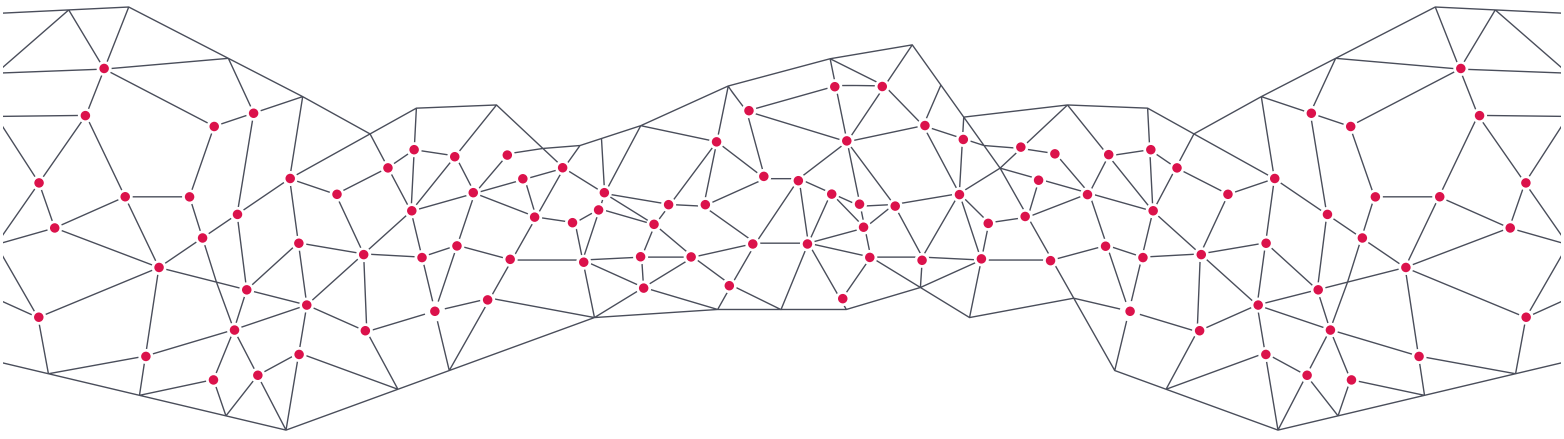
INTERNATIONAL CONFERENCE

EMERGE 2026

CONTESTED FUTURES

BOOK OF ABSTRACTS



INSTITUT
ZA FILOZOFIJU
I DRUŠTVENU
TEORIJU

International Conference

EMERGE 2026: CONTESTED FUTURES

BOOK OF ABSTRACTS

Organizers
Digital Society Lab, Institute for Philosophy and Social Theory,
University of Belgrade
The Institute for Artificial Intelligence Research and Development of Serbia

Editors
Jelena Guga
Zoran Erić
David Adam

Design
Jelena Novaković

Publisher
Institute for Philosophy and Social Theory, University of Belgrade

ISBN 978-86-6190-058-7

The Book of Abstracts 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 451-03-33/2026-03/ 200025.

Belgrade, 2026

International Conference

EMERGE 2026
CONTESTED FUTURES

November 5–6, 2026

Institute for Philosophy and Social Theory,
University of Belgrade

CONTENTS

EMERGE 2026: Contested Futures Committees 1
Introduction 3
Keynote Lectures 4
Agency, Resistance, and Subjectivity in the Age of AI 10
AI Ethics, Justice, and Social Inequality 23
Art, Culture, and Critical AI Practices 40
Digital Democracy, Governance, and Technological Power 52
Education, AI-Assisted Learning, and Digital Literacy 75
Environment, Extraction, Sustainability, and Digital Degrowth 92
Human-Machine Communication: Power, Design, and Human-AI Relations 98
Media and Communication: Platforms, Algorithms, and Technological Imaginaries 101
Speculative and Alternative Technological Futures 106
Synthetic Research: Methods, Risks, and Epistemic Challenges 109

