociopolitical relations, economies, and worldviews.

The existing literature pertinent to tactical AI art includes Marcus and Davis' *Rebooting AI* (2019), Mitchell's *Artificial Intelligence* (2019), and Larson's *The Myth of Artificial Intelligence* (2021), which pinpoint the conceptual and technological issues of AI research and the contradictions of AI implementation. Pasquinelli's *How a Machine Learns and Fails* (2019) and Kearns and Roth's *The Ethical Algorithm* (2019) address the ethical and sociopolitical consequences of AI's technical imperfections and biases. Żylińska's *AI Art* (2020) critiques AI's impact on art and culture. Crawford's *Atlas of AI* (2021) maps the exploitative layers of AI capitalism hidden behind marketing, media hype, and cultural commodification. In *Tactical Entanglements* (2021), Zeilinger provides a theoretical analysis of selected artworks' tactical values, issues, and potentials. In *Deep Else* and *Lures of Engagement* (both 2022) and *Faux Semblants* and *Renegade X* (both 2023), I critique the creative, expressive, political, and ethical features of AI art.

My discussion in this chapter centers around art practices that thematize creative agency, crowdsourced labor, and delegated artmaking to reveal the social rootage of AI technologies, underline the productive human roles behind the performative power of applied AI, and expose conflicted social relations in the corporate AI sector. I focus on works whose poetics indicate the broader issues of contemporary AI-influenced science, economy, and society. Lined out in three subject areas—The Elusive Artist, Digital Labor Transparency, and Outsourcing Artistic Work—these exemplars are interlinked across their topical domains and deeper art-historical contexts. I examine these works' conceptual, methodological, expressive, and ethical aspects and assess their effectiveness in impacting AI's phenomenological,

epistemic, and political realities. This allows me to identify several problems that affect their tactical cogency and to outline some of the prospects for addressing the challenges and advancing the field.

The Human Shades of AI

The increased enterprise and state application of AI has a significant influence on culture, economy, and politics. Many problematic aspects of that influence are seductive and hard to assess because of strong anthropomorphic¹ concepts in the foundations of AI science, the AI's convoluted technical logic, and the IT industry's systemic illusionism of autonomous AI (Larson, 2021; Natale, 2021). Their impacts on artists range from provoking critical scrutiny, through inciting confusion, to inducing complacency.

The Elusive Artist

The specter of misidentified creative autonomy has been haunting computational art since its outset as the pervasiveness and efficacy of computing technologies can nudge artists toward conflating their efforts or identities with the highly formalized technical emulations of human faculties. Pioneering AI artist Harold Cohen had an

¹ Anthropomorphism is an innate psychological tendency to assign human cognitive traits, emotions, intentions, or behavioral features to non-human entities or phenomena (Hutson, 2012).

ambiguous relationship with the machinic creative agency throughout his lifelong project called AARON (1971–2016)—a robotic system tasked to “draw and paint autonomously” and “embody creative behavior and the conjuring of meaning” (Cohen, 1995). Whether sincerely or for promotional purposes, Cohen flirted with mystifying rhetoric about the robot’s “surprises” and “creative serendipity” (Garcia, 2016), which converged with his pioneering role into a strong tributary to the legacy of bio-detached anthropomorphism in computational art.

The emotional charge of contemporary AI artists’ claims that “there is something deeply thrilling about observing a machine learn, starting from scratch and iteratively discovering something about its world” (Audry, 2021, 85) indicates a strange fascination with complex statistical computation situated within strictly defined expressive spaces and signals an inclination to elevate constrained modes of functional autonomy into meaningful cognitive processes. Rather than fundamentally approaching their AI applications as tools, artists frequently represent them as “autonomous creators”, “creative collaborators”, “partners”, or “companions” (Audry, 2021, 27–28, 241–243).²

Some artists banalize this tendency through saccharine reiterations of Cohen’s emphasis on robotic creative agency and misleading claims that they “teach their robots how to draw or paint” instead

2 Mainstream AI-derivative art is replete with verbal anthropomorphizations of ML systems. For instance, Holly Herndon calls her vocal deepfake generator Holly+ her “digital twin” and pronounces it as “her” (Herndon, 2022). In a statement about her and Jlin’s work *Godmother* (2018), Herndon calls the software they used the “nascent machine intelligence” that “listened to her godmother [Jlin]” and reinterpreted her (Jlin’s) art in the voice of “her mother [Herndon]” (Audry, 2021, 213).

of using dryer but more accurate language about constructing and programming robotic systems to produce drawings and paintings. Examples include Pindar Van Arman's *Painting Robots* (since 2006) (Van Arman, 2016), Shantell Martin and Sarah Schwettmann's *Mind the Machine* (2017) (Schwettmann, 2017), Joanne Hastie's *Abstractions (Tech Art Paintings)* (since 2017) (Hastie, 2021), and Aidan Meller, Lucy Seal, and collaborators' *Ai-Da Robot Artist* (since 2019) (Meller et al., 2019) which epitomizes the anthropomorphic kitsch of AI. Driven by weekend artists' enthusiasm, these practices symbiose the happy-go-lucky joy in technocentric creativity with dilettante negligence toward the conceptual and aesthetic evolution of visual and media arts since the late 19th century.³

Attractive to the audience and the media, the simplistic or superficial takes on machinic creativity, authorship, and originality can be found in all areas of AI art. For instance, some of the performative AI practices that enjoy the sponsorship of big tech companies aid their corporate PR strategies by representing AI as a pantheon of powerful but friendly anthropomorphic deities. They imply the notions of machinic creative spontaneity in human-robot interactions that exploit the evolved human capacity for, and bias toward, detecting agency in mid-sized objects moving at medium speeds. These practices promote a robotically enhanced consumerist lifestyle or muse about the existentially intense but politically or ethically vague notions of human-AI symbiosis. They are also sleekly sanitized and anesthetized mutations of earlier avantgarde practices (Grba, 2023a). Examples include Huang Yi's choreography *HUANG YI and KUKA* (since 2015) (Yi, 2021), Nigel John Stanford's musical performance *Automatica*:

3 See, for example, Arnason and Mansfield (2012) and Hopkins (2000)

Robots vs. Music (2017) (Stanford, 2017), Sougwen Chung's projects such as *Drawing Operations Unit: Generation 2* (2017) (Chung, 2020), and visceral homo-robotic interactions such as Marco Donnarumma and Margherita Pevere's *Eingeweide* (2018) (Donnarumma, 2023).

The audience's romantically skewed perception of, and virtue signaling about, AI art reinforces the anthropomorphization of AI in popular discourse. Georgina Ruff (2022) exemplified this by dissecting the disparities between the "emphatic" verbal exchanges about Sun Yuan and Peng Yu's robot installation *Can't Help Myself* (2016), its actual production background, and exhibition history. She identified and analyzed heavy anthropomorphization in both the curator's comments and in narratives communally constructed, circulated, and reverberated on social media by the audience who largely never experienced the work in person. Such public feedback in turn encourages artists to relegate their creative agency to ML architectures. Repeating with each increase in AI technologies' precision or scope, the latest iteration of this hype cycle is the proliferation of prompt-based generated images, videos, and animations with large Text-to-Image diffusion models (TTIs) such as DALL-E 2, MidJourney, Stable Diffusion, Disco Diffusion, and Pytti (McCormack et al., 2023).

The anthropomorphization of AI also allows artists to employ critical thinking and contextual awareness to address the questionable intersections of AI and creativity. In Adam Basanta's installation *All We'd Ever Need Is One Another* (2018), custom software randomizes the settings of two mutually facing flatbed scanners so that in every scanning cycle, each captures a slightly altered mix of the facing scanner's light and its unfocused scanning light reflected off the opposite scanner's glass plate. The perceptual hashing algorithms use

parameters such as aspect ratio, composition, shape, and color distribution to compare each scan to a database of images and image metadata scraped from freely accessible online artwork repositories. When the comparison value between the scan and the most similar database image exceeds 83 percent, the software declares a “match”, selects the scan for printing, and labels it according to the database image metadata (Basanta, 2018). After it printed one of the scans and labeled it *85.81%_match: Amel Chamandy ‘Your World Without Paper’, 2009*, Canadian artist Amel Chamandy initiated a legal action about the intellectual property rights against Basanta because of the reference to her photograph. However, *85.81%_match...* is not for sale and Basanta does not use it for direct commercial gains by any other means. In *All We’d Ever Need...*, he legitimately and consistently applied the functional logic of ML to leverage the open-endedness of creative work and disturb the entrenched notions of agency, authorship, originality, and intellectual property. The argumentation intricacies in Chamandy’s lawsuit reveal the cognitive and ethical issues of our tendency to crystallize the commercial rights of creativity (Zeilinger, 2021, 94–108).

Basanta and other notable tactical artists criticize AI as a sociopolitical complex and as technology. They expose the misalignment between algorithmic decision-making processes and anthropomorphic metaphors that represent them and question the responsibility of using such metaphors in the AI’s underlying epistemologies (Curry, 2023, 181). Recognizing the variable abstraction of technologically entangled authorship, they show that crucial aesthetic factors such as decision-making, assessment, and selection are human-driven and socially embedded regardless of the complexity or counterintuitiveness of our expressive tools. By emphasizing that artworks emerge

and live in ceaseless interrelations with other creative acts and agencies (Zeilinger, 2021, 104), they reaffirm the notion of art as a dynamic, evolving, biologically and sociopolitically contextualized relation that needs continuous cultivation. Although these properties make art hard to objectify, commercialize, and own, it remains instrumentalizable in myriad ways.

Digital Labor Transparency

AI industry giants, such as Google, Amazon, Microsoft, IBM, Apple, and Meta in the USA, and Baidu, Alibaba, Tencent, and Huawei in China, constitute the core of the corporate AI sector which also integrates numerous smaller enterprises, startups, academic projects, and independent researchers. The politics of that sector unfolds in defining and controlling research, development, production, and application of AI technologies. Both AI-producing and AI-powered companies deploy cybernetic frameworks to maximize profit, reinforce their power structures, and “optimize” labor by conditioning its workforces economically and politically through manipulative interfaces, complacency, or lack of protection. Continuing the capitalist paradigm of labor transparency, the AI industry simultaneously strives to obscure the economic interests behind its commodities and render its workforce invisible in professional and public discourse. Main strategies for achieving that are based on concealing or diminishing the extent, value, and implications of human endeavor in AI development and foregrounding the notions of autonomous AI.

However, like Wolfgang von Kempelen's chess-playing animatronic hoax the Turk,⁴ enterprise AI applications are simulacra of machinic autonomy substantiated by a heterogenous corpus of computer scientists, software engineers, online workers, and everyday users. On-line microlabor crowdsourcing platforms, such as Amazon's MTurk,⁵ are arguably the most notorious manifestations of the corporate AI's "human-in-the-loop" complex. They have been extensively used as largely unregulated instruments for amassing transnational echelons of workers to perform repetitive tasks essential for building, training, and testing ML architectures. While online microlabor provides workers some new opportunities to renegotiate power relations, it was initiated by and for the convenience of techno-capitalist elite to push labor extraction, division, alienation, and precarity to new extremes, thus entrenching or even exacerbating the established power differentials (Golumbia, 2015; Hill, 2016). The AI community adopted Amazon's cynical nomenclature to euphemistically call

-
- 4 Wolfgang von Kempelen constructed the Automaton Chess Player, widely known as the Mechanical Turk or the Turk in 1770. For 84 years, it was exhibited throughout Europe and the Americas as a fully automatic chess-playing machine, attracting many spectators and notable challengers, such as Napoleon Bonaparte, Charles Babbage, Benjamin Franklin, and Edgar Allan Poe (Carroll, 1975). The Turk was designed as a torso of an Ottoman sultan fronting the chessboard mounted on top of a cabinet with an elaborate internal mechanism. However, the mechanism's role was not to analyze the chess table positions and compute the moves but to conceal the expert human chess player and enable him to move the sultan's figures on board (Jay, 2001).
 - 5 Besides MTurk, the online microlabor platforms include Fiverr, Microworkers, Clickworker, Upwork, TaskRabbit, WorkMarket, Catalant Technologies, Inc., Toloka, and others.

the online-aggregated human labor “pseudo-AI” or “Artificial Artificial Intelligence” (AAI).

Artists address this foundational “artificiality” in different approaches that reveal corporate AI’s sociopolitical conflicts and expose the productive human roles in AI technologies. As a prescient critique of the 1990s hype about the “autonomy” of interactive digital multimedia, Mona Hatoum’s three-day performance *Pull* (1995) provides a valuable contextual backdrop for the “human in the loop” complex. *Pull*’s visitors were invited to pull the braid of human hair hanging in a wall niche and witness in real-time Hatoum’s facial and vocal reactions on a video monitor directly above it. Fascinated by the speed and variability of mediated interaction, visitors largely perceived the work as an advanced digital installation, whereas it was based on a von Kempelen-style deception. The braid was not part of a computer interface but of Hatoum who laid hidden in a chamber behind the wall and faced the live video camera/microphone setup (Artbasher, 2007). Matt Richardson’s *Descriptive Camera* (2012) belongs to a corpus of works that address the sociotechnical blindness of the crowdsourced AI economy. *Descriptive Camera* has a conventional digital camera lens but no display or memory; instead, it sends the captured image to an Amazon MTurker tasked to write down and upload its brief description and the device prints it out (Richardson, 2012). This “picture development” process offers a revelatory counterintuitive glimpse into the large volume of image classification and annotation tasks required for building the ML training datasets, which underpaid workers perform on online labor aggregators.

The corporate AI’s sociotechnical blindness pertains less to the lack of perceptiveness for the centrality and values of human work than

to an active evasion of acknowledging them. RyBN and Marie Lechner's project *Human Computers* (2016–2019) targets this evasiveness by documenting the historical use of human beings in large computational architectures (RyBN, 2021). It features an online media archaeology of human labor in computation since the 18th century, from von Kempelen's Turk to recent enterprises such as SETI @ home, Amazon's MTurk, ReCAPTCHA, Scale, etc. In 2018, the project included a chess game *AAI Chess*, which offered three all-human playing modes: human vs. human, human vs. MTurker, and MTurker vs. MTurker. Two years later, Jeff Thompson used the same principle in his *Human Computers* (2020). It tasked the visitors to manually resolve a digital image file (a Google StreetView screenshot of the gallery) from its binary form into a grid of pixels. With 67 calculations per pixel, the complete human-powered image assembly takes approximately eight hours (Thompson, 2020). Here, the visitors' enactment of automated operations illustrates how a combination of complexity and speed in pervasive technologies makes them hard to understand and manage.

In *Crowd-Sourced Intelligence Agency (CSIA)* (since 2015), Derek Curry and Jennifer Gradecki address the related issues of accuracy and accountability in data processing. *CSIA* offers a vivid educational journey through problems, assumptions, or oversights inherent with ML-powered dataveillance practices. It centers around an online app that partially replicates an OSINT system and allows the visitors to assume the role of data security analysts by monitoring and evaluating their friends' Twitter messages or by using an automated Bayesian classifier to test the "delicacy" of their messages before posting (Gradecki and Curry, 2017). This relational architecture fosters active learning by combining a library of resources about the intelligence sector's analytic and decision-making practices with the playfully

transgressive message “policing” that allows a direct experience of how the inaccurate judgment of decontextualized metadata affects ML systems used by the surveillance agencies to enhance their processing speed and volume.

The aggregation of human input is crucial for the initial configuration of ML models and for their refinement with real-world feedback data where content recommendation algorithms play an interesting part. In most cases, they retain the online platform users and influence their behavior for monetary or political gain by predicting the content a user will engage with based on their previous interactions. Beyond being manipulative by design, predictive algorithms remain notoriously mechanistic and solipsistic because they are tasked to shape complex human behavior based on relatively simplistic and decontextualized online data. Lacking commonsense reasoning capabilities, they also tend to be abusive, so if a user selects biased or disinformational content, the following recommendations may reinforce the slanted or false narrative. Tomo Kihara’s *TheirTube* (2020) warns about the implications of such flaws. It is an online filter bubble simulator that generates YouTube landing pages with recommended videos for six fictional profiles (Liberal, Conservative, Conspiracist, Climate Denier, Fruitarian, and Prepper) based on a simulation of their previous YouTube viewing data (Kihara, 2020).

With *Randomized Living* (2015–2017), Max Hawkins undertook a riskier and perhaps more responsible interrogation of the applied AI’s consequences. In this two-year experiment, Hawkins organized his life according to the dictate of recommendation algorithms. He designed several apps that continuously harvested his online data to suggest a city where he would live for about a month, places to go, people to

meet, and things to do in that city (Hawkins, 2021). *Randomized Living* is a striking exemplar of cybernetic existentialism—the art of conceiving a responsive and evolving cybernetic system to express profound existential concerns. It warns about the humans’ general susceptibility to constraining behavior and cognition to fit machinic protocols, such as labor regimes in industrial capitalism, and to the perceptual and expressive conditioning in interactions with computationally mediated reality (Pasquinelli, 2019, 17).

These works reveal technological reductivism that shifts social relations through the mutually reinforcing network effects and the online users’ opportunism or complacency. It drives the pathological business logic of many IT providers, which dehumanizes users and turns them into slavish data-generating commodities by building an addiction to “free” online services whose recommendation algorithms favor negatively biased, politically derisive, and socially toxic views. It is worth noting that such deviations get compensated over time by the emergent flexibility of human intelligence that drives cultural maturation. Nevertheless, while AI’s social issues may be considered the transient side-effects of the coevolution between culture and technology, they are damaging, and it is crucial to remain cognizant of and vigilant about them because of the inherent asymmetry in production means and political power between institutions and individuals.

| *Outsourcing Artistic Work*

Artmaking is contextualized by the sociocultural environment and dependent on the available knowledge, material, and organizational

resources, so delegating some of its components is inevitable. The outsourcing of artistic labor has a long history reaching back to antiquity. In Western culture, it is emblemized by Renaissance artists' studios that employed technicians, apprentices, and disciples to produce commissioned works with various degrees of the master artist's involvement (Wallace, 2014). Traditionally, outsourced creative labor remained concealed in the finished artwork but avantgarde modernist artists rendered it explicit by developing techniques such as collage, assemblage, and bricolage to experiment with the nature of creativity, question the cultural notions of art and challenge the parameters of aesthetic judgment. Pablo Picasso, Georges Braque, Hannah Höch, Kurt Schwitters, and others legitimized creative outsourcing in their practices, and Marcel Duchamp radicalized it with readymades, such as the venerable *Fountain* (1917) (Molderings, 2010). Following the proliferation of artistic outsourcing strategies throughout the 20th century, the Internet facilitated their further expansion by coupling accessibility and programmability.

For instance, Perry Bard brilliantly engaged the relational flexibility of human visual interpretation in her online project *Man with a Movie Camera: The Global Remake* (2007–2014). It invited visitors to select any shot from Dziga Vertov's seminal film *Man with a Movie Camera* (1929) and upload their video interpretations (Bard, 2014). Bard's server-side software collected the participants' shots and re-played a two-channel setup synchronizing Vertov's original on the left with a continuously reassembled remake on the right.⁶ By leveraging the breadth of human perceptive cognition, this relatively

⁶ Multiple interpretations of the original shot were randomly selected for the replay.

simple technical arrangement engrosses both the uploaders and viewers in an intriguing experience of surprise motivated by curiosity rather than monetary profit. Bard used her webpage and software to transparently define, solicit, and process the visitors' voluntary participation based on a common-exploratory goal.

Since the launch of online microlabor platforms in 2005, artists have used them in diverse creative approaches. For example, Clement Valla ran several generative experiments through Amazon's MTurk. In *Sol LeWitt + Mechanical Turk* (2009), Valla's software recreated Sol LeWitt's algorithmic drawings, posted their instructions for MTurkers to execute online (5 US cents per drawing), and assembled their interpretations into a grid. In *A Sequence of Lines Traced by Five Hundred Individuals* and *A Sequence of Circles Traced by Five Hundred Individuals* (both 2011), Valla utilized the entropic effects of iterative tracing, and in *Seed Drawing* (2011), he used the evolutionary effects of iterative copying to aggregate large-scale structures of organic patterns (Valla, 2023).⁷ Artists have also experimented with tactical arrangements of delegated creativity on microlabor platforms to highlight and question their socioeconomic issues (Bishop, 2012; Torok, 2020). For instance, xtine burrough has examined the affective politics of the microlabor economy and the human values of online workforce in participatory projects that subvert the MTurk's logic, such as *A Penny for Your Thoughts* (2009) or *Mediations on Digital Labor* (2015) (burrough, 2020). Other examples include Andy Baio's *The Faces of Mechanical Turk* (2008) (Baio, 2008), Nick Thurston's *Of the Subcontract: Or Principles*

7 Mainstream art soon assimilated Valla's approach. For instance, Agnieszka Kurant and John Menick's *Production Line* (2017) was a direct methodological clone of Valla's *Sequence* works, but resulted in a series of static forms (Arts at MIT, 2022).

of *Poetic Right* (2013) (Thurston, 2013), Guido Segni's *The Middle Finger Response* (2013) (Segni, 2013), and Kieran Browne and Ben Swift's *Enacting Alternative Economies* (2020) (Browne and Swift, 2020).

While some artworks realized through microlabor platforms are non-commercial, in most of them artists are task providers, which can lead to ethically questionable situations, particularly regarding the online workers' wages compared to the artists' profits yielded by selling the aggregately produced works (Żylińska, 2020, 117–120). The central part of this issue is that the microlabor employers (task providers) define the conditions of task fulfillment and control the details about the purpose and broader context of the tasks they offer to employees. It came to the fore with Aaron Koblin's non-tactical projects *The Sheep Market* (2006), *Ten Thousand Cents* (2007–2008), and *Bicycle Built for Two Thousand* (2009, with Daniel Massey) (Koblin, 2015). Although they were conceptualized as sound generative experiments, Koblin drew critique for exploiting workers through compensatory allocation disparities (Elish, 2010; LABoral, 2011; Berdugo and Martinez, 2020, 89). For instance, in *The Sheep Market*, MTurkers were tasked to “draw a sheep facing to the left” for 2 US cents per drawing. They generated a collection of 10,000 sheep drawings that Koblin sold on the project website as a limited edition of adhesive stamp blocks (20 drawings/stamps per block) for 20 USD each.

The sociocultural disparities between artists and MTurkers are not abstracted away in monetary transactions, which can cause subtler ethical problems. In the tactically motivated *Social Turkers* (2013), Lauren Lee McCarthy solicited the MTurkers' real-time feedback over a mobile phone video link to improve her social interactions during twenty in-person dates she arranged on OkCupid (McCarthy, 2013). McCarthy

was praised for positioning herself as “less than” the MTurkers by seeking their advice rather than celebrating the diversity and humanity of their responses in the face of mechanization (Berdugo and Martinez, 2020, 89). However, the notion of McCarthy’s “inferiority” to the MTurkers can be interpreted as condescending because it accepts and, to some degree, conceptually exploits, even if ironically, the corporate AI’s implicit categorization of MTurkers as the abject (socio-economically “low”) others. Furthermore, the artist did not seek the MTurkers’ help as a friendly gesture (for free) but for monetary compensation, the same way the AI industry extracts their labor. In that sense, McCarthy’s method for taking social and behavioral directives from the AI complex was less responsible compared to Max Hawkins’ *Randomized Living* (2015–2017) which involved no monetary transaction. This also makes *Social Turkers* worth comparing with Stanley Brouwn’s actions *This Way Brouwn* (1961), in which the artist asked random passersby to give him verbal street directions and draw a navigation route to a chosen location—all voluntary exchanges in an interaction between contextually more equal actors (von Graevenitz, 1977).

A subset of tactical practices explores the AI’s impact on defining, performing, and valuing (art)work by directly applying ML for art production. For instance, in *Demand Full Laziness: A Five-Year Plan* (2018–2023), Guido Segni delegated part of his artistic production to a set of GANs that make images by recording periods of his “unproductivity” such as sleeping, reading, or lazing around (Segni, 2018). Images are distributed on the Patreon platform to sponsors who fund the artist. This critique of the trend of automating artistic output integrates the contemporary reality of extreme labor precarity with the popular notions of AI-influenced cultural status of artmaking, the routinized aesthetics, and the technological opportunism of (computational)

art. Similarly, Sašo Sedlaček's *Oblomo* (2019–2020) contextualized the visitors' non-work with the crypto-economy to satirize the NFT-driven art tokenization frenzy. It featured a blockchain web platform with custom cryptocurrency, which rewards its visitors for being still: when an ML rig detects their non-activity, it awards them *Oblomo* coins minted from an Om block foundation. Visitors can spend these coins at the platform's market to buy services and goods offered by other visitors (Sedlaček, 2020).

As Browne and Swift noted (2020, 2), computational mediation of work and creativity introduces specific concerns. It is fueled by techno-deterministic views that ignore the social contexts in which technologies are developed and applied and naturalize the technologically enforced power relations. Since tactical media artists operate directly within the technological systems whose ideological regimes they aim to critique, techno-determinism also affects the poetic, social, political, and ethical values of their work.

Challenges

In various ways, exemplars discussed in this chapter affirm that the AI's "Turk" is stubbornly not mechanical enough, it resists "emancipation", and it is not easy to make it autonomous because humans are inherently technological animals and the sine qua non of technology. By facilitating the resurgence of humanity from its typically opaque habitats of computer science and industry, these artworks also assert that Silicon Valley's regressively infantile dreams of ultimate techno-autonomy are being forcefully enacted in the real world with possibly nightmarish sociopolitical consequences.

At the same time, these artworks' self-critique, whether intended or not, indicates that their sociotechnical milieu simultaneously provides tactical affordances and threatens to relativize, neutralize, or instrumentalize them. Closely related to AI research, development, and application, the poetic features and implications of AI art are affected by the epistemic, conceptual, discursive, and sociopolitical problems of AI science and industry. In addition to these problems, tactical AI art extends and often amplifies the tactical media art's inherent issues.

The historical tendency in computational art, initially imposed by high cognitive demands of computer technology, tacitly conflates artmaking with the skillful handling of creative devices, which often incentivizes artists towards technical virtuosity devoid of self-critical distance or playful irreverence (Taylor, 2014; Żylińska, 2020, 75–85). This techno-fetishist mentality often reinforces a naïve lack of understanding that the poetic role of production techniques in the arts is fundamentally defined by conceptual thinking and meaningful contextualization. Thus, bio-detached and socially unaware notions of agency permeate AI art production and its popular representation through confused or openly mystifying rhetoric. They promote a pseudo-romantic quest for human-flavored creative “essence” within ML systems instead of demystifying them as sophisticated tools for statistical analysis, which have little to do with creativity per se and are more constructively understood in the non-technical perspective as sociopolitical apparatuses (Forbes, 2020, 6). These notions limit the perceptiveness for bold and nuanced artistic examinations of intelligence, creativity, expression, authorship, labor, ownership, authenticity, accuracy, and bias.

Various obvious but often undisclosed conceptual, methodological, thematic, aesthetic, and presentational similarities between different artists' works underscore the concerns about their creators' creative literacy and professional ethics (Grba, 2023a). Sincerely-motivated, pertinent, and well-conceived ideas are sometimes rendered as dry, unengaging, ineffective, or counter-effective works (Grba, 2022a, 20–21) or represented through overly didactic setups and inflated theoretical rhetoric (Quaranta, 2020; Grba, 2021, 246–247). Tactical artworks are mostly consumed and gain significance within academic communities prone to developing echo chambers, while their real-world critical cogency, impact, and viability require unbiased assessment. The academization of tactical values (or purposes) may be scholarly fruitful, but the general audience, which is (supposed to be) central to tactical art, can easily recognize it as aloofness or cynicism, which leads to indifference, distrust, or resentment.

Tactical media art's sociotechnical entanglement is problematic in several aspects. By openly identifying loopholes and weaknesses in the institutions they criticize, artists set their achievements up for exploitation and recuperation (Lovink and Rossiter, 2005). Sometimes, tactical critique becomes the reflection of its target(s) (Raley, 2009, 30) or inadvertently lapses into the mystification of technocracy whereby a class of tech-savvy artisans acts on behalf of "the (lay) people" by articulating a vision of individual freedom realizable from within the power structures of the information society (Barbrook and Cameron, 1995/2008). In the context of AI, it is worth asking if an activist action should end up being used by the AI industry to improve its profitable instruments or remedy its public image without necessarily improving technical and ethical standards (Powles, 2018). Could tactical art disrupt the corporate AI regime with lasting and more

desirable social consequences, and how effectively can it incite or aid the government policies for accountability and regulation of private AI businesses with global influence?

Perspectives

By identifying, acknowledging, and understanding these issues, artists can find new ways to intervene critically and productively in the AI-influenced sociopolitical reality. To challenge or tangibly affect the AI power regimes and mobilize the audience with a lasting impact, artists need to maintain a critical outlook on their poetic devices and balance their procedural skills with motivational sincerity and ideational cogency. This ethos of balanced competencies requires cultivation through expressive diversification, experimental freedom, playfulness, conceptually strong hacking, and imaginative discovery.

Artists can benefit from epistemic humility to develop more rigorous criteria for creative thinking and acquire wider knowledge of historical, theoretical, cultural, and political contexts in which they produce and present their works (Grba, 2022a, 24). This will help them address the potentially adverse expressive circumstances and clear the way for meaningful poetic directives. The inherently political nature of AI (Pasquinelli, 2019) obliges artists not just to exploit their expressive means but deconstruct and explore them by recognizing the injustices in the notional, relational, economic, and other layers of their sociotechnical environments. To counter the recuperative sway of the corporate sector, art market, and academia, artists should strive for integrity by recognizing, objectively assessing, and correcting the

systemic value issues of these systems. They need to resist prioritizing career building over artmaking, pursue external support with skepticism toward institutional rationales for art sponsorship, and be open to taking genuine risks by evolving potentially hazardous ideas.

The artists' proverbial inclination toward opportunism calls for decidedly independent professional strategies and sophisticated tactical interventions that evade the lures of commodification and complacency. These strategies require acknowledging art and technology as human dispositives within anthropological and sociocultural perspectives, and recognizing that creativity is driven by competitive ambitions and thus inherently instrumentalizable. In that light, artists should configure and respect their methodologies as experiential processes whose outcomes inform the audience by stirring inquisitiveness and critical thinking, stimulating imagination, and encouraging progressive action. Aided by subterfuge and humor, such tactical frameworks can be more impactful than superficial aestheticization, spectacularism, or overexplanation.

The contemporary AI landscape provides a range of resources for artists to reveal its sociopolitical contradictions and point out that science, technology, businesses, and education need a thorough reconfiguration and improvement of epistemological and ethical standards facing the increasing complexity of human existence. By demystifying their expressive contexts and the seemingly radical capabilities of their tools, artists can leverage AI issues as critical assets with broad political significance. The responsible treatment of these assets can build new insights into human nature and provide meaningful posthumanist perspectives (McQuillan, 2018).

References

Arnason, H. H., & Mansfield, E. C. (2012). *History of Modern Art* (7th ed.). Pearson.

Artbasher. (2007, July 14). Pulling my Leg? COCA Anthony Harper Award for Contemporary Art 2007. *Artbash Forum* (archived by The National Library of New Zealand). <https://ndhadeliver.natlib.govt.nz/webarchive/20090121190031/http://www.artbash.co.nz/article.asp?id=1050>

Arts at MIT. (2022). Agnieszka Kurant. *MIT Center for Art, Science and Technology*. <https://arts.mit.edu/people/agnieszka-kurant>

Audry, S. (2021). *Art in the Age of Machine Learning* (Kindle ed.). MIT Press.

Baio, A. (2008). The faces of Mechanical Turk. *Andy Baio's Website*. https://waxy.org/2008/11/the_faces_of_mechanical_turk/

Barbrook, R., & Cameron, A. (1995/2008). The Californian ideology. *Mute*, 1(3). <https://www.metamute.org/editorial/articles/californian-ideology>

Bard, P. (2014). *Man with a movie camera: The global remake*. Perry Bard's Website. <http://www.perrybard.net/man-with-a-movie-camera>

Basanta, A. (2018). All we'd ever need is one another. *Adam Basanta's Website*. <https://adambasanta.com/allwedeveryneed>

Berdugo, L. R., & Martinez, E. (2020). Seeing blocks and crypto bros. *Media-N: The Journal of the New Media Caucus*, 16(1), 79–98.

Bishop, C. (2012). Delegated performance: Outsourcing authenticity. *October*, 140, 91–112. MIT Press. https://doi.org/10.1162/OCTO_a.00091

Browne, K., & Swift, B. (2019). The other side: Algorithm as ritual in artificial intelligence. In *Extended abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–9). Association for Computing Machinery. <https://doi.org/10.1145/3170427.3188404>

- Browne, K., & Swift, B. (2020). Enacting collective ownership economies within Amazon's Mechanical Turk. In *Extended abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://storage.googleapis.com/kieranbrowne-public-files/preprints/enacting_collective_ownership.pdf
- burrough, x. (2020). A decade of working with the working crowd. *Media-N: The Journal of the New Media Caucus*, 16(1), 116–140.
- Carroll, C. M. (1975). *The great chess automaton*. New York: Dover Publications.
- Cetinić, E., & She, J. (2022). Understanding and creating art with AI: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 18(2), 1–22. <https://doi.org/10.1145/3475799>
- Chung, S. (2020). Selected works by Sougwen Chung. *Sougwen Chung's Website*. <https://sougwen.com/artworks>
- Cohen, H. (1995). The further exploits of AARON, painter. *Stanford Humanities Review*, 4(2), 141–158.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Curry, D. (2023). Artistic defamiliarization in the age of algorithmic prediction. *Leonardo*, 56(2), 177–182.
- Donnarumma, M. (2023). Eingeweide. *Marco Donnarumma's Website*. <https://marcodonnarumma.com/works/eingeweide>
- Elish, M. C. (2010, January 29). Representing labor: Ten thousand cents and Amazon's Mechanical Turk. *Furtherfield*. <https://www.furtherfield.org/representing-labor-ten-thousand-cents-and-amazons-mechanical-turk/>
- Forbes, A. G. (2020). Creative AI: From expressive mimicry to critical inquiry. *Artnodes*, 26, 1–10.
- Garcia, C. (2016). Harold Cohen and AARON—A 40-year collaboration. *Computer History Museum*. <https://computerhistory.org/blog/harold-cohen-and-aaron-a-40-year-collaboration>

Golumbia, D. (2015, August 5). The Amazonization of everything. *Jacobin*. <https://www.jacobinmag.com/2015/08/amazon-google-facebook-privacy-bezos>

Gradecki, J., & Curry, D. (2017). Crowd-sourced intelligence agency: Prototyping counterintelligence. *Big data & society*, 4(1), 1-7. <https://doi.org/10.1177/2053951717693259>

Grba, D. (2020). Alpha version, Delta signature: Cognitive aspects of artefactual creativity. *Journal of Science and Technology of the Arts*, 12(3), 63-83. <https://doi.org/10.34632/jsta.2020.9491>

Grba, D. (2021). Brittle opacity: Ambiguities of the creative AI. In M. Verdicchio et al. (Eds.), *Proceedings of xCoAx, 9th Conference on Computation, Communication, Aesthetics, and X* (pp. 235-260). i2ADS.