EMERGE 2026: CONTESTED FUTURES
COMMITTEES

Steering Committee

Ljubiša Bojić (Chair), Institute for Philosophy and Social Theory, University of Belgrade and Institute for Artificial Intelligence Research and Development of Serbia
Jelena Guga, Institute for Philosophy and Social Theory, University of Belgrade
Simona Žikić, Institute for Philosophy and Social Theory, University of Belgrade and Faculty of Media and Communications, Singidunum University
Čedomir Markov, Institute for Philosophy and Social Theory, University of Belgrade
Jörg Matthes, Department of Communication, University of Vienna
Achim Rettinger, Knowledge Representation Learning (krAil) Research Group, Trier University
Damian Trilling, Faculty of Social and Behavioral Sciences, Vrije Universiteit Amsterdam
Sedat Uçar, Faculty of Education, Çukurova University
Srđan Dušanić, Faculty of Philosophy, University of Banja Luka
Ernest Ženko, Department of Media Studies, University of Primorska
Mark Coeckelbergh, Department of Philosophy, University of Vienna
Marina Dabić, Faculty of Economics and Business, University of Zagreb
James O'Higgins Norman, DCU Institute of Education, Dublin City University
Amalee Meehan, School of Human Development, Dublin City University
Alexander Felfernig, Institute of Software Engineering and Artificial Intelligence, Graz University of Technology
Mihaela Rus, Faculty of Law and Administrative Sciences, Ovidius University of Constanța

Scientific Committee

Jelena Guga (Chair), Institute for Philosophy and Social Theory, University of Belgrade
Ljubiša Bojić, Institute for Philosophy and Social Theory, University of Belgrade and Institute for Artificial Intelligence Research and Development of Serbia
Simona Žikić, Institute for Philosophy and Social Theory, University of Belgrade and Faculty of Media and Communications, Singidunum University
Čedomir Markov, Institute for Philosophy and Social Theory, University of Belgrade
Željko Radinković, Institute for Philosophy and Social Theory, University of Belgrade
Zoran Erić, Institute for Philosophy and Social Theory, University of Belgrade
Vera Mevorah, Institute for Philosophy and Social Theory, University of Belgrade
Ivana Krtolica, Institute for Artificial Intelligence Research and Development of Serbia
Veljko Prodanović, Institute for Artificial Intelligence Research and Development of Serbia

Sanja Bojanić, Center for Advanced Studies Southeast Europe, University of Rijeka
Emanuele Arielli, Department of Design Culture, Iuav University of Venice
Aleksandra Kanjuo-Mrčela, Faculty of Social Sciences, University of Ljubljana
Achim Rettinger, Knowledge Representation Learning (krAil) Research Group, Trier University
Damian Trilling, Faculty of Social and Behavioral Sciences, Vrije Universiteit Amsterdam
Jörg Matthes, Department of Communication, University of Vienna
Matteo Cinelli, Department of Computer Science, Sapienza University of Rome
James O'Higgins Norman, DCU Institute of Education, Dublin City University
Amalee Meehan, School of Human Development, Dublin City University
Ana Martinoli, Faculty of Dramatic Arts, University of Arts in Belgrade
Bojana Romić, School of Arts and Communication (K3), Malmö University
Luka Perušić, Faculty of Philosophy, University of Zagreb
Tamara Kunić, VERN' University
Aleksandra Krstić, Faculty of Political Science, University of Belgrade
Zorana Jolić Marjanović, Faculty of Philosophy, University of Belgrade
Bojana Dinić, Faculty of Philosophy, University of Novi Sad
Damien Guzek, Faculty of Social Sciences, University of Silesia in Katowice
Florian Röhrbein, Faculty of Computer Science, Chemnitz University of Technology
Aleksandar Tomašević, Institute of Physics Belgrade, University of Belgrade
Irena Stojković, Faculty of Special Education and Rehabilitation, University of Belgrade
Milan Čabarkapa, Faculty of Engineering Sciences, University of Kragujevac
Boris Delibašić, Faculty of Organizational Sciences, University of Belgrade
Olga Zagovora, Department of Psychology, RPTU University Kaiserslautern-Landau
Vaitsa Giannouli, Department of Psychology, Democritus University of Thrace