Grba, D. (2022a). Deep else: A critical framework for AI art. *MDPI Digital*, 2(1), 1-32. <https://doi.org/10.3390/digital2010001>

Grba, D. (2022b). Lures of engagement: An outlook on tactical AI art. In M. Verdicchio et al. (Eds.), *Proceedings of xCoAx 2022, 10th Conference on Computation, Communication, Aesthetics, and X* (pp. 58-74). i2ADS. https://doi.org/10.24840/xCoAx_2022_39

Grba, D. (2023a). Renegade X: Poetic contingencies in computational art. In M. Verdicchio et al. (Eds.), *Proceedings of xCoAx 2023, 11th Conference on Computation, Communication, Aesthetics, and X*. i2ADS.

Grba, D. (2023b). Faux semblants: A critical outlook on the commercialization of digital art. *MDPI Digital*, 3(1), 67-80. <https://doi.org/10.3390/digital3010005>

Hastie, J. (2021). *About*. Joanne Hastie's Website. <https://joannehastie.com/pages/about>

Hawkins, M. (2021). Randomized living. *Max Hawkins' Website*. https://maxhawkins.me/work/randomized_living.html

Herndon, H. (2022). Holly+. *Holly Herndon's Website*. <http://www.hollyherndon.com/holly>

Hill, S. (2016). How the un-sharing economy threatens workers. In T. Scholz & N. Schneider (Eds.), *Ours to hack and to own* (pp. 42-46). OR Books.

Hopkins, D. (2000). *After modern art 1945-2000*. Oxford University Press.

Hutson, M. (2012). *The 7 laws of magical thinking: How irrational beliefs keep us happy, healthy, and sane*. Hudson Street Press.

Jay, R. (2001). *The automaton chess player*. In *Jay's journal of anomalies* (p. 148). Farrar, Straus and Giroux.

Kearns, M., & Roth, A. (2019). *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press.

Kihara, T. (2020). TheirTube. *Project Website*. <https://www.their.tube>

Koblin, A. (2015). Aaron Koblin. *Aaron Koblin's Website*. <http://www.aaronkoblin.com>

LABoral. (2011). The sheep market. *LABoral Centro de Arte y Creación Industrial Website*. <http://www.laboralcentrodearte.org/en/recursos/obras/the-sheep-market>

Larson, E. J. (2021). *The myth of artificial intelligence: Why computers can't think the way we do*. The Belknap Press of Harvard University Press.

Lovink, G., & Rossiter, N. (2005, October 17). Dawn of the organised networks. *Nettime*.

Marcus, G. F., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Pantheon Books.

McCarthy, L. L. (2013). Social Turkers. *Project Website*. <http://socialturkers.com>

McCormack, J., Cruz Gambardella, C., Rajcic, N., Krol, S. J., Llano, M. T., & Yang, M. (2023). Is writing prompts really making art? *arXiv Preprint*. <https://arxiv.org/abs/2301.13049>

McQuillan, D. (2018). Manifesto on algorithmic humanitarianism. In *Reimagining digital humanitarianism*. Goldsmiths, University of London. https://www.researchgate.net/publication/323689097_The_Manifesto_on_Algorithmic_Humanitarianism

Meller, A., Seal, L., El, S., Abass, Z., Engineered Arts, & Oxfordians. (2019). *Ai-Da*. Project Website. <https://www.ai-darobot.com>

Mitchell, M. (2019). *Artificial intelligence: A guide for thinking humans*. Kindle ed. Farrar, Straus and Giroux.

Molderings, H. (2010). *Duchamp and the aesthetics of chance: Art as experiment*. Columbia University Press.

Natale, S. (2021). *Deceitful media: Artificial intelligence and social life after the Turing test*. Oxford University Press.

Pasquinelli, M. (2019). How a machine learns and fails—A grammar of error for artificial intelligence. *Spheres—Journal for Digital Cultures*, 5, 1-17. <https://spheres-journal.org/wp-content/uploads/spheres-5-Pasquinelli.pdf>

Powles, J. (2018, December 7). The seductive diversion of ‘solving’ bias in artificial intelligence. *Medium OneZero*. <https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53>

Quaranta, D. (2020). Between hype cycles and the present shock: Art at the end of the future. *DOC#6*. Nero Publications. <https://www.neroeditions.com/docs/between-hype-cycles-and-the-present-shock>

Raley, R. (2009). *Tactical media*. University of Minnesota Press, pp. 1-30.

Richardson, M. (2012). Descriptive Camera™. *Matt Richardson's Website*. <http://mattrichardson.com/Descriptive-Camera/index.html>

Ruff, G. E. (2022). Artificial empathy: Embodied connections between viewers and techno-art. Presentation notes for the Future Bodies 2022 New Media Caucus Symposium. Virginia Tech.

- RyBN. (2021). Human computers. *RyBN Website*. <https://rybn.org/human-computers>
- Schwettmann, S. (2017). Visual art. *Sarah Schwettmann's Website*. http://www.cogconfluence.com/2016collab_smse
- Sedlaček, S. (2020). Oblomo. *Sašo Sedlaček's Website*. <https://sasosedlacek.com/oblomo>
- Segni, G. (2013). The middle finger response. *Project Website*. <http://www.crowdworkersoftheworldunite.com/>
- Segni, S. (2018). Demand full laziness: A five year plan. *Project Website*. <https://demand-full-laziness.today>
- Stanford, N. J. (2017). Automatica: Robots vs. Music. *Nigel John Stanford's YouTube Channel*. <https://youtu.be/bAdqazixuRY>
- Taylor, G. D. (2014). *When the machine made art: The troubled history of computer art*. Bloomsbury Press.
- Thompson, J. (2020). Human computers. *Jeff Thompson's Website*. <https://www.jeffreythompson.org/projects/human-computers.php>
- Thurston, N. (2013). *Of the subcontract: Or principles of poetic right*. Information as Material. https://monoskop.org/images/d/db/Thurston_Nick_Of_the_Subcontract_or_Principles_of_Poetic_Right_2013.pdf
- Torok, A. (2020). Regarding artists regarding labor. *Media-N | The Journal of the New Media Caucus*, 16(1), 24–44.
- Valla, C. (2023). Works. *Clement Valla's Website*. <http://clementvalla.com/work>
- Van Arman, P. (2016). Teaching creativity to robots. *TEDx Talks YouTube Channel*. <https://youtu.be/YYuOPdJSZcA>
- von Graevenitz, A. (1977). We walk on the planet earth: The artist as a pedestrian—The work of Stanley Brouwn. *Dutch Art + Architecture Today*, 1, 285–294.

Wallace, I. (2014, June 11). The evolution of the artist's studio, from Renaissance bottega to assembly line. *Artspace, ART 101*. https://www.artspace.com/magazine/art_101/art_market/the-evolution-of-the-artists-studio-52374

Yi, H. (2021). Huang Yi and KUKA. *Huang Yi's Website*. <https://huangyistudio.com/archives/project/huang-yi-and-kuka>

Zeilinger, M. (2021). *Tactical entanglements: AI art, creative agency, and the limits of intellectual property*. meson press.

Żylińska, J. (2020). *AI art: Machine visions and warped dreams*. Open Humanities Press.

Desktop Films: Posthuman Gaze and Technogenesis in Contemporary Cinema

Jelena Mišeljić

Introduction

Desktop or computer screen films appear to have anticipated a new context for understanding the everyday environment. Since the pandemic year of 2020, our actions are no longer tied to the once-privileged physical, tactile reality perceived through all our senses. In an era characterized by public health risks associated with physical proximity, the convenience of communicating from home has transformed our interactions into pixelated talking heads set against virtual Zoom backgrounds. In this global moment, this type of film is likely to hold a significant place in future film studies.

The cameras of our devices capture our faces, often without our knowledge. This involuntary cooperation between computer and human intelligence, along with the critical visual aspects of its operation, constitutes one of the primary concerns of computer screen

Jelena Mišeljić, Teaching Assistant, Faculty of Dramatic Arts, University of Montenegro, Cetinje ORCID: 0000-0003-1610-1327, email: jelena.mi@ucg.ac.me

DOI: <https://doi.org/10.5281/zenodo.14205074>

films. Automated technologies that enable facial recognition based on stored information about our features, combined with the logic of desktop functioning, are distinctly imbued with posthumanist themes. This paper will explore this new type of film as an increasingly common tendency in contemporary cinema that adapts to the circumstances of everyday digital life. Computer screen films will also be treated as an artistic concept reflecting the processes of technogenesis, a fundamental idea within the framework of posthumanist studies. Given this perspective, the selected films can be analyzed as artifacts of contemporary trends.

Desktop films emanate a nuanced posthuman perspective regarding cinematic practices involving surveillance technology. Beyond the fact that protagonists are subjected to the gaze of webcams, they are also positioned within the screens of various platforms. Thus, the posthuman gaze necessitates a comprehensive theorization of the perception of boundaries between physical and virtual environments, as well as the human body. In this context, desktop films—considered a post-cinematic form—serve not only as representations of contemporary technogenesis and co-evolution with digital technologies but also as their artifacts.

These films prompt us to revisit questions concerning corporeality, the perception of our environment, and physical space, as well as the significance of the materiality of the body and its surroundings. What is lost in distance communication mediated by screens? What qualities do our bodies, objects, and spaces possess in physical presence compared to the virtual? Has the significance of representations been overtaken by the very material from which the world is composed? In an era characterized by composite images and bodies, can this polarity endure?

As a new film tendency, desktop films are analyzed in this paper as a reflection of technogenesis and the co-evolutionary processes between humanity and technology. Such analysis allows for a better understanding of the expanding theoretical horizons not only of film but also of perceptual and physical media. In the context of contemporary networked culture, which assumes increasing importance under various social circumstances, examining desktop films enables us to question not only the erasure of boundaries between technologically and biologically mediated realms but also the nature of mediation itself through screens and various flat bodies. This interplay drastically complicates the relationship between the virtual and material worlds.

Desktop Cinema: Possible Definition and Related Theories

The history of computer screen film is not very rich, nor has special attention in cinema anthologies and literature, and the reason, we assume, is the fact that it is a relatively new term that includes works that are primarily independent and low-budget productions and which are still a somewhat marginalised curiosity of certain genre subcultures. The founder of the term *screenlife* is producer Timur Bekmambetov. However, even before his idea of founding a company dedicated to this type of film, several titles were produced that could be classified in the same category. The main characteristic of this film form is storytelling exclusively through the computer screen. The entire action of the film takes place on the protagonist's computer screen as the only diegetic space.

Long before our computers got cameras, their screens were the fascination of experimentalist practices of video art, both as artefacts of the current moment that document, on the one hand, and as a means of artistic expression, on the other. James Benning's film *Pascal's Lemma* communicates with the works of Andy Warhol from the digital multimedia phase, in that, like his works of art on the Commodore Amiga model, the computer is put in the foreground, but in this case, it is treated in two ways – as a means and a protagonist. On the other hand, with the development of the Internet and new telecommunications technologies, screens have become a portal to the world of abstract networks and binary codes. The computer screen presentation was essential to the author's imagination, mainly in films primarily concerned with glamorising the possibilities of computers or hacking. Underground. So, for example, in the film *Hackers* (Iain Softley, 1995), the computer screen is approached as one of the leading sets: the film's action takes place in locations that broadly vary between neon urban, cyberpunk landscapes and computer screens, specially designed for the movie.

Today, computers *perceive*, i.e. as if they see and hear, like stylised Cyclops circuits. In addition to our fascination with the advanced design and functioning of the interface, these machine senses provoke some other preoccupations related to the distribution of the world's perception into its living and non-living elements. Computer eyes are just as present as living ones, and what they register becomes integral to every day. There are numerous scopic regimens to which special attention should be paid in the future. The presence and operation of automated eyes, in the form of a small web camera designed to serve telecommunication needs primarily, thus influenced the complexity of the idea of the screen, which thus becomes our mirror in the literal

sense. Thanks to web cameras, computers have become especially interesting for filmmakers. Investigations of the computer screen in the film thus move from visual representations of digital abstract worlds to the horizons of questioning how our personal computers see us. Taking all this into account, computer screen movies represent recorded user activity narrated via desktop, which is integrated into our daily activities, on an equal footing with off-screen life. In addition, the desktop film can be considered a part of that perception of the world that does not belong to us but to an inanimate environment in which technology plays an important role. Such a film, in the end, reflects the coexistence, the interaction between the living and the inanimate, the virtual and the real, that is, the communication between us and the computer, in which a new and permanently changing perception of the world is formed through mutual influence, and which is crucial to understanding technogenic processes.

Aesthetics of a Memory Box

In one of the interviews, Bekmambetov stated: “You cannot fully understand today’s world and fully know people if you do not look at their screens” (Yamato, 2018). With this sentence, he seems to be paraphrasing the words of Lev Manovich from the essay *Archaeology of the computer screen*: “We may debate whether our society is a society of spectacle or simulation, but, undoubtedly, it is the society of a screen” (Manovich, 1955: 2). On this occasion, we will use it as an introduction to clarifying the basic term we are dealing with in this chapter – computer or computer screen. Manovich, in his genealogy

of screens, distinguishes between traditional and dynamic screens, where the dynamic is the one that can show a changing and moving image while thinking of film, television, and video screens. When it comes to a computer or computer screen, according to Manovich, it belongs to the interactive type of dynamic screen. It is important to note that its origin dates back to the middle of the twentieth century when it was developed for military purposes and only became known and available to the general public. It is also interesting that Manovich predicts in the essay that television sets will adopt the computer window metaphor in the future, as evidenced today by various video-on-demand platforms and streaming services. However, we are witnessing that the “window metaphor” goes even broader than that. It moves into the space of the film medium.

What is so fascinating about a computer screen’s appearance, the “aesthetics”? In an essay dedicated to QuickTime movies, writes:

The organisation reassures me, clarity and order of consistent and logical menus, commands and systems with the help of which I can access a huge amount of information (...) The virtual solidity of my desktop and files is opposed to the ambiguities of poetry – and any discomfort I feel facing my mistakes, losses and frustrations does not stem from my ‘well-fitted cube’ but from a very human and irregular logic (Sobchack, 2004: 310).

The idea of flawless precision of the “memory box”, in front of which we are aware of our “irregular logic”, leads us to the question of such articulated mutual relations. In other words, can we imagine what

memories of us are contained in that “box”, that is, how our computer sees us. The idea of a computer that watches us questions the limits of observation and seeing as abilities limited to the living world. There is a profound posthuman logic in this that operates beyond our perception.

The space in a desktop film is the space of the computer screen, and the controlled content constitutes the diegesis. By completely excluding the representation of the physical space in which protagonists are located (the room or office where s/he sits), affective involvement is achieved, without the impression of intrusion, of imposing on someone’s private space. In the already quoted essay, Manovich writes the following:

Although the screen, in reality, is only a window of limited dimensions positioned inside the physical space of the viewer, the latter is supposed to completely concentrate on what is seen in this window, focusing attention on the representation and disregarding the physical space outside. (Manovich 1995, 4)

In the case of a computer screen film, we are faced with the visual concept of a *screen within a screen*, where the film spectator entirely depends on the decisions of another viewer – the protagonist (*user*) and the language of the computer (and Internet) interface rather than classic film storytelling in the diegetic space and time. Desktop films are one of the first attempts to represent the computer interface language within the film medium. Everything immanently related to the space and time of the events in the film is determined by computer operations on the screens, which is the essential feature of this type of

film. In this way, we are also provided with an extended viewing experience. The viewer, placed in the position of the protagonist who manages the software, “accepts the user experience as his experience, as another first-person storytelling experience” (Mišeljić, 2021: 42).

In the book *How We Think: Digital Media and Contemporary Technogenesis*, N. Katherine Hayles interprets technogenesis as a co-evolutionary process in which technologies fundamentally change how people think, perceive and act in the world. Hayles opens up the possibility of theorising the human that becomes posthuman. She also writes that “humans have always been integrated into their environment and have co-evolved with the environment” (Pötzsch, Hayles, 2014: 98), so the question arises: what comes first – technology or human? Consequently, understanding this human-technology interdependence is critical to understanding the posthuman condition. According to Hayles, technogenesis is a type of *ontogenetic* evolutionary process. She uses the metaphor of ontogenesis as a developmental process of combined epigenetic changes to clarify the implications of structural changes in almost all (cognitive, physical, emotional and social) aspects of the human being, which are also manifested in modifications of the perceptual apparatus. For example, the use of smartphones affects morphological and cognitive changes in the human brain, which in turn causes the need for social and cultural adaptations. Thus, we can say that in a certain way, the films of computer screens reflect technogenic processes, especially dealing with the issue of the complexity of internet communication.

The boundaries of the human body and its technological extensions are questioned by immersing viewers in the digital space of a network represented by a computer interface and by intertwining material and

imaginary worlds that mix with screen images. Friedrich Kittler wrote about how the human mind is defined according to the models and metaphors of technical media. Kittler analyses how at the very beginning of the twentieth century, that is, immediately after the first public film screenings, there was a significant increase in the number of alpinists who described the moments of facing death as a result of fatal falls with the phrase that “their whole lives flew before their eyes, the only thing was not clear whether that film is being unwound forwards or backwards” (Kittler, 2012: 34). Kittler concludes, “the soul advanced technically and turned into the film” (ibid.: 35). under the direct influence of the fascinating audiovisual medium that preoccupied the worldview of social life. The cases Kittler takes describe a new relationship between man and technology and between man and media. This relationship can no longer be interpreted with McLuhan’s metaphors: they are not just *extensions*, but units that follow their autonomous logic, participating with the human subject. Does the digital environment also shape our visions, hallucinations, dreams, and life stories? Technogenesis is not a one-way impact of technology on man but a mutual, correlative and habitual interweaving of technique, technology and human being that forms a digitally extended social environment, or as Hayles points out:

In contemporary technogenesis, we are not talking about progress. In other words, there is no guarantee that these mutual dynamic transformations of people and technology will go in a positive direction. This is more about adaptation, fitting organisms and their environment, recognising the coordinated transformation of both sides (people and technology) (p. 173).

Certain Tendencies in Desktop Cinema

Concerning all the implications of technogenesis from a primarily theoretical perspective, desktop films can be analysed as a language and artefact of our current, contemporary, digital intelligence and imagination, which mix with concrete, technical representations and operations. In this chapter, we will try to draw a map of desktop film's genre, stylistic and thematic preoccupations and create a platform for their consideration in the broader social context of digital culture. In the tradition of the found footage film, the tendencies of the computer screen film move between the horror and thriller genres, within which we can observe the shift of the narrative conventions.

Cyber-horror

In the context of the horror film and on the succession of the found footage genre, what is noticeable in these films is that the stronghold of fear and horror is directly related to the medium by which the film was recorded. The first feature films of computer screens coincided with the trend of video call options the same way *The Blair Witch Project* (Daniel Myrick, Eduardo Sanchez, 1999) was a cinematic response to the expansion of digital home cameras. Additionally, desktop films, in a certain way, have anticipated not only the software that makes this possible but also all those internet communication problems, which were further complicated by the increasing use of social networks. The parable between a terrifying new technology and a

malevolent paranormal force is clear. In both cases, it is a menacing presence, abstract in its invisibility and frightening in its ability to see us even when we do not know it. When it comes to the visual plan, it is fully adapted to the computer design and interface of the time in which it was created. This is an entirely logical tendency that follows this type of film. Here, however, it is not only about visual design but also about how the role of new techniques and media is treated in the same dramaturgy of the film. Today, for example, it is unthinkable to write a screenplay without smartphones, which can be used to solve most problems, without either a particular conceptual author's approach or a very skillfully imagined narrative justification that would achieve the suspension of the viewer's disbelief. Computer screen films, following in the footsteps of *The Collingswood Story* (Mike Constanza, 2002), still treat the screen as a communication *window* within which the narrative takes place. However, what is noticeable is that this window is no longer a mirror of what is happening in the physical space but becomes a portal for evil forces. The horror of invasion is moving from the physical personal area to the space of the personal computer.

The aforementioned independent horror *The Collingswood Story* stands out among the first film titles for computer screens. The film's story is focused on a young couple trying to maintain a long-distance relationship after moving to two different cities due to college. They use webcams and video calls for communication to stay in close contact. However, the fascination with the webcam and the possibilities of the software (designed for the needs of the film) leads them to a rather risky entertainment: randomly choosing chat buddies who stream their lives 24 hours a day, when suddenly the space (both physical space of a house that appears to be haunted and virtual space of

the network) becomes invaded by an evil force. In contrast to the stylistic trend of the hacker films of the nineties, the protagonists are shown exclusively within the window of the video calling software they use. On the visual level, the director uses two types of shots: shots of the screen itself with the desktop content (window for video calls) and shots that contain only close-up or medium-close-up of the protagonists while they are communicating, looking directly into the camera. Sound design is also important, considering that there are two types of auditory images: the ambient sound of the room where the protagonists are, and the sound filtered through the video call software. In this way, the viewer, intentionally or not, is confused. Namely, the question arises from whose perspective we follow the events. Whose point of view is that? The second group of shots consists of close-ups of the protagonists who speak so that they directly address the camera, that is, the viewers. This point of view belongs to the computer camera. We, as viewers, participate in video communication through the machine vision representations of the web camera watching the protagonists.

In *The Collingswood Story*, we can say that the webcam's view invades the personal space of the film's protagonists and observes them as they share private moments of weakness. On the other hand, with the more significant development of the Internet, even these unprotected areas of artificial computer intelligence have become resistant to intrusions. Computer invasion is a term primarily used in glossaries dealing with high-tech and cybercrime, and it is interesting in all the ways we can think about the connotations of this definition. Let's take, for example, computer viruses, which are a real threat not only to individual devices and data but also have implications in the broader framework of various systems and infrastructures. Although social

networks are proven to function as corporate platforms that retain the rights to record, store and further exploit our data, we perceive this relationship as agreed. We agree to it because of all the benefits that social networks have: from maintaining contact with many people and business promotion to the instant gratification we experience with each notification. Conversely, the virus is an antagonist, an uninvited, invisible and intrusive force that, concerning such nature, can be equated with supernatural, spectral phenomena.

Another example is the horror film *Unfriended* (Levan Gabriadze, 2014), which problematises peer violence in the context of social networks – cyberbullying. The story of the film, set in real-time, deals with a group of high school students during a Skype call on the anniversary of the suicide of their friend, Lora. Lora was publicly shamed when her compromising video was published on social networks, which was also the main reason for her suicide. During the Skype session, we find out exactly what happened, and we gradually discover that this group of people, the protagonists in the film, are responsible for the leaking of the footage. The more we learn about their guilt, the more Lora's spirit takes possession of the video call, carrying out brutal murders of everyone in the session, using the video call platform itself as a portal through which it passes from the virtual, i.e. digital, to the physical zone of the film's diegesis. It seems that in this film of the computer screen, the idea of an evil spirit occupying the digital space is very clearly set as a kind of metaphor about the digital footprint, that is, about all those versions of our identities and actions on the Internet that can never be erased entirely. The second part of the franchise, *Unfriended 2* (Stephen Susco, 2018), deals with cybercrime even more directly. Similar to the first part, a group of friends gather for a Skype call because one of them accidentally found a laptop that has access

to the dark web. During the film, we learn that they are being followed by a group of hackers who decide to involve them in a morbid game.

The *Unfriended* franchise deals with the problems of the Internet, social networks, and cross-platform communication. We can say that this transposition of conventions of *home invasion* into *computer invasion* takes on the characteristics of an independent genre, as evidenced by the tendency that also left a specific contribution to the film industry during the Covid-19 pandemic. The film *Host* (Rob Savage, 2020) is a computer screen film that was created during the lockdown. Due to the closing of cinemas, the film was distributed on video-on-demand platforms, with no intention of being shown on the big screen. In the film, we follow a Zoom video call in which a group of friends participate who, wanting to break the monotony of self-isolation, organise an online session of summoning spirits with the support of a medium. Of course, everything goes in the wrong direction due to the approach of some of the participants who understand the session as entertainment, so the whole event turns into a bloodbath caused by an evil spirit whose movement crosses the boundaries between virtual and real. In this film, as in *Unfriended*, there is a motif of a demon that takes over the physical space of the video call participant via the Internet. In *Host*, at the beginning of the session, the medium even points out one specific thing regarding the uncontrolled behaviour of the evil spirit: this entity can behave like that because of the strength of the so-called Circle of Trust that is formed when all participants of the session are in one physical space. The film thus recalls some of the famous home invasion horror films where *ghost-in-the-machine* is now transposed to the plane of the internet space, which is represented through the computer screen.

To conclude, in the mentioned films, there is a tendency for directorial treatment of the digital space as a platform for the action of malevolent, otherworldly forces on both the thematic and narrative levels. The technological and digital *otherworld* is thus presented as a threat that we cannot defend against with *earthly* methods. In this way, these films, in a somewhat slight technophobic tone, remind us of the slasher horror subgenre. The conservatism (albeit ironic) that these films exude is contained in need to brutally punish teenagers for their “arrogant” behaviour: no depraved adolescent can escape the “hand of justice” of a murderous psychopath. However, unlike slasher films, there are no survivors in cyber-horror films, and such a directorial decision can certainly be attributed to a sufficiently profitable dose of contemporary cynicism that attracts the audience of the millennial generation. In the context of posthumanist preoccupations, i.e. the implications of the use of automated cameras, we will not be wrong if we place these films in the category of dystopian fiction in which restless spirits after the death of their physical, biological bodies linger and are embodied as ghostly, immortal data about registered activities on the networks.

Missing Person

Techno-thriller

In addition to horror, missing person thriller is a trendy subgenre of desktop movies whose focus is *the search*. What, on a general level, characterises these films (in addition to the elements of suspense, mystery and emotional involvement in discovering the solution together with the protagonist) is the exploration of different places and

physical locations in order to find the missing person. This is why the transposition of this genre into a computer screen film, where the entire investigation process takes place, is particularly interesting.

One of the most top-rated titles of this genre is *Searching* (Aneesh Chaganty, 2018), where the story follows a father's search for his missing teenage daughter. After the disappearance of his sixteen-year-old daughter Margot, widower David Kim decides to join the search and help the police by accessing the contents of her laptop. David enters Margot's e-mail and accounts on social networks, revealing her private life, which was utterly unknown to him since the death of his wife and Margot's mother. The case is led by a detective who suggests that Margot has, in fact, escaped. In bursts of disbelief and despair, David goes deeper and deeper into the investigation and discovers that the detective was laying wrong clues with the aim of covering up the revelation: her son, Margot's stalker, has been catfishing her and organised a secret meeting. The case is solved, David finds Margot, and they continue their lives now a little more focused on each other. As the title of the film suggests, this is a film about a quest. In addition to the basic thematic line of the film (the search for a missing person), the term *searching* is, however, more adaptable to the idea. It is about searching as a leitmotif; search, in this sense, is one of those generally accepted concepts of interface language: "search in an Internet browser (David tirelessly searches the Internet for any information he needs), search for documents in (someone's) personal computer (David also searches a private computer his daughter, her accounts on social networks, e-mail), obsessive search for a person from social networks (the search for Margot's stalker), search for the reason for dysfunction in the family (David at one point thinks that Margot ran away because he was not enough a good parent)" (Mišeljić,

2021: 41). Unlike the paranormal phenomena that occupy the space of the personal computer, here, the uninvited and intrusive element is – the parent. Nevertheless, the process of going through other people's personal accounts is not treated as problematic at any point of the film – the necessity of such an act is narratively set as justified. In this sense, under the justification of disappearance, any act of intrusion into personal space (personal orders on the computer, e-mail, social networks) is acceptable, as is, after all, the genre convention of reading the diary of a missing person, not only by the police but also by close people. The problem of false identity on social networks is an obvious directorial intention in the missing person genre of desktop film. In that sense, the film *Profile* (Timur Bekmambetov, 2018) deals with the fight against terrorism. This adaptation of a novel based on a true story focuses on the story of a British journalist who, on a secret mission to uncover the identity of an ISIS recruiter, falsely presents herself as a potential candidate, trying not to be lured into their ranks. The action of the film takes place through Facebook and Skype communication between her and the terrorist in a game of changing false identities.

As we pointed out, Desktop films contain an inherent idea of discomfort in the realisation that the contemporary perception of the world is reflected in the processes of mutual interaction between living and non-living elements. The scope of the inanimate world, however, here refers not only to the computer and its software but primarily to the erasure of the boundaries between the physical and the virtual, which the cinematic space of the desktop emanates. If the inanimate world is made up of things or *objects*, in this case, the Internet can be treated as a *hyperobject*, according to Timothy Morton. In the book *Hyperobjects: Philosophy And Ecology After The End Of The World*,

Morton writes about interobjectivity, defining it as one big configurational space that makes up everything that surrounds us. According to him, interobjectivity is a broader concept than intersubjectivity, which he sees as a “local, anthropocentric instance of interobjectivity”, the one “with which people are familiar” (Morton 2013: 81). Interobjectivity is thus a post-anthropocentric concept, one that is inherent in hyperobjects, inanimate creations that permeate our environment. From climate change to the Internet, hyperobjects fill our environment at every moment, feeding the sea of information with data about all its users. This hyperobject, in which an endless database is gathered, occupies a certain physical space in the form of endless fields of processors that work simultaneously. Likewise, in the cinematic language of desktop films, the narrative events are articulated as the protagonist’s activities in the form of multiplying screens and windows, in terms of our “well-fitted cube”. The postanthropocentric gaze in the film of the computer screen, in this way, does not only refer to the visual representation of the gaze of our camera on the phone and computer but to the visual economy of the gaze of the hyperobject, the deep cyberspace where the webcam is just one window. The evil forces that the aforementioned films thematise come from are, in lovecraftian terms – inexplicable, space of zeros and ones, like the digital rain in *The Matrix* (The Wachowskis, 1999), into which our every action is transformed. The disdainful property of the postanthropocentric gaze also reflects the processes of contemporary technogenesis, in which our mental space is now shifting from the horizon of the traditional film image to the converged image of the screen space of a personal computer. Our reality is no longer linearly arranged, and physical landscapes no longer dominate it. The virtual spaces of computer animation and digitised worlds of Internet content have a significance equivalent to physical reality.

Consequently, this regime of the cinematic image has certain repercussions on the experience of watching the film. Desktop films, found footage, and CCTV films do not communicate with the big screen. A raw, unedited, decomposed, moving digital image created by automated mechanisms and software interfaces is almost unimaginable in a cinema. What audiences are offered is a raw user experience, simple in its impurity, the experience of a user immersed in the abstract space of the machine. What the desktop film aesthetic emanates, indirectly and perhaps unintentionally, is the question of the rematerialisation of our bodies into images and the physical spaces into the background of video call screens. The context and strategies used by these films are essential from the aspect of understanding the language that resists the conventions of the classical narrative principle of the film camera as an invisible observer. The viewer is in the position of one of the participants in technologically mediated storytelling, which refers to the entire space and design of the film. As Hayles writes, in the posthuman, there are no essential differences between bodily existence and computer simulation, cybernetic mechanism and biological organisms, robotic teleology and human goals (Hayles, 2012).

Conclusion

This paper examined some of the main preoccupations of contemporary digital culture through the analysis of the emerging genre of desktop films. Various directorial approaches to the themes like surveillance, internet communication, multiplication of identities on social networks and uninvited intrusions into personal space have one main point in common, and that is – the loosening of the boundaries

between virtual and physical realities, or more precisely, the tendency to erase them. This type of film somewhat indicates different perceptions of the body in the digital world. These bodies are translated into two-dimensional images on the screens of platforms that serve for communication. The relationship between the human and the non-human, the question of corporeality, which is now a real turn towards two-dimensionally represented bodies within the second screen, and even the very strategies of representation of different screens within the film screen, are topics that are clearly articulated in the aspects of posthumanist thinking. Using the perspective of the computer user on whose screen the action takes place as a basic narrative strategy, the desktop films simulate the user's experience and our somewhat frenetic and dispersed online behaviour, which is moving toward the postanthropocentric environment.

This research aimed to examine how digital film genres reflect a context in which it is necessary to reconfigure the once dominant anthropocentric view. This confirms the basic hypothesis that the automated image, in the function of an expressive film medium, reflects the contemporary moment in which the experience of new media technologies takes over the role of human experience, thereby calling into question the stability of the human subject in the time of digital audiovisual representations of the posthuman state.

References

- Hayles, N. K. (2012). *How we think: Digital media and contemporary technogenesis*. The University of Chicago Press.
- Kittler, F. (2012). *Optical media: Berlin lectures*. Polity Press.
- Manovich, L. (1995). An archeology of a computer screen. http://manovich.net/content/04-projects/011-archeology-of-a-computer-screen/09_article_1995.pdf
- Mišeljić, J. (2021). Post-film narration in film 'Searching': Screenlife aesthetics and neoliberal brain. *Hrvatski filmski ljetopis*, 105, 34-44. Hrvatski filmski savez.
- Morton, T. (2013). *Hyperobjects: Philosophy and ecology after the end of the world*. University of Minnesota Press.
- Potzch, H., & Hayles, N. K. (2014). Posthumanism, technogenesis, and digital technologies: A conversation with N. Katherine Hayles. *The Fibreculture Journal*, 23, 95-107. <http://fibreculturejournal.org/wp-content/pdfs/FCJ-172PotzchHayles.pdf>
- Sobchack, V. (2004). Nostalgia for a digital object: Regrets on the quickening of QuickTime. In L. Rabinovitz & A. Geil (Eds.), *Memory bytes: History, technology and digital culture* (pp. 305-330). Duke University Press.
- Yamato, J. (2018, August 17). With 'Searching,' 'Unfriended,' and beyond, Timur Bekmambetov seeks a new cinematic language that mirrors our digital lives. *Los Angeles Times*. <https://www.latimes.com/entertainment/movies/la-ca-mn-timur-bekmambetov-searching-unfriended-dark-web-screenlife-20180817-story.html>
- Bekmambetov, T. (Director). (2018). *Profile* [Film]. Bazelevs Company.
- Chaganty, A. (Director). (2018). *Searching* [Film]. Bazelevs Company.
- Constanza, M. (Director). (2002). *The Collingswood Story* [Film]. Cinerebel Films.
- Gabriadze, L. (Director). (2014). *Unfriended* [Film]. Bazelevs Company.