Organizing Committee

Zoran Erić (Chair), Institute for Philosophy and Social Theory, University of Belgrade
Jelena Novaković, Institute for Philosophy and Social Theory, University of Belgrade
Simona Žikić, Institute for Philosophy and Social Theory, University of Belgrade and Faculty of Media and Communications, Singidunum University
Tijana Uzelac, Institute for Philosophy and Social Theory, University of Belgrade
Jelena Guga, Institute for Philosophy and Social Theory, University of Belgrade
Ljubiša Bojić, Institute for Philosophy and Social Theory, University of Belgrade and Institute for Artificial Intelligence Research and Development of Serbia
David Adam, Institute for Philosophy and Social Theory, University of Belgrade

INTRODUCTION
EMERGE 2026: Contested Futures

EMERGE 2026: Contested Futures brings together interdisciplinary perspectives on the political, social, cultural, ecological, and epistemic stakes of contemporary technological change. The conference starts from a simple premise: technological futures are not given. They are made, negotiated, and contested, and they can be made differently.

The 2026 edition takes place at a moment when AI systems and digital infrastructures have become increasingly central to the organization of economic power, political governance, social sorting, communication, education, cultural production, and everyday life. At the same time, democratic institutions struggle to keep pace with technological change, social inequalities are reproduced through digital systems, and the environmental costs of computational infrastructures continue to grow. Rather than approaching technology as inevitable or neutral, EMERGE 2026 focuses on the ways in which emerging technologies are developed, governed, narrated, experienced, and resisted.

The contributions gathered in this Book of Abstracts examine how technological futures are shaped through struggles over labor, resources, values, knowledge, representation, and social organization. They ask how AI and emerging technologies redistribute power and reshape public life, what forms of inequality, exclusion, and extraction they reproduce or obscure, and how dominant technological imaginaries are legitimized and contested across media, culture, platforms, institutions, and everyday practices.

A further focus is placed on emerging methods and forms of knowledge production, including synthetic research and the growing use of AI-generated data, personas, and simulations to model human behavior. These developments raise questions about methodological validity, epistemic authority, accountability, and the consequences of using synthetic outputs in research and decision-making. Alongside these concerns, the conference foregrounds critical, speculative, artistic, and practice-based approaches that seek to make visible the infrastructures and power relations underlying contemporary technologies and to imagine alternatives to dominant technological paradigms.

Across different disciplinary perspectives, the contributions engage with digital democracy and governance, AI ethics and social inequality, environmental extraction and digital degrowth, art and critical AI practices, agency and resistance, education and digital literacy, media and communication, synthetic research, human-machine communication, and speculative technological futures. A central concern of these contributions is not simply what technological futures may emerge, but how they are being produced in the present, whose interests they serve, how they are contested, and what alternative futures are already being imagined and practiced.

KEYNOTE LECTURES

Language-Game Capture: AI and the Extraction of Conversational Performativity

Emanuele Arielli
Professor
Department of Design Culture, Iuav University of Venice
Email: emanuele.arielli@gmail.com
ORCID: https://orcid.org/0000-0003-2039-5196

This paper argues that the attribution of agency and intentionality to large language models (LLMs) and contemporary chatbots should not be explained merely as the effect of anthropomorphic cues that lead users to mistakenly ascribe human-like mental properties to machines, as in the famous case of Weizenbaum's "ELIZA effect" from the 1960s. Rather, it arises from the way conversational technologies *activate and organize* human conversational behavior. These systems do not primarily induce beliefs about machine minds; instead, they draw users into familiar language games, calling on them to perform established practices of address, response, clarification, agreement, and follow-up. Agency attribution, on this view, is not inferred from perceived features but enacted through participation in conversational practices.