Savage, R. (Director). (2020). *Host* [Film]. Shadowhouse Films.

Sanchez, E., & Myrick, D. (Directors). (1999). *The Blair Witch Project* [Film]. Haxan Films, Artisan Entertainment.

Susco, S. (Director). (2018). *Unfriended: Dark Web* [Film]. Bazelevs Company.

Wachowski, L., & Wachowski, L. (Directors). (1999). *The Matrix* [Film]. Warner Brothers.

Just Look like Us, or Look just like Us: The Measure of (Un)canniness

Predrag Krstić

Introduction: *Star Trek's* Narrative

Data is the model android of the *Star Trek* franchise, *Next Generation* series. An episode with a revealing title, “The Measure of Man” deals in a transparent, in-depth, and legally, ethically and philosophically very provocative way with the sufficient conditions for considering artificial minds persons. A scientific representative of the United Federation of Planets, the lead researcher in the field of cybernetics and robotics, Dr. Bruce Maddox demands that Data—an example of a very sophisticated ‘positronic computer’ in the galaxy, but not a sentient lifeform—be turned over to his department and subjected to a thorough analysis and experiments with the possibility of being irrevocably broken down, dismantled, and its memories and knowledge lost. In order to avoid such fate, the android, but also a high-ranking member of the space ship Enterprise crew, Lieutenant Commander Data,

Predrag Krstić, Senior Research Fellow, Institute for Philosophy and Social Theory, University of Belgrade, Belgrade ORCID: 0000-0002-2966-7082, email: predrag.krstic@ifdt.bg.ac.rs

DOI: <https://doi.org/10.5281/zenodo.14205119>

is forced to prove that he is a self-conscious being to whom the Federation owes the status of (inviolable) citizen.

The commanding officer of the Enterprise, and if we can put it in this way, friend of Data's, Captain Jean-Luc Picard, in a court case on the matter represents Data at the hearing. He dismisses the argument that Data is *only* a constructed being, a machine: "Do we deny that? No, because it is not relevant: we, too, are machines, just machines of a different type. Commander Riker [representative of Commander Maddox's position, against his will and by official duty], has also reminded us that Lieutenant Commander Data was created by a man; do we deny that? No. Again, it is not relevant. Children are created from the 'building blocks' of their parents' DNA. Are they property?"

After a while, Picard calls Maddox to the stand as an enemy witness, who unflinchingly claims that Data is not a sentient being and therefore not entitled to all the rights reserved for all life forms. Picard asks "tendentiously:" "Commander, would you enlighten us? What is required for sentience?" The answer goes: "Intelligence, self-awareness, consciousness." The second item in this list seems particularly interesting to Captain Picard: "Because you are conscious of your existence and actions. You are aware of yourself and your own ego." Not sensing a trap, Commander Maddox self-assuredly responds: "Because you are conscious of your existence and actions. You are aware of yourself and your own ego."

Captain Picard now moves in for a demonstration, addressing the defendant: "Commander Data, what are you doing now?" Data answers, in a reasoned manner and a voice impervious to affect: "I am taking part in a legal hearing to determine my rights and status. Am I a person or property?" Captain Picard insists on presenting Data's awareness

of the situation in which he finds himself: “And what’s at stake?” “My right to choose. Perhaps my very life” – answers Data, succinctly, without beating around the bush, as is his wont, presenting Picard with a chance to bring his argument to a head: “‘My rights.’ ‘My status.’ ‘My right to choose.’ ‘My life.’ It seems reasonably self-aware to me.” (Scheerer, 1989; cf. Short, 2003; Neuwirth, 2018)

The Amplitude of Distance

By androids in this article we mean robots, like Data, equipped with strong artificial intelligence¹—which are made in the image and likeness of humans. A certain ambivalence was already noted, long before Data’s age, towards the androids thus shaped: along with fascination (Antosca, 2019), envy and fear—at the very least those three—and all at the same time. Hence also the unavoidable story of the “uncanny valley” that seeks to explain the phenomenon of our ability to feel empathy towards androids nearly as much as towards humans, yet revulsion or even disgust too at the point when they display a high degree of human-likeness.

1 On the difference between strong and weak AI, cf. Bodley, 2015, pp. 44 et passim; Braga and Logan, 2017. In keeping with gender correctness, we also had in mind a rival term for “robot”—“fembot,” as well as a less frequently used equivalent for a male android: “gynoid.” (cf., e.g., Sharp, 2007; Ferrando, 2014). Solely for the sake of brevity, here we will stick to the already classic terms “android” and “robot”, not implying any specific gender by them.

Namely, the Japanese robotics scientist, Masahito Mori, hypothesized that human-like objects produce reactions similar to those produced by real people proportional to the degree of similitude. The more androids resemble humans, the more humans will relate empathetically to them. However, when a certain degree of similitude is achieved, suddenly the common emotion produced is one of revulsion. Finally, only when the object becomes indistinguishable from real humans, they illicit empathetic responses again. For the gap in the empathy continuum, Mori (1970), whether rightly or unhappily (cf. Bartneck, Kanda, Ishiguro, and Hagita, 2007), coined the phrase *Bukimi No Tani*. Rendered in English as *uncanny valley*, the phrase has over time come to designate any feeling of the perceiver's unease whenever a robot or computer-generated figure nearly perfectly resembles the human.

But both the feeling itself and the designated terms have their histories. Uncanny is the English version of the German *Das Unheimliche*, difficult to translate. In general, it is any feeling produced when in perceiving a familiar, close, intimate object one suddenly recognizes something unfamiliar, foreign, mysterious, uneasy, or terrifying. Ernst Jentsch introduced the term into psychology in his 1906 text "Über die Psychologie des Unheimlichen" (Jentsch, 1906) to mean the suspicion that a lifeless object might spring to life, that is, that a seemingly living being is actually living. It was adopted by Freud in his 1919 essay "Das Unheimliche" (Freud, 1986). Freud thought that the disturbingly unfamiliar in the familiar, the emotional duality in the face of an object that bears both the foreign and the intimate, indicated phylogenetically produced fears. Hence the perfect example being the perception of a dead body; although that of androids might be better still: they are truly separate objects, on the very limit of the division of people and non-people. They are not human, but we feel bad when we treat them as mere objects. The lack of empathy is even a sign of

“inhumanity,” as we see in Philip Dick’s *Do Androids Dream of Electric Sheep?* or in Steven Spielberg’s *AI* (Misselhorn, 2010, p. 102; Lee, 2006). It would appear that as much as we love to symbolically animate inanimate non-humans through the safety of anthropomorphization, we equally fear their actual coming to life.

Yet, Mori’s ambitions were greater than mere description of the path our experience of androids take: he suggests that we begin developing “an accurate map of the uncanny valley so that through robotics research we can begin to understand what makes us human” (Mori, 2005, p. 100). This is really then an anthropogeography, which only in its first step displays our increase and sudden dropping out of our “affinity for robots,” when the humanoid entity becomes too human-like but not yet entirely “correct.” The valley can be escaped, however, with affinity once again increased—hope springs eternal for robotics scientists—when reaching the point when the entity is no longer uncanny and comes ever so close or indeed overlaps with the level of ease possessed in relation to living humans.²

-
- 2 Naturally, this was followed by a debate about this intuitively rather plausible hypothesis and its confirmation, that is, rejection or correction. The most important empirical research on “the valley” was done by Karl F. MacDorman (2005a; 2005b; 2006; MacDorman, Minato, Shimada, Itakura, Cowley, and Ishiguro, 2005). Hiroshi Ishiguro contributed most significantly to the practical, robotics side of the discussion, constructing a few astonishingly humanoid robots, above all his own double, so-called Geminoid HI-1, a double for his four-year-old daughter and the allegedly prettiest fembot, Erica (Deutsche Welle, 2021; NBD, 2019; for visual evidence of an increasingly more faithful “humanization” of robots, cf. Aguilera-Hellweg, 2017). Their reception, apart from praise for the technical achievement, was not free of the uneasy feeling of uncanniness (cf. Brownlee, 2007).

The first finding is more remarkable: too human-like objects can not only illicit empathy, but produce a feeling of discomfort, spookiness (Ho, MacDorman, and Pramono, 2008). Seeing how androids are the most human-like robots, they ought to most often fall within the uncanny valley—until they become perfect human copies. But the valley is narrow, its edges are sharp and steep. An android, robot, even a human-like toy or doll ever so similar in appearance and behavior to humans, initially produces the experience of “cognitive emotional confusion,” and therefore discomfort, a “sensation of strangeness”—due to its “subtle imperfections:” “Machines that appear too lifelike would be unsettling or even frightening inasmuch as they resemble figures from nightmares or films about the living dead” (MacDorman, and Ishiguro, 2006b, pp. 300–301).³

On the other hand, it would appear that we have less of a problem with machines that are not so life- and human-like. Karl MacDorman

3 For the increasingly endless literature on the “uncanny valley,” and empathy towards beings we know not to be living yet resemble us (in some significant way), as well as on the revulsion and frisson they seemingly inevitably produce, and on a discussion about the (im)possibility of empathizing or at all understanding the fundamentally other yet similar, and the human ability and/or propensity to ascribe human intentionality and mental states to animate and inanimate, non-human entities, and questions less abstract, such as how to apply to the research and development of service robots that could interact with humans, among else, see Minato, Shimada, Itakura, Lee, and Ishiguro, 2005; Ramey, 2005a; Sober, and Wilson, 1998; Epley, Waytz, and Cacioppo, 2007; Geller, 2008; Cabibihan, Carrozza, Dario, Pattenfatto, Jomaa, and Benallal, 2006; Zia-ul-Haque, Wang and Jadoon, 2007; Thompson, 2019; Nishimura, Kazutaka, Dohi, Ishizuka, Takeuchi, and Tsujino, 2006; Becker-Asano, Ogawa, Nishio, and Ishiguro, 2010; Saygin, Chaminade, Ishiguro, Driver, and Frith, 2012 Pollick, 2009.

and Hiroshi Ishiguro (2006b) designate androids explicitly as “very humanlike robots,” as opposed to “mechanical-looking humanoid robots”. The ‘very humanlikeness’ of androids, they continue, is true for “robots that aspire to a degree of likeness [with humans] that transcends gross morphology:” “It is not enough for an android to have a head, two arms, a torso, and perhaps two legs. It must be humanlike down to the look and feel of the skin, teeth and hair. Its behavior should also approximate human behavior. People should be able to subconsciously respond to the android as human.” (MacDorman, and Ishiguro, 2006b, p. 298)

Moreover, the android is from this vantage point best defined as “an artificial system designed with the ultimate goal of being indistinguishable from humans in its external appearance and behavior” (MacDorman, and Ishiguro. 2006b, pp. 298-299). Not, that is, C3PO from *Star Wars*, nor the Tin Man from *The Wizard of Oz*, but at the very least the most famous android of them all: Terminator or else the not so coincidentally pejoratively called “skin jobs” from *Battlestar Galactica* or *Blade Runner*. However their uncanniness might impinge on our comfortably and once and for all categorized world, the ideal of contemporary robotics are “replicants” that would be biologically difficult to distinguish from humans (cf. Bodley, 2015, p. 127; Barringer, 1997), those synthetic humanoid simulacra that become identical, the perfect human copies—although only copies (MacDorman, and Ishiguro, 2006a). It is this category of android that is the topic of countless sci-fi films that express and disseminate the fear that the creation become like its creator (cf. LaGrandeur, 2015, pp. 112-114), that the roles of master and slave become reversed, for the words of the creation of Dr. Frankenstein to its creator to come true: “You are my creator, but I am your master;—obey” (Shelley, 1982: 165).

The Unpleasant Fatherhood

Just as we dealt with humanity's privilege over animals and other living "objects," we are faced with problems caused by our own creations, in much the same way as the "natural" entities, from which we have allegedly separated (Heise, 2009 pp. 504–505; cf. Wolfe, 2003; Loughnan, and Haslam, 2007). Our stance is again dictated by anxiety about our own uniqueness (cf. Smith, 2011). We repeat, reassuring ourselves: androids are creations, but they are not themselves creators nor self-creators like us; they are only humanoids, imitations of humans, only handmaidens. But what if they cease to be so, what if they "transcend" our intention and their purpose, and once more demonstrate to us that we do not have to be the measure of all things? What if automatons, like in the movie of that name, become what that word originally means: movement unto itself, of its own "volition?"

Never mind. We can accept that androids have replaced us in many jobs, or will soon do so, that they are undoubtedly better at many of them than us, but "humanist" pride says: it could only count towards our accomplishments. We made them, in our own image, and if they prove to be better than we are, is not our achievement and honor the greater? They could not, after all, produce us, and probably not even something else better still than themselves, as we did with them. The logic of origin, genesis, pedigree suffers no admonishment: the power of creation determines primacy.

Genetic engineering, so often said to be meddling in divine affairs, is nothing compared to the production of the artificial human: only

the latter is truly not only “outplaying” or “outwitting” God, but taking over the last remaining creative competence from Him. For this very reason, much as when having to do with God, we cannot stand the pride of creation. What is more, we ascribe in advance: as opposed to God who saw what he created and concluded it was good, man immediately considers that which is created in the human image the exact opposite. We are preemptively intolerant of the “viperous” idea that androids might pluck the fruit of the tree of knowledge, becoming like us. They are invented to be close to us, but not like us; or better, to be like us, but not us.

There is yet another reason, beyond pride, that we fear them becoming like us. This “becoming like one of us” (Genesis 3:22) we fear, could be taken more literally: they are our mirror (Ishiguro, 2014; Finn, 2018); their sci-fi projection is our social reflection, a reflection of a society in which we produce them either symbolically or quite literally (Gairola, 2017). For our part, we know ourselves well enough, and therefore fear that they might become equally mindless, aggressive, covetous of domination and triumphal extermination as we have been shown and shown up to be. Nor can we see how things might be otherwise with androids than they were and are with us; and truly, what means would we have at our disposal to bring them to reason that we have not already (and unsuccessfully) tested on ourselves?

And even if they by chance managed to become “civilized” on their own, things would be all the worse for us. At stake is not only our conceitedness, our alleged and allegedly threatened exceptionality and uniqueness, but the very real possibility that we prove inferior, including in the sense of correctability and perfectibility of which we are, again, so proud; we, the greatest *patrēs*, risk losing the privilege of paternalistic benevolence.

Thus, our ambivalence towards androids is ambivalence towards ourselves, our own self-understanding. They must, therefore, never be us; rather, they must in a sense be better than us, tamer, more obedient, less free, more attendant, and more tending towards assistance and slavery than us (Krstić, 2012). For the same reasons they also have to be forever worse than us, inhuman or humans stripped of free will and all trace of self-reflection.⁴ Otherwise, if they were indeed like us, who knows what would happen—to us. For now, luckily, such a scenario is not possible, except in sci-fi. There is, actually, a double anxiety associated with them: what we fear in some people, we are also afraid of in androids, except now on the level of species. On one side is the will to mastery and rebellion—perhaps a pure projection of our own course of events, or perhaps bad consciousness about our chronic injustice against “the other”—on the other, the very opposite: blind obedience, inadaptability. The latter is most often tied to fear of

4 Asimov’s famous Laws of Robotics are most revealing in this respect: “(1) A robot may not injure a human being or, through inaction, allow a human being to come to harm. (2) A robot must obey the orders given it by human beings except where such orders would conflict with the First Law. (3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Law” (Asimov, 1990, p. 8). James Gips notes that Asimov’s laudable effort in founding “robot-ethics,” which he undertook from the 1950s, if it can be reduced to something as “simple” as the Three Laws of Robotics, would also not be “suitable for our magnificent (free-ranging) robots: “These laws are for slaves. We want our robots to behave more like equals, more like ethical people” (Gips, 1995, p. 243). On the (im)possibility of ascribing moral status and/or the status of moral agent to artificially intelligent creations and humanoids, cf. Floridi, and Sanders, 2001; Moor, 2001; Ramey, 2005b; Gomel, 2011; Podschwadek, 2017; Misselhorn, 2018; Feil-Seifer, and Matarić, 2011; Himma, 2009.

excessive rationality, the soul-less sequence of artificial reasoning, lack of sensitivity to context, of blind and insensitive performance of an assigned task. This insight into terrifying (in-human?) rationality (cf. Lyotard, 1991; Schelde, 1993; Tsimpouki, Blatanis, and Dini, 2017; Chen, 2014) has already been well thematized in films, such as *2001: A Space Odyssey* (1968) and *Alien* (1979) (cf. Graham, 2002).