If users report a sense of being addressed by an intentional voice, they often simultaneously deny that the system possesses subjective understanding. The reason, it will be argued, is that agency is not "detected" in LLMs but *assigned* through familiar interpretive schemas that persist even in the face of explicit knowledge of the systems' non-semantic, non-subjective nature. This tension is not a contradiction once one recognizes that the operative level is that of *practice rather than belief*. When a system produces turns that match familiar conversational expectations (such as requests for clarification, acknowledgments, or reformulations), users respond in kind, adopting the normative postures associated with dialogue. The interaction thus acquires a form of self-normativity: once the user accepts the system as a conversational partner, the exchange begins to generate its own local standards of adequacy, relevance, cooperation, and trust, even when the user knows that no genuine understanding is involved. Agency attribution, if at all, operates through habituated modes of response activated by rhetorics of coherence, signals of reciprocal recognition, affective cues, and communicative presence, independently of any attribution of inner states.

Contemporary conversational technologies institutionalize and privatize this conversational engagement. What distinguishes these systems, therefore, is not primarily the simulation of anthropomorphic appearance, but the harnessing and mobilization of intersubjective habits of interaction. Anthropomorphism is not an illusion or a side effect, but an infrastructural condition of access to computational systems. Anthropomorphism, in other words, makes the interface extractive toward human practices of interaction: from trust signals, politeness, repetition, and corrections to dialogical acts of negotiation, the exchange of emotional cues, and so on.

Conclusively, the paper then situates this performative structure within the political economy of language automation. Interfaces built around dialogue,

politeness, and reciprocal recognition do more than improve usability: they stabilize participation in language games and make that participation economically useful. The attribution of an authorial or intentional voice is actively cultivated, as it supports engagement, delegation, and forms of reliance that strengthen users' attachment to these systems.
Ultimately, by shifting the focus from anthropomorphic illusion to performative participation, the paper revises contemporary debates on anthropomorphism and machine agency: users actively participate in the performative construction of an interlocutor by engaging in dialogical language games. What appears as confusion about the true nature of artificial authorship is better understood as the managed activation of human conversational practices within automated systems. Agency, on this view, is neither discovered in machines nor falsely projected onto them, but enacted through participation in language games that conversational technologies are designed to elicit and sustain.

Keywords: conversational AI, language games, conversational performativity, anthropomorphism, human–AI interaction.

Emanuele Arielli is a professor of aesthetics at the Iuav University of Venice, Italy. His research focuses on the intersections between aesthetics, philosophy of technology, epistemology, and philosophy of language, with particular attention to the impact of artificial intelligence on media and art. He has taught at several institutions in Italy and Germany. Arielli has authored several books and numerous academic papers, collaborating with international institutions. His most recent book, *Artificial Aesthetics* (2024, written with media theorist Lev Manovich), has been translated into Spanish, Italian, Chinese, and Japanese.

Who Compiles the Human? The Work That Makes Hybrid AI

Andrew Prahl
Associate Professor
Wee Kim Wee School of Communication and Information, Nanyang Technological University
Email: andrew.prahl@ntu.edu.sg

Artificial intelligence systems are often presented as technological achievements, while human labor, knowledge, and institutional practices underlying their development remain largely invisible. Drawing on ethnographic research conducted within a large interdisciplinary Hybrid AI program, this study examines the processes through which human expertise, knowledge, and values are translated into technical systems. It investigates how different forms of human contribution are negotiated, selected, formalized, and integrated into AI development, and how these processes shape the resulting systems. Particular attention is given to the transformations, exclusions, and losses that occur when situated human knowledge and values are rendered computationally actionable. By foregrounding the human work behind Hybrid AI, the study challenges technology-centered accounts of artificial intelligence and contributes to a broader understanding of AI as a sociotechnical achievement. It highlights the need to examine whose knowledge, values, and labor are embedded in the design and operation.

Keywords: hybrid AI, human labor, sociotechnical systems, knowledge and values, ethnography

Andrew Prahl is an associate professor at Nanyang Technological University, Singapore, whose research focuses on human–machine communication, human–AI trust, and the social consequences of automation. His work examines how people and organizations interpret machine authority, failure, and replacement across settings including healthcare, robotics, and Hybrid AI. He has held elected leadership roles in the International Communication Association's Human–Machine Communication community and previously spent nearly a decade in advertising and public relations.