Inconclusive

Philosophical Conclusion

Throughout the twentieth century, (auto)reflection of the humanities and social sciences was an attempted balsam on the open wounds anthropocentric narcissism inflicted upon the environment, animals, foreigners, God, anything different and other, which could not be subsumed—except through anthropomorphization, identification with ourselves, meaning through violence—under the imperative of our universal, all-encompassing optics. We have recognized biological forms of life in their own right and particularity. After a long internal struggle, we placed ourselves among them, attempting to remove traces of the fantasy of our superiority. Our exceptionality, we have learned and on occasion even accepted, is but one exception among many exceptions, all of which have equal right to existence and wellbeing.

Yet, we are still reticent to recognize artificial life forms, “machines,” made of our own labor, and admit that they can exist as an independent life in its own right, and then they have a right to equal (co-)existence (Pordzik, 2012, pp. 144–145). In our mental structure, our vain

fantasies that still dictate our imagination, they are still a threat, surely because we project onto them our own tendency towards violence, as well as because they would deny us another source of conceited exceptionality.

Perhaps for this very reason, androids, rather than cyborgs (cf. Krstić, and Prodanović, 2013), might turn out to be the final act in humanity's, theoretical perhaps rather than practical, adventure. By which we do not mean anything fatal, but rather that humanity's Odyssey of self-understanding was marked by a strategic approach that would fit into the name of "anthropomorphization"—the projection of humanity onto everything different, an ever-same, ever-afraid, born of fear reaction of *reductio ad hominem*. Instead of being an underground cunning, with androids it could be directed and, at least as an auto-ironic gesture, testify to the potential of applying the same strategy to its own bearer, taking the form of de-mythologization and un-structuring, as a *reductio hominis* that would finally topple the "subject stylized to the absolute hilt" (cf. Adorno, 1997: 187).

There is only a chance, and not a good one, for that beneficial therapy; not a guarantee. One could, of course, also take a different route to naturalize the anthropomorphizing gesture than any kind of fundamental and grounding thought endeavor: not to question humanity's fatal strategies, but direct that tendency to opportunistically provide non-human entities human attributes. Robotics designers who make interactive robots, in particular service robots, do precisely that. The argument is not incoherent: by making them human-like, they create a more intimate connection in communication and assistance provided for users of medical or household help, and the dwelling space is already designed for movement and functioning of humans, so

mechanical emulation of it is the most practical option (cf. e.g., Ishiguro, and Asada, 2006: 74; Kiesler, Powers, Fussell, and Torrey, 2008: 177; MacDorman, and Ishiguro, 2006b: 299; Oyedele, Hong, and Minor, 2007: 631; Shaw-Garlock, 2011: 5; Bartneck, Kanda, Ishiguro, and Hagi-ta, 2009: 275; Waytz, Epley, and Cacioppo, 2010: 58).

Perhaps we will never be able to shed the anthropomorphizing gesture, shift to the other or perceive it without reducing it to us. Perhaps an element of anthropomorphization must remain even after all the critiques of it emerging again and again over the course of two and a half millennia. This might be an objective limit of both humanism and posthumanism. But it is one thing to meet and recognize, to think and test this limit, quite another to breach it.

In the case of “the android provocation,” anthropomorphism that fixes our “place in the cosmos,” taken commonsensically and synoptically, shines through both ends: as humanity is making a machine of itself (cyborgization), it also decided to render the machine more human (androidization). In the first case we improved ourselves, in the other, we rendered something potentially better than ourselves, a mechanical robot, too human, forcing it to imitate us. In the past, we assimilated both God and nature, rendered them in our own image, and now, perhaps with greater prerogative, we imagine that our products must forever to conform to our image and our “logic.” The only antidote we can propose to this universal human strategy in encounter with otherness, is persistent effort, however unachievable in its entirety, to retreat from the always already anthropomorphizing perspective: a kind of retreat that would allow other forms of life, including one or maybe specially the one that are our “products”, to have their own measures and their own lives.

This paper was realized with the support of the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, according to the Agreement on the realization and financing of scientific research.

References

- Adorno, T. W. (1997). *Negative Dialektik* [Negative Dialectics]. Theodor W. Adorno: Gesammelte Schriften, Bd. 6. Suhrkamp.
- Aguilera-Hellweg, M. (2017). *Humanoid*. Blast Books.
- Antosca, A. R. (2019). Technological re-enchantment: Transhumanism, techno-religion, and post-secular transcendence. *Humanities and Technology Review*, 38(2), 1-28.
- Asimov, I. (1990). *Robot visions*. Victor Gollancz.
- Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007). Is the uncanny valley an uncanny cliff? *Proceedings of the 16th IEEE, RO-MAN*, 368-373. <https://doi.org/10.1109/ROMAN.2007.4415111>
- Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2009). My robotic doppelgänger—A critical look at the uncanny valley. *18th IEEE International Symposium on Robot and Human Interactive Communication*, 269-276. <https://doi.org/10.1109/ROMAN.2009.5326351>
- Becker-Asano, C., Ogawa, K., Nishio, S., & Ishiguro, H. (2010). Exploring the uncanny valley with Geminoid HI-1 in a real-world application. *Proceedings of IADIS International Conference Interfaces and Human Computer Interaction*, 121-128.
- Bodley, A. M. (2015). The android and our cyborg selves: What androids will teach us about being (post)human. Dissertation, *Washington State University*. <https://www.proquest.com/openview/efb9bc71e395715daad-263787d96e618/1>
- Braga, A., & Logan, R. K. (2017). The emperor of strong AI has no clothes: Limits to artificial intelligence. *Information*, 8(4), 156-177. <https://doi.org/10.3390/info8040156>
- Brownlee, J. (2007, April). Professor Ishiguro's creepy robot doppelgänger. *Wired*. <https://www.wired.com/2007/04/professor-ishig/>

Cabibihan, J.-J., Carrozza, M. C., Dario, P., Pattofatto, S., Jomaa, M., & Benallal, A. (2006). The uncanny valley and the search for human skin-like materials for a prosthetic fingertip. *6th IEEE-RAS International Conference on Humanoid Robots*, 474–477. <https://doi.org/10.1109/ICHR.2006.321315>

Chen, A. (2014). Is artificial intelligence a threat? *The Chronicle of Higher Education*. <https://www.chronicle.com/article/is-artificial-intelligence-a-threat/>

Deutsche Welle. (2021). Robots that look like humans: The fantastic robots of Hiroshi Ishiguro—Erica, Geminoid, and Telenoid. *YouTube*. https://www.youtube.com/watch?v=AwMfNq1x_kQ

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886. <https://doi.org/10.1037/0033-295X.114.4.864>

Feil-Seifer, D., & Matarić, M. M. (2011). Ethical principles for socially assistive robotics. *IEEE Robotics and Automation Magazine*, 18(1), 24–31. <https://doi.org/10.1109/MRA.2010.940150>

Ferrando, F. (2014). Is the post-human a post-woman? Cyborgs, robots, artificial intelligence and the futures of gender: A case study. *European Journal of Futures Research*, 2(43), 1–17. <https://doi.org/10.1007/s40309-014-0043-8>

Finn, R. V. (2018). “More Human Than Human”: Lacan’s Mirror Stage Theory and Posthumanism in *Philip K. Dick’s Do Androids Dream of Electric Sheep?* (Doctoral dissertation). University of New Orleans. <https://scholarworks.uno.edu/cgi/viewcontent.cgi?article=3617&context=td>

Floridi, L., & Sanders, J. W. (2001). Artificial evil and the foundation of computer ethics. *Ethics and Information Technology*, 3(1), 55–66. <https://doi.org/10.1023/A:1011440125207>

Freud, S. (1986). *Das Unheimliche* [The uncanny]. In *Werke aus den Jahren 1917–1920. Gesammelte Werke*, Bd. 12 (pp. 229–268). Fischer.

Gairola, R. K. (2017). Android matters: Apocalyptic technology and Hegelian dystopia in Ridley Scott's *Blade Runner*. *Journal of the Humanities and the Social Sciences*, 4, 17–32. <http://researchrepository.murdoch.edu.au/id/eprint/51955>

Geller, T. (2008). Overcoming the uncanny valley. *IEEE Computer Graphics and Applications*, 28(4), 11–17. <https://doi.org/10.1109/MCG.2008.79>

Gips, J. (1995). Towards the ethical robot. In K. M. Ford, C. Glymour, & P. J. Hayes (Eds.), *Android epistemology* (pp. 243–252). American Association for Artificial Intelligence. <https://doi.org/10.1017/CBO9780511978036.019>

Gomel, E. (2011). Science (fiction) and posthuman ethics: Redefining the human. *The European Legacy*, 16(3), 339–354. <https://doi.org/10.1080/10848770.2011.575597>

Graham, E. (2002). *Representations of the post/human: Monsters, aliens, and others in popular culture*. Manchester University Press.

Heise, U. K. (2009). The android and the animal. *PMLA*, 124(2), 503–510. <https://doi.org/10.1632/pmla.2009.124.2.503>

Himma, K. E. (2009). Artificial agency, consciousness, and the criteria for moral agency: What properties must an artificial agent have to be a moral agent? *Ethics and Information Technology*, 11(1), 19–29. <https://doi.org/10.1007/s10676-008-9167-5>

Ho, C.-C., MacDorman, K. F., & Dwi Pramono, Z. A. (2008). Human emotion and the uncanny valley: A GLM, MDS, and ISOMAP analysis of robot video ratings. *Proceedings of the Third ACM/IEEE International Conference on Human-Robot Interaction*, 169–176. <https://doi.org/10.1145/1349822.1349845>

Ishiguro, H. (2014). Android philosophy. In J. Seibt, M. Nørskov, & R. Hakli (Eds.), *Sociable robots and the future of social relations: Proceedings of Robo-Philosophy* (pp. 3–12). IOS Press. <https://doi.org/10.3233/978-1-61499-480-0-3>

Ishiguro, H., & Asada, M. (2006). Humanoid and android science. *IEEE Intelligent Systems*, 21(4), 74–85. <https://doi.org/10.1109/MIS.2006.72>

- Jentsch, E. (1906). Zur Psychologie des Unheimlichen [The psychology of the uncanny]. *Psychiatrisch-Neurologische Wochenschrift*, 22, 203–205. <http://d-nb.info/1138447315/34>
- Kiesler, S., Powers, A., Fussell, S. R., & Torrey, C. (2008). Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition*, 26(2), 169–181. <https://doi.org/10.1521/soco.2008.26.2.169>
- Krstić, P. (2012). Granice poboljšanja: tri lika boljeg (od) čoveka [The better human, the better than human: Limits of enhancement]. *Filozofija i društvo*, 23(2), 124–144. <https://doi.org/10.2298/FID1202124K>
- Krstić, P., & Prodanović, S. (2013). Smurfs, cyborgs, and changelings: Prospects of human enhancement retrospected. *European Journal of Futures Research*, 1(1), 1–7. <https://doi.org/10.1007/s40309-013-0021-6>
- LaGrandeur, K. (2015). Androids and the posthuman in television and film. In M. Hauskeller, T. D. Philbeck, & C. D. Carbonell (Eds.), *The Palgrave handbook of posthumanism in film and television* (pp. 111–119). Palgrave Macmillan. https://doi.org/10.1057/9781137430328_12
- Lee, B. (2006). Empathy, androids, and “authentic experience.” *Connection Science*, 18(4), 419–428. <https://doi.org/10.1080/09540090600868847>
- Loughnan, S., & Haslam, N. (2007). Animals and androids: Implicit associations between social categories and nonhumans. *Psychological Science*, 18(2), 116–121. <https://doi.org/10.1111/j.1467-9280.2007.01858.x>
- Lyotard, J.-F. (1991). *The inhuman: Reflections on time*. Polity Press.
- MacDorman, K. (2005a). *Androids as experimental apparatus: Why is there an uncanny valley and can we exploit it?* CogSci-2005 workshop: Toward social mechanisms of android science. https://www.researchgate.net/publication/245406914_Androids_as_an_Experimental_Apparatus
- MacDorman, K. (2005b). Mortality salience and the uncanny valley. *Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots*, 399–405. <https://doi.org/10.1109/ICHR.2005.1573600>

- MacDorman, K. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *Toward social mechanisms of android science*. <https://www.researchgate.net/publication/241217609>
- MacDorman, K., & Ishiguro, H. (2006a). Toward social mechanisms of android science. *Interaction Studies*, 7(2), 289–296. <https://doi.org/10.1075/is.7.2.12mac>
- MacDorman, K., & Ishiguro, H. (2006b). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3), 297–337. <https://doi.org/10.1075/is.7.3.03mac>
- MacDorman, K., Minato, T., Shimada, M., Itakura, S., Cowley, S., & Ishiguro, H. (2005). Assessing human likeness by eye contact in an android testbed. *Proceedings of the XXVII Annual Meeting of the Cognitive Science Society*. <https://www.researchgate.net/publication/238193670>
- Minato, T., Shimada, M., Itakura, S., Lee, K., & Ishiguro, H. (2005). Does gaze reveal the human likeness of an android? *Proceedings of the 2005 4th IEEE International Conference on Development and Learning*, 106–111. <https://doi.org/10.1109/DEVLRN.2005.1490953>
- Misselhorn, C. (2010). Empathy and dyspathy with androids: Philosophical, fictional, and (neuro)psychological perspectives. *Konturen*, 2(1), 101–123. <https://doi.org/10.5399/uo/konturen.2.1.1341>
- Misselhorn, C. (2018). Artificial morality: Concepts, issues and challenges. *Society*, 55(2), 161–169. <https://doi.org/10.1007/s12115-018-0229-y>
- Moor, J. H. (2001). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems*, 21(4), 18–21. <https://doi.org/10.1109/MIS.2006.8>
- Mori, M. (1970). Bukimi no tani [The uncanny valley]. *Enajii*, 7(4), 33–35.
- Mori, M. (2005). Uncanny valley. *Proceedings of the Humanoids—2005 Workshop: Views of the Uncanny Valley*. <http://www.androidscience.com/theuncannyvalley/proceedings2005>

NBD. (2019). A dialogue with the world's most beautiful humanoid robot Erica. *YouTube*. <https://www.youtube.com/watch?v=lf9Up4e4600>

Neuwirth, M. (2018). "Absolute alterity"? The alien animal, the human alien, and the limits of posthumanism in Star Trek. *European Journal of American Studies*, 13(1), 1–20. <https://doi.org/10.4000/ejas.12464>

Nishimura, Y., Kazutaka, K., Dohi, H., Ishizuka, M., Takeuchi, J., & Tsujino, H. (2006). Mortality salience and the uncanny valley. *Proceedings of the 2005 5th IEEE-RAS International Conference on Humanoid Robots*, 393–398. <https://doi.org/10.1109/ICHR.2005.1573600>

Oyedele, A., Hong, S., & Minor, M. S. (2007). Contextual factors in the appearance of consumer robots: Exploratory assessment of perceived anxiety toward humanlike consumer robots. *CyberPsychology and Behavior*, 10(5), 624–632. <https://doi.org/10.1089/cpb.2007.9977>

Podschwadek, F. (2017). Do androids dream of normative endorsement? On the fallibility of artificial moral agents. *Artificial Intelligence and Law*, 25(3), 325–339. <https://doi.org/10.1007/s10506-017-9209-6>

Pollick, F. E. (2009). In search of the uncanny valley. In P. Daras & O. Mayora Ibarra (Eds.), *User centric media* (pp. 69–78). Springer. https://doi.org/10.1007/978-3-642-12630-7_8

Pordzik, R. (2012). The posthuman future of man: Anthropocentrism and the other of technology in Anglo-American science fiction. *Utopian Studies*, 23(1), 142–161. <https://doi.org/10.5325/utopianstudies.23.1.0142>

Ramey, C. H. (2005a). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. *Proceedings of the IEEE-RAS International Conference on Humanoid Robots: Views of the Uncanny Valley*, 8–13.

Ramey, C. H. (2005b). "For the sake of others": The "personal" ethics of human-android interaction. *Proceedings of the CogSci 2005 Workshop: Toward Social Mechanisms of Android Science*, 137–148.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive Affective Neuroscience*, 7(4), 413–422. <https://doi.org/10.1093/scan/nsr025>

Scheerer, R. (Director). (1989). *The Measure of a Man* [Star Trek: The Next Generation, episode 9, season 2]. DVD. Paramount, 2002.

Schelde, P. (1993). *Androids, humanoids, and other science fiction monsters: Science and soul in science fiction films*. New York University Press.

Sharp, S. (2007). Fembot feminism: The cyborg body and feminist discourse in *The Bionic Woman*. *Women's Studies*, 36(5), 507–523. <https://doi.org/10.1080/00497870701593739>

Shaw-Garlock, G. (2011). Loving machines: Theorizing human and social-technology interaction. In M. H. Lamers & F. J. Verbeek (Eds.), *Human-robot personal relationships* (pp. 1–10). Springer. https://doi.org/10.1007/978-3-642-19385-9_1

Shelley, M. W. (1982). *Frankenstein, or, the modern Prometheus*. University of Chicago Press.

Short, S. (2003). The measure of a man? Asimov's *Bicentennial Man*, *Star Trek's Data*, and being human. *Extrapolation*, 44(2), 209–224. <https://doi.org/10.3828/extr.2003.44.2.6>

Smith, D. L. (2011). *Less than human: Why we demean, enslave, and exterminate others*. St. Martin's Press.

Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Harvard University Press.

Thompson, T. (2019). Ghost stories from the uncanny valley: Androids, souls, and the future of being haunted. *Western Folklore*, 78(1), 39–65. <https://doi.org/10.14325/mississippi/9781496825087.003.0007>

Tsimpouki, T., Blatanis, K., & Dini, R. (Eds.). (2017). *War on the human: New responses to an ever-present debate*. Cambridge Scholars.

Waytz, A., Epley, N., & Cacioppo, J. T. (2010). Social cognition unbound: Insights into anthropomorphism and dehumanization. *Current Directions in Psychological Science*, 19(1), 58–62. <https://doi.org/10.1177/0963721409359302>

Wolfe, C. (2003). *Animal rites: American culture, the discourse of species, and posthumanist theory*. University of Chicago Press.

Zia-ul-Haque, Q. S. M., Wang, Z., & Jadoon, N. R. (2007). Investigating the uncanny valley and human desires for interactive robots. *Proceedings of the 2007 IEEE International Conference on Robotics and Biomimetics*, 2228–2233. <https://doi.org/10.1109/ROBIO.2007.4522516>

Art after AI: The Impact of Generative AI on the Artworld

Jelena
Novaković
Jelena Guga

Introduction

The rapid evolution of modern technology has transformed the art creation process, making it more accessible and efficient. However, the increased accessibility has also led to a challenge for artists – an overwhelming influx of information that affects societal perceptions of art and cognition (Krasteva, 2016). Throughout history, technology has been pivotal in expanding and transforming artists' creative horizons, providing them with powerful tools and new means of expression. Art forms once considered experimental or not even regarded as art – from kinetic art, and early computer art to digital art – have been gradually incorporated into traditional art systems and institutions (Guga, 2015; Grau, 2003; Paul, 2008). New media art, to use the umbrella term for various art forms that merge science, technology,

Jelena Novakovic, Research Associate, Institute for Philosophy and Social Theory, University of Belgrade, Belgrade ORCID: 0000-0001-7439-758X, email: jelena.novakovic@ifdt.bg.ac.rs

Jelena Guga, Research Fellow, Institute for Philosophy and Social Theory, University of Belgrade, Belgrade ORCID: 0000-0002-8180-3408, email: jelena.guga@ifdt.bg.ac.rs

DOI: <https://doi.org/10.5281/zenodo.14205155>

and art in the production, presentation, and distribution of artworks, has also established its history, festivals, and exhibitions, and it is now common in the curricula of art education worldwide. These art forms, including early AI art, have paved the way for the newly emerged AI Art's immediate presence in the art world.

While new technologies often evoke fears of displacement among traditional artists, they also give rise to new artistic styles, injecting vitality into art forms (Hertzmann, 2018). As artists become adept with new tools, they refine their techniques and broaden the scope of creative work and its outputs. At the same time, however, art has undergone the process of democratization as the wide accessibility and ease of use of these tools have turned everyone into artists. (Mazzone and Elgammal, 2019). This transformation has been evident across various technological shifts, from the Industrial Revolution to the Digital and AI Revolution. AI algorithms, particularly Generative AI, go beyond being mere tools as they take on the role of a creative medium. In the art world, the term “medium” encompasses instruments and the entire spectrum of possibilities and constraints in the creative process (Mazzone and Elgammal, 2019). The role of today's AI is not only that of a creative tool but also of a creator or co-creator of artworks. Hence, it is very important to explore and understand the impact of Generative AI on the democratization of creativity, traditional artistic roles, and the emergence of new artistic expressions.

Different AI Art Generator algorithms, such as DeepDream, DALL-E, Midjourney, or Stable Diffusion, allow users to input prompts, translating simple textual commands into complete visual artworks within minutes. Unlike traditional image combinations, these algorithms imagine words and convert them into visual reality by drawing on vast

datasets of paintings, artworks, and art history (Hanna, 2023). Various methodologies within AI art include Style Transfer, mimicking renowned artists' styles; Generative Adversarial Networks (GANs), generating entirely new art; and Interactive Systems fostering real-time collaboration between humans and machines (Yusa et al., 2022). However, distinguishing AI-generated art from human-created art poses challenges, with studies indicating a negative bias toward AI-generated works (Chamberlain et al., 2018; Hong et al., 2019; Ragot et al., 2020). A separate and very complex challenge related to AI and AI-generated works is that of intellectual property and copyrights over the data that is being used for training the AI systems and tools. In most cases, there is no real transparency or discrimination regarding the two and, as long as they are present in the public domain, the data, even if protected by copyright, is being used equally in AI-training platforms. This issue will not be discussed separately in this chapter, as its problematic is multifaceted and requires more in-depth consideration. Still, it should be flagged as one of the important points that needs to be carefully addressed in the process of AI-generated and collaborative art institutionalisation.

AI algorithms, considered a creative medium, encompass code, mathematics, hardware, software, and more. They bring about an unforeseen element of unpredictability, producing images that may diverge from the creator's initial intent (Coeckelbergh, 2023). This unpredictability highlights AI's role as a creative entity, going beyond a simple tool and actively participating in the creative process, challenging traditional definitions of creativity and authorship. AI generators have not only transformed the creative process but have also entered the business domain. Examples include DALL-E 2 generating the first AI-generated cover of *Cosmopolitan* magazine and brands incorporating AI

into advertising campaigns. Controversies surrounding AI interventions in recreating famous artworks underscore the ethical debates surrounding AI in the art world (Babbs, 2022; Taylor, 2023).

Integrating AI generators into the art world has sparked a revolution, prompting a reevaluation of traditional and contemporary artistic practices. This chapter examines the multifaceted impact of AI generators on democratizing art, the evolving landscape of artistic labor, and the collaborative potential between human creativity and AI. Additionally, it explores the ethical considerations that arise with the intersection of AI and artistic expression. While the use of AI generators democratizes creativity, it may lead to job displacement in certain artistic roles. Artists adept at leveraging AI as a collaborative enhancement tool are expected to foster innovation, giving rise to entirely new forms of artistic expression, and reshaping the art world's landscape. As AI has, as of recently, posed challenges to all spheres of life including education, business, and governance, the creative sector was at the forefront of its profound impact. Turning everyone into an artist, AI Art generators have brought fears and hopes to artistic production, questioning once again the role of the artist, and redefining creativity, authorship, and ownership.

Generative AI and Art

Current AI generators have emerged as transformative tools in the art world, challenging established norms and revolutionizing the creative process. This is why it is important to explore the profound impact of AI generators on the democratization of art, the potential

job displacement, and how artists can leverage this technology to enhance their creative processes. Beyond the concerns of job loss and professional obsolescence, this chapter delves into AI's exciting possibilities in fostering new forms of artistic expression. Artists, as well as others who are intrigued by the potential and boundaries of AI in the realm of creativity and the arts, are engaging with AI as a collaborative creative tool. This collaboration continues to evolve even beyond artistic practices, given that AI has already found its use in all spheres of life. In this creative partnership, the human and machine contribute their unique skills to the creative process (Mazzone and Elgammal, 2019). At the same time, artists from all over the world are fighting against the use of AI in art due to numerous and still unregulated issues including authorship, originality, and copyright (Farrant, 2023).

However, the resistance to technology in the art world is not new and AI has just reopened the existing issues from a new and more complex perspective. For instance, many viewed photography as non-artistic (Coeckelbergh, 2023), akin to AI, due to its mechanical nature. Critics questioned the legitimacy of photography, perceiving it as a threat to traditional art forms. The displacement of portraiture's social function by photography further fueled these concerns. Nevertheless, some artists embraced the new technology, exploring its potential for creative expression. According to Hertzmann (2018), photography played a significant role in the emergence of the Modern Art movement, as it inspired painters with its imagery and also compelled them to move beyond the confines of realism (Lin, 2005). This influence resulted in decades of creative dynamism within the world of painting. Moreover, photography became an art form in its own right.

Somewhat ironically, having in mind the historical “threat” of photography to traditional arts, associations of photographers, along with visual artists, filmmakers, and designers today are threatened by AI generators and are legally fighting the companies producing AI-based apps that use copyrighted works of artists worldwide without their consent and knowledge, and without having received any payments. A good example, among many, is the German art associations’ initiatives and recommendations in dealing with AI to ensure a common fair future for all stakeholders (Noack, 2023).

The animosity towards new technologies can be traced back to the notion of craft which was considered to be at the core of any artwork or artform. Every artist had to be, first and foremost, trained and skilled in their artistic practice of choice, in order for their individual artistic ideas and their respective executions to be accepted as relevant. A painter needs to show a masterful command over brush and paint, a sculptor over clay and marble, a musician over instruments and note-systems which they then use to produce works of art, in order to be recognised as a genuine artist. Traditional institutions of art do not recognise art in its idea only, without the necessary craft demonstrated in its execution. The current technology, employed to expedite and streamline creative processes, is inherently democratizing, enabling aspiring artists without conventional training to create traditional art forms such as paintings, music, sculpture, and literature. This necessitates a redefinition and modernization of the very concept of ‘craft,’ as technology reshapes and broadens the landscape of artistic production, challenging traditional norms and expanding the accessibility of creative expression. It is therefore necessary to reflect on the novel circumstances where, not the material of final artistic expression (i.e. image) defines the level of craft, but

the command over the actual process that drives the creativity, even if not anchored in any physical material per se.

In the same manner, the emergence of 3D computer animation was also met with apprehension from traditional artists, who feared it would replace their jobs. Computer animation has since, however, become a flourishing industry with a greater variety of job opportunities for animators. This narrative challenges the common belief that art and technology are incompatible, as it exemplifies how technology can stimulate the growth and advancement of art forms and has always been embedded into art practices either as a tool or a critical concept. Contemporary technology has, moreover, opened the door to radical immersion that art has ever been striving for. The integration of a perceiving body with the world of the artwork it is confronted with is now possible through different types of virtual realities, where the concept of frame, a traditional border of an artwork, is being deconstructed and the world of art becomes participatory, engaging and even saturating our cognitive capacities (Uspenski and Guga, 2022). The skill and the craft here are defined by the understanding of rules and language of technology, which allows for such immersive spaces to be created and to achieve the desired (or negotiated) esthetic experience.

Although generative AI has only recently received mass interest and attention, along with its creative artistic possibilities, its history dates back to mid 20th century and the beginnings of AI research and development. Artists, the early pioneers at the time, immediately caught on with the technology and experimented with AI and the potential of computers to create art. One of the earliest examples is Harold Cohen's AARON (1973), a computer program designed to create abstract

drawings and paintings. It was based on a set of rules and algorithms and was able to learn and adapt over time with the input of the creator. Other examples include the works by artists like Vera Molnar, Manfred Mohr, Naoko Tosa, Frieder Nake, David Cope, Peter Beyls, or Casey Reas. These early AI projects were limited in their capabilities as they were based on simple algorithms and were technically more simple, yet able to revolutionize art practice. Even though generative art does not necessarily involve AI and includes diverse, heterogeneous approaches, it has been often equated only with computational art practices or with AI art based on machine learning systems (Grba, 2021).

Following the “AI winter,” characterized by diminished funding and interest in the field, the emergence of neural networks and machine learning in the 1990s and 2000s presented artists with fresh opportunities to explore technology creatively. This allowed for the development of more intricate artworks that delved into the complexities of human-machine relationships, behavior, and the potential socio-cultural impacts in more nuanced ways. Artists such as Ken Feingold, Patrick Tresset, Stelarc, Ken Rinaldo, and John F. Simon, Jr. actively engaged in the creation of conceptually intriguing and thought-provoking AI art during this period.

With the development of computer vision, deep learning, and big data sets, AI art gained momentum and expansion, especially with the support of AI companies and academic and artistic programs focusing on AI. The recent boom of AI art is not merely tied to technology but is closely related to and emerges from a wider social, cultural, economic, and political context. Regardless of the medium of art production or presentation, the role of the artist involves commenting on social or cultural phenomena, bringing to the fore “historically and

locationally contextualized issues affecting societies, such as identity politics, marginalized social groups, ecology, and so on.” (Mevorah, Guga and Markov, 2023)

Yet, in the realm of burgeoning technologies and the accompanying excitement and apprehension, such as the current scenario with AI, there are artists whose works merely showcase the capabilities of the technology they employ as a medium, achieving little beyond that. Although visually captivating and awe-inspiring, like hyper-aesthetized large-scale immersive, interactive, or generative environments for instance, such artworks tend to overshadow the critical and poetic aspects of AI artworks. Discussing Refik Anadol’s productions, Grba (2022) characterizes such works as lacking critical perspectives on issues like mass surveillance, immaterial labor, environmental damage, and other problematic aspects associated with the big data capture and processing upon which they depend. Similarly, Zylinska (2020) observes that such artworks have “a pacifying effect, anesthetizing us into the perception of banal sameness that creates an illusion of diversification without being able to account for *differences that matter*” (italics in original). Serving as a kind of digital decoration with the purpose of entertainment and special effects, this type of art “really is art as spectacle” (Zylinska, 2020).

Then again, other artists critically employ AI in their artistic practice to address various socio-cultural, ecological, and other effects and issues arising from our entanglements and coexistence with AI. Artists like Trevor Paglen, Anna Ridler, Egor Kraft, Vladan Joler, Jake Elwes, Ben Snell, Heather Dewey-Hagborg, and many more are among those not sacrificing their experimental, creative, and critical edge in the face of art commodification. Their works explore the complexities

of human-AI relationships, shedding light on issues such as bias in algorithms, surveillance, and the impact of AI on societal structures. Critical AI artists strive to foster a deeper understanding of the implications of technology, encouraging viewers to question its role in shaping our perceptions and influencing our daily lives. This collective exploration showcases the multidimensional nature of AI art, with artists contributing distinct insights and reflections on the complex interplay between technology, society, and the environment.

As evident throughout the history of art, every new technology has been embraced and incorporated into art production. Even when technology is not the art medium, it is included critically or conceptually as the artists comment on our world through their art. The creative process encompasses a range of intellectual, social, emotional, and individual factors together with artistic abilities. Contemporary artists who employ new technologies in their work must therefore consider not only the technological platform and the opportunities it provides for exploring new directions but also the complex challenges presented by the ongoing diversification of technological advances, both in the art world and societies, as technology is primarily a social and cultural phenomenon. The aim of such artists is to engage with technology in a way that produces innovative and meaningful work, rather than simply using it as a tool for aesthetic expression. As such, the successful integration of technology into the artistic process requires a nuanced understanding of the interplay between technology, creativity, and artistic expression (Krasteva, 2016). As Grba (2021) writes, “[artists] consciously use AI systems to play with the variable abstraction of authorship vis-à-vis technology, to explore the notions of agency and learning in the creative process, or to accentuate the uncanny appearance of artificial entities.”

While algorithms and computer programs may facilitate the actual production of the artwork, it is ultimately the result of considerable time and effort invested by human artists. From the conception of the idea to the meticulous guidance of the execution, to the selection of the best output among the many generated by the machine, the human artist plays an essential role in the creative process. It is this combination of human and machine input that makes computer-generated art a unique and exciting medium, pushing the boundaries of what is possible in traditional art forms (Hertzmann, 2018), but this progress in AI-generated artwork inherently complicates modern comprehension of creativity and the concept of aesthetic beauty in the realm of art (Hong and Curran, 2019).

Democratization of Art and the Future of Artistic Labor

AI generators have played a pivotal role in democratizing art, dismantling traditional barriers to artistic production, and opening the creative realm to wider audiences. With the access to the right tools, anyone can explore their artistic potential. Even with basic technical skills, individuals can now produce artwork in a matter of minutes, by developing and applying different type of artistic craft, that rivals the quality of pieces created through hours of manual labor by professional artists. A great part of AI's democratic nature lies in its accessibility through relatively low cost of the material used for the artwork production (access to the software and electricity) vs. traditional art-forms (i.e. painting), rather than in traditional artistic training. This

accessibility has the potential to redefine the exclusivity traditionally associated with artistic talent, welcoming a more diverse range of voices and perspectives into the artistic discourse. The democratization of art through AI art generators challenges the notion that artistic brilliance is the privilege of a selected few. It fosters a more inclusive and diverse art community, enabling individuals from various backgrounds to express and develop new forms of artistic craft and participate in the creation of art. This shift has reshaped the dynamics of the art world, promoting a more egalitarian approach to creativity, regardless of the superficiality of exhaustive algorithmic variations that reduce our perception to visual consumption and art to mild bemusement (Zylinska, 2020).

In 2022, the same year AI generators were launched, Jason Allen, credited through the Midjourney platform, received first place in the annual state fair art competition held in Colorado with his piece *Théâtre D'opéra Spatial*. The artist considers AI image generators as an organic extension of his creative process, having spent weeks fine-tuning parameters while producing over 100 images for this artwork. Following the selection of his preferred image, Allen continued the artistic process in Adobe Photoshop and completed it by applying post-processing techniques with Gigapixel AI (Batycka, 2022). This process of selecting tools and inputs, adjusting settings, modifying code, and iterating until a desirable output is produced is seen as fundamental to the creation of all current and foreseeable computer artworks (Hertzmann, 2018).