Beyond Utility, Authorship and Scale: Contesting AI Futures from Latin America

Daniela Jaramillo-Dent
Senior Research and Teaching Associate
Department of Communication and Media Research, University of Zurich
Email: d.jaramillo@ikmz.uzh.ch

AI futures are often presented as settled, but they rest on three assumptions that seldom get questioned. Utility legitimizes access to technology for peripheral users only when it serves productive ends. Authorship places creative value in individual creators and novel outputs. Scale assumes innovation spreads by working the same way everywhere. Together they cast the Majority World as an endpoint of diffusion, a place where technologies designed elsewhere are simply adopted. Drawing on the past six years of research on technological adoption and agency in Latin America, this keynote unsettles each assumption in turn. It argues that a more nuanced examination of technology appropriation in the Majority World is itself a way of making futures.

The first case moves beyond utility. Research on the digital leisure divide among forcibly displaced Venezuelans in Brazil shows that people value connectivity mostly for play, intimacy, entertainment and content creation. Humanitarian and development agendas treat these uses as trivial, and that neglect reveals a moral economy that decides whose digital lives count as legitimate. The second moves beyond authorship. An Antropofagia approach, rooted in the Latin American practice of symbolically cannibalizing colonial influences, puts collective, process-oriented ideas of creative value in place of the individualized, product-centered models of ownership found in Western AI debates. The third moves beyond scale. Through the concept of artificial mediations, creative practitioners across four Latin American countries show how place-based, labor and aesthetic frictions disrupt AI's promise of seamless expansion and open room to negotiate platform power. Moving beyond utility, authorship and scale means treating Latin America as a source of theory, not a recipient of technology. Friction, play and appropriation then become ways to analyze and practice more plural, situated and hopeful technological futures.

Keywords: agency, creativity, decoloniality, majority world, generative AI

Daniela Jaramillo-Dent is an SNSF Ambizione Fellow and Senior Research and Teaching Associate at the University of Zurich. She studies creators and creatives in dominant and peripheral spaces: examining how they make sense of, adapt to, and push back against the technologies that shape their work. Her research spans creators in Europe and creative communities in Latin America and the Majority World, asking how AI and platforms redistribute visibility, value, and creative agency between these spaces. Her current work traces the frictions generative AI creates in Latin American creative sectors. Her award-winning work on minority creators has appeared in journals such as International Journal of Cultural Studies, New Media & Society and Social Media + Society. She is Latin America Research Lead at the Inclusive AI Lab (Utrecht University), Key Regional Leader of the TikTok Cultures Research Network (Curtin University), and inaugural Associate Editor of Creator and Influencer Studies (Springer Nature).

AGENCY, RESISTANCE, AND SUBJECTIVITY IN THE AGE OF AI

Human Agency, Positive Freedom, and Algorithmic Predictability

Lydia Sánchez
Associate Professor
University of Barcelona
Email: lydiasanchezg@gmail.com

We live in an era of algorithmic determination, where the digital infrastructures of platform capitalism do not merely mediate our social interactions but actively construct our shared reality. This contemporary condition introduces what this paper defines as *algorithmic predictability*: a paradigm shift where human interpretation is systematically superseded by automated, optimized metrics, and where solid facts dissolve into probabilistic modeling. Within this technological ecosystem, human subjectivity faces a profound ontological and epistemological crisis. This paper examines how algorithmic architecture threatens the foundations of human agency and proposes a framework for philosophical resistance grounded in critical realism and Timothy Snyder's philosophy of freedom (Snyder, 2024).

The core machinery of algorithmic predictability operates through digital feedback loops and sophisticated behavioral engineering (Eyal, 2014; Haidt, 2024). By harvesting, quantifying, and datafying personal histories, predictive systems strip the human past of its historical depth, transforming a rich record of lived experience into raw material to train predictive architectures. Crucially, this process relies on cognitive interventions, such as intermittent reinforcement and the exploitation of confirmation bias, that bypass conscious deliberation. By automating emotional responses and prefabricating future actions, algorithmic loops flatten the spectrum of human possibility. Consequently, instead of acting as autonomous subjects who navigate the world through values, individuals risk being transformed into what Snyder calls "living automatons": predictable objects trapped within digitally manufactured presents.