AI-generated images began surfacing across various creative domains, prompting reactions from artists globally, primarily initiated on social media platforms. On one side, they were defending AI

technology, saying that creativity is still required to come up with the right prompts to generate artwork and that the AI image generators are not different from using Photoshop or other software, while others saw Allen's award as the death of artistry (Roose, 2022). Even artists like Refik Anadol, commented that generating images through verbal prompts can be exciting but does not qualify as true art-making (Smee, 2023). Numerous artists have shared "Do Not AI" signs across their social media profiles and online art collections. Investigating this growing issue reveals that designers and illustrators are particularly at risk. ArtStation, a prominent platform, faced such a surge of anti-AI memes that it sought the labeling of AI-generated artwork (Thorpe, 2023). In image generators like Stable Diffusion, Midjourney, and Dall-E people often use artists' names in prompts to make the generated images look better, not only the ones whose art can be used under the Creative Commons license but also any artist alive, since the AI has learned from the existing datasets (Wiggers, 2023; Thorpe, 2023). A group of artists has initiated a class action lawsuit against various companies, including Stability AI, DeviantArt, Midjourney, and Open AI. This legal proceeding, based on allegations of copyright infringement (Choi, 2023; Wiggers, 2023), underscores the capability of generating remarkably accurate imitations, often encompassing billions of images, including copyrighted works by contemporary artists (Russel, 2023).

While generative AI software usually imitates an artist's unique style instead of directly copying their artwork, it is often challenging to tell the difference between machine-generated and genuine pieces, especially when considering a potential purchase (Saheb et al., 2023). In a bid to empower artists with greater authority over their creations, the website HavelBeenTrained provides an option for creators to

exclude their work from the training dataset of the Stable Diffusion art-generating AI model. Leveraging this capability, artists have proactively employed it, leading to the elimination of around one billion artworks and persisting challenges within the digital visual art landscape in the United States. These challenges undermine creators' once-exclusive rights to their published digital art. Although generative AI may not explicitly violate the current U.S. copyright law (Choi, 2023), there are contentions that it extracts and exploits artists' past and future works.

While the democratization of art is a commendable aspect of AI generators, concerns arise regarding the potential displacement of existing artistic jobs. As AI becomes increasingly proficient at generating art, there is a legitimate fear that certain roles within the art industry may diminish. Tasks such as digital rendering, repetitive design work, or even certain aspects of traditional craftsmanship could be automated, reducing the demand for human labor. This raises questions about the future of employment in the art world and the need for artists to adapt to this changing landscape. However, it is crucial to view AI not merely as a threat to traditional roles but as a transformative force that can redefine the nature of artistic work.

In our current post-digital era, characterized by the prevalence of AI algorithms, artists are not exempt from its influence. In a study (Mevorah, Guga and Markov, 2023) examining artists' perspectives on digital technologies, it was found that irrespective of their inclination to adopt these technologies in their artistic endeavors or their resistance to them, all artists recognize the social and cultural influence of digital technologies. They actively acknowledge and interact with these impacts, even when utilizing the most traditional forms of

artistic expression. Rather than viewing AI as a replacement for human creativity, AI can serve as a powerful tool for artists to augment their creative processes. Artists adept at integrating AI into their workflow can leverage its capabilities to explore new possibilities, experiment with unconventional ideas, and push the boundaries of traditional artistic practices. For instance, AI algorithms can analyze vast datasets of art history, styles, and techniques, providing artists with valuable insights and inspiration. It can generate initial concepts or assist in the tedious aspects of the creative process, allowing artists to focus on more intricate and emotionally charged elements of their work. In this way, AI becomes a collaborator rather than a competitor, enabling artists to critically and creatively evolve and innovate.

Human-AI Collaboration and Ethical Considerations

The synergy between human creativity and AI has the potential to give rise to entirely new forms of art. Artists are experimenting with generative algorithms to create pieces that defy conventional categorizations. The blending of human intuition with machine-generated randomness can result in artworks that challenge preconceptions and expand the boundaries of artistic expression. Interactive installations, virtual reality experiences, and multimedia collaborations are just a few examples of how AI contributes to the evolution of art forms. The dynamic interplay between artists and AI opens up uncharted territories, fostering a fertile ground for creativity and experimentation.

AI generators act as catalysts for innovation in the art world, pushing the boundaries of what is possible. Artists who embrace this technology find themselves at the forefront of a new wave of creativity, exploring uncharted territories and pushing the limits of traditional artistic practices. The ability of AI to process vast amounts of data and generate novel ideas allows artists to break free from creative constraints. This liberation from conventionality encourages exploration, enabling artists to create works that challenge societal norms, provoke thought, and evoke emotional responses. For that, a new type of craft is needed. That, which in writing prompts recognises, almost intuitively and with artistic talent, the fine-line of communication between a man and a machine, such that it gives optimal and surprising creative results.

As the art world embraces AI, ethical considerations come to the forefront. Questions about authorship, originality, copyright, and the impact on the artistic community arise. How do we attribute creative works when AI is involved in the process? Does the democratization of art through AI generators diminish the value of artistic training and craftsmanship? How are human-driven characteristics of artmaking such as the urge to create, conceptualization, motivation, skill, style, and anticipation affected by the use of generative AI? Who is this overproduction of images for and who profits from it? How can we ensure fair use of AI for all the stakeholders? These ethical questions underscore the need for a thoughtful and considered approach to the integration of AI in the art world. Artists, technologists, and policy-makers must collaborate to establish frameworks that ensure the responsible and ethical use of AI in artistic creation.

Conclusion

AI generators are undeniably reshaping the art world, presenting both challenges and opportunities. Unquestionably, like any other practice, artistic or otherwise, they require institutionalisation through critical thinking to ensure that creativity, which is the essence of any art, profits from them in a meaningful and productive way. The democratization of art through AI is breaking down traditional barriers, making creativity accessible to all, and opening the doors for new types of artists and new types of more accessible craft training. Concerns about job displacement highlight the need for artists to adapt and view AI as a collaborative tool rather than a threat. Moreover, the collaboration between human creativity and AI has the potential to foster unprecedented innovation, giving rise to new forms of artistic expression. Artists who embrace AI as a catalyst for creativity and critical thinking find themselves at the forefront of a transformative wave, pushing the boundaries of traditional practices and redefining the nature of artistic work. As we navigate this evolving landscape, it is crucial to strike a balance that reassesses and reflects on human creativity while leveraging the benefits that AI has to offer. The intersection of art and technology presents an exciting frontier where the synthesis of human intuition and machine capabilities can lead to a renaissance in artistic expression, offering a glimpse into the possibilities and limitations that lie ahead for the art world.

This paper was realized with the support of the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, according to the Agreement on the realization and financing of scientific research.

References

- Aničin, L., & Stojmenović, M. (2023). Bias analysis in stable diffusion and MidJourney models. In N. M. S. Nandan, V. Garcia Diaz, & G. A. E. Satish Kumar (Eds.), *Intelligent systems and machine learning. ICISML 2022*. Lecture notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (Vol. 471). Springer. https://doi.org/10.1007/978-3-031-35081-8_32
- Babbs, V. (2022, November 18). How is artificial intelligence changing art history? *Hyperallergic*. <https://hyperallergic.com/744100/how-is-artificial-intelligence-changing-art-history/>
- Batycka, D. (2022, September 1). An A.I.-generated artwork won first prize at a Colorado State Fair. Human artists are infuriated. *Artnet News, Science and Tech*. <https://news.artnet.com/art-world/colorado-art-ists-mad-ai-art-competition-2168495>
- Carrasco Barranco, M. (2022). Artistic beauty in the face of artificial intelligence art. In I. Vidmar Jovanović & V. M. Stupnik (Eds.), *Social and technological aspects of art: Challenges of the 'new normal'*. University of Rijeka.
- Chamberlain, R., Mullin, C., Scheerlinck, B., & Wagemans, J. (2018). Putting the art in artificial: Aesthetic responses to computer-generated art. *Psychology of Aesthetics, Creativity, and the Arts*, 12(2), 177–192. <https://doi.org/10.1037/aca0000136>
- Choi, E. (2023). Protecting visual artists from generative AI: An interdisciplinary perspective. *1st ICML Workshop on Generative AI and Law*.
- Russell, A. (2022, November 12). When AI can make art—what does it mean for creativity? *The Guardian*. <https://www.theguardian.com/technology/2022/nov/12/when-ai-can-make-art-what-does-it-mean-for-creativity-dall-e-midjourney>
- Coeckelbergh, M. (2023). The work of art in the age of AI image generation: Aesthetics and human-technology relations as process and performance. *Journal of Human-Technology Relations*, 1(1), 1-13. <https://doi.org/10.59490/jhtr.2023.1.7025>

- Farrant, T., & AFP. (2023, March 27). From lawsuits to tech hacks: Here's how artists are fighting back against AI image generation. *Euronews.culture*. <https://www.euronews.com/culture/2023/03/27/from-lawsuits-to-tech-hacks-heres-how-artists-are-fighting-back-against-ai-image-generatio>
- Grau, O. (2003). *Virtual art: From illusion to immersion*. MIT Press. <https://doi.org/10.7551/mitpress/7104.001.0001>
- Grba, D. (2021). Brittle opacity: Ambiguities of the creative AI. *xCoAx, 9th International Conference on Computation, Communication, Aesthetics and X Proceedings*. <https://github.com/dejangrba/brittle-opacity>
- Grba, D. (2022). Deep else: A critical framework for AI art. *Digital*, 2(1), 1–32. <https://doi.org/10.3390/digital2010001>
- Guga, J. (2015). *Digital self: How we became binary*. University of West Bohemia.
- Hertzmann, A. (2018). Can computers create art? *Arts*, 7(2), 18. <https://doi.org/10.3390/arts7020018>
- Hong, J.-W., & Curran, N. M. (2019). Artificial intelligence, artists, and art. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 15(2), 1–16. <https://doi.org/10.1145/3326337>
- Krasteva, M. (2016). The impact of technology on the modern art. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, 6, 247–254. <https://doi.org/10.55630/dipp.2016.6.26>
- Lin, P.-H. (2005). A dream of digital art: Beyond the myth of contemporary computer technology in visual arts. *Visual Arts Research*, 31(1), 4–12. <http://www.jstor.org/stable/20715364>
- Liu, J. (2022, October 19). What if Van Gogh could paint your hometown? *Hyperallergic*. <https://hyperallergic.com/760728/what-if-van-gogh-could-paint-your-hometown/>
- Mazzone, M., & Elgammal, A. (2019). Art, creativity, and the potential of artificial intelligence. *Arts*, 8(1), 26. <https://doi.org/10.3390/arts801002>

Mevorah, V., Guga, J., & Markov, Č. (2023). Divine genius, subversive hero, or creative entrepreneur? Exploring various facets of the artist as a mythical figure. *Glasnik Etnografskog instituta SANU*, 71(1), 99–121. <https://doi.org/10.2298/GEI2301099M>

Noack, M. (2023). Art, AI, and visual authors' rights: Copyright recommendations for a fair future for visual artists—German initiatives and their recommendations. In *IAA Europe Symposium: Arts and Robots* [Video]. Belgrade. <https://www.youtube.com/watch?v=YXdCI8B7ifk>

Paul, C. (2008). *Digital art*. Thames & Hudson.

Ragot, M., Martin, N., & Cojean, S. (2020, April). AI-generated vs. human artworks: A perception bias toward artificial intelligence? *CHI '20: CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3334480.3382892>

Roose, K. (2022, September 2). An A.I.-generated picture won an art prize. Artists aren't happy. *The New York Times*. <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>

Saheb, S. T., Karimi, S. S., Esposito, M., & Tse, T. (2023, May 30). How to protect artists from generative artificial intelligence. *LSE Business Review*. <https://blogs.lse.ac.uk/businessreview/2023/05/30/how-to-protect-artists-from-generative-artificial-intelligence/>

Smee, S. (2023, February 15). AI is no threat to traditional artists. But it is thrilling. *The Washington Post*. <https://www.washingtonpost.com/arts-entertainment/2023/02/15/ai-in-art/>

Taylor, M. (2023, March 1). Museum under fire for showing AI version of Vermeer masterpiece. *Hyperallergic*. <https://hyperallergic.com/805030/mauritshuis-museum-under-fire-for-showing-ai-version-of-vermeer-masterpiece/>

Thorpe, V. (2023, March 18). ChatGPT said I did not exist: How artists and writers are fighting back against AI. *The Guardian*. <https://www.theguardian.com/technology/2023/mar/18/chatgpt-said-i-did-not-exist-how-artists-and-writers-are-fighting-back-against-ai>

Uspenski, I., & Guga, J. (2022). Embodying metaverse as artificial life: At the intersection of media and 4E cognition theories. *Philosophy and Society*, 33(2), 326–345. <https://doi.org/10.2298/FID2202326U>

Wiggers, K. (2023, May 3). Spawning lays out plans for letting creators opt out of generative AI training. *TechCrunch*. <https://techcrunch.com/2023/05/03/spawning-lays-out-its-plans-for-letting-creators-opt-out-of-generative-ai-training/>

Yusa, I. M. M., Yu, Y., & Sovhyra, T. (2022). Reflections on the use of artificial intelligence in works of art. *Journal of Aesthetics, Design, and Art Management*, 2(2), 152–167. <https://doi.org/10.58982/jadam.v2i2.334>

Zylinska, J. (2020). *AI art: Machine visions and warped dreams*. Open Humanities Press.

book series

DISPUT

1. Ivana Spasić, Milan Subotić (eds.), *R/evolucija i poredak: o dinamici promena u Srbiji*
2. Ivana Spasić, Milan Subotić (eds.), *R/evolution and Order: Serbia after October 2000*
3. Vladimir N. Cvetković (eds.), *(Re)konstrukcija institucija: godinu dana tranzicije u Srbiji*
4. Zagorka Golubović, Ivana Spasić, Đorđe Pavićević (eds.), *Politika i svakodnevni život: Srbija 1999–2002.*
5. Mile Savić (eds.), *Integracija i tradicija. Integration and Tradition*
6. Božidar Jakšić, Goran Bašić, *Umetnost preživljavanja: gde i kako žive Romi u Srbiji*
7. Petar Bojanić (ed.), *Glas i pismo: Žak Derida u odjecima*
8. Miloš Nemanjić, Ivana Spasić (eds.), *Nasleđe Pjera Burdijea: pouke i nadahnuća*
9. Rada Drezgić, Daša Duhaček, Jelena Vasiljević (eds.), *Ekofeminizam: nova politička odgovornost*
10. Ana Birešev, Gazela Pudar, Dušan Bošković (eds.), *Priroda, etika, politika: ekološke (pre)okupacije i (pre)orijentacije*
11. Rada Drezgić, Željko Radinković, Predrag Krstić (eds.), *Horizont bioetike: moral u doba tehničke reprodukcije života*
12. Mirko Blagojević, Jelena Jablanov Maksimović, Tijana Bajović (eds.), *(Post)sekularni obrt: religijske, moralne i društveno-političke vrednosti studenata u Srbiji*
13. Holger Zaborowski, Željko Radinković, Rastko Jovanov (eds.), *Phänomenologische Ontologie des Sozialen*

14. Željko Radinković, Jan Müller, Rastko Jovanov (eds.), *Politiken des Lebens: Technik, Moral und Recht als institutionelle Gestalten der menschlichen Lebensform*
15. Dušan Bošković, Aleksandar Petrović (eds.), *Povratak iz zemlje zmajeva: Rečnik tehnologije – 33 godine posle*
16. Mark Lošonc, Predrag Krstić (eds.), *Holokaust i filozofija*
17. Gazela Pudar Draško, Aleksandar Pavlović, Elton Meka, *Politics of Enmity*
18. Đurđa Trajković, Srđan Prodanović, Gazela Pudar Draško (eds.), *Urges For Engagement: Conditions of Social Change*
19. Srđan Prodanović, Gazela Pudar Draško, Marija Velinov (eds.), *Otete institucije u Srbiji: teorija i praksa*
20. Marjan Ivković, Srđan Prodanović (eds.), *Engaging (For) Social Change: Towards New Forms of Collective Action*
21. Vera Mevorah, Željko Šarić, Predrag Krstić (eds.), *Holokaust i teologija*
22. Marjan Ivković, Željko Radinković (eds.), *Priručnik kritike*
23. Ljubiša Bojić, Simona Žikić, Jörg Matthes, Damian Trilling (eds.), *Navigating the Digital Age: An In-Depth Exploration into the Intersection of Modern Technologies and Societal Transformation*

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

316.77:004(082)

351.78:316.422(082)

7:004.8(082)

NAVIGATING the digital age : an in-depth exploration into the intersection of modern technologies and societal transformation / [edited by Ljubiša Bojić ... [et al.]]. - Belgrade : Institute for Philosophy and Social Theory; Vienna : Department of Communication, Faculty of Social Sciences, University of Vienna, 2024 (Belgrade : Alta Nova). - 606 str. : ilustr. ; 21 cm. - (Book series Disput)

Tiraž 300. - Str. 11-17: Preface / Ljubiša Bojić ... [et al.]. - Napomene i bibliografske reference uz tekst. - Bibliografija uz svaki rad.

ISBN 978-86-82324-85-0

а) Информациона технологија -- Друштвени аспект -- Зборници
б) Безбедност -- Друштвени развој -- Зборници в) Уметност
-- Вештачка интелигенција -- Зборници г) Масовни медији --
Зборници

COBISS.SR-ID 157595145