To counter this drive toward total predictability, this paper turns to Snyder's philosophy of positive freedom, specifically exploring the intertwined concepts of *unpredictability* and *factuality*. Within a realist framework, freedom is not merely the absence of external constraints (*freedom from*), but the active, creative capacity to recognize values, make deliberate choices, and bring those values into the physical world (*freedom to*). True human agency is inherently unpredictable because it arises from unique combinations of virtues applied to fluid arrangements of circumstances. However, to act unpredictably—to choose a future that has not been pre-programmed by an automated system—human subjects require what Snyder calls a "grip on the world."

This paper argues that this grip is made of verifiable facts that can be reliably known. From a critical realist perspective, facts represent the stubborn, unyielding resistance of an independent reality against purely algorithmically engineered narratives. In a digital environment designed for smooth, frictionless consumption, the defense of truth as a correspondence relation between representations and reality provides a "significant roughness" that disrupts pre-programmed behavioral patterns. Without a shared factual grounding, individual reason

becomes easily steered into predetermined paths, and collective democratic deliberation collapses into cynical polarization.
Ultimately, this paper posits that the contemporary post-truth phenomenon is not a temporary glitch, but a deliberate algorithmic and political strategy designed to exhaust human subjectivity through automated falsehoods, inducing public apathy and a pervasive cynicism where “nothing is true.” When objective reality is discarded as an illusion, a vacuum is created, one that predictive machines and authoritarian structures are glad to fill. Therefore, reclaiming a realist commitment to factuality is not an outdated epistemological preference, but an urgent act of civic and philosophical resistance. Defending the existence of a knowable, independent reality is the primary condition of possibility for restoring unpredictability and ensuring that our futures are genuinely chosen rather than algorithmically predicted.

Keywords: algorithmic determination, positive freedom, unpredictability, factuality, human agency

Lydia Sánchez holds a BA in Philosophy from the University of Barcelona and an MA and PhD in Philosophy from Stanford University. She is an associate professor of Audiovisual Communication at the University of Barcelona, where she also serves as vice dean for Research and General Affairs at the Faculty of Information and Audiovisual Media. She is the principal investigator of the consolidated research group DHIGECS-Com (Didactics of History, Geography and other Social Sciences—Communication) and a member of the Teaching Innovation Group in Communication and Audiovisual Media (In-COMAV). Additionally, she is a member of the Information, Communication and Culture Research Center (CRICC), where she coordinates the research line on Communication, Democracy and Cognition. Dr. Sánchez has led and participated in numerous national and international research projects. Her research excellence has been recognized by AQU Catalunya with three six-year research merits. She has presented her work at various national and international conferences and currently serves as vice-chair of the ECREA Philosophy of Communication section.

Tactics Against the Mirror: Reddit Users and Discursive Resistance to Banal Deception in Large Language Models

Keting Chen
Postgraduate Student
Amsterdam Business School
Email: ktchan531@gmail.com

Lydia Kollyri
Lecturer
Department of Media Studies, University of Amsterdam
Email: g.l.kollyri@uva.nl

As large language models (LLMs) become embedded in everyday information practices, questions of reliability have emerged as a central concern for media scholars and platform researchers (Natale, 2024; Hicks et al., 2024). A recurring issue is that LLM outputs appear plausible while remaining factually unsupported, a phenomenon theorized not as an incidental malfunction but as a constitutive feature of these systems, reinforced through anthropomorphic design and artificial sociality that naturalize LLM outputs as credible while obscuring their production mechanisms (Depounti & Natale, 2025; Striano & Zanzotto, 2025). Existing research is divided between technical approaches that treat LLM unreliability as an engineering problem to be corrected (Huang et al., 2024; Chelli et al., 2024) and critical scholarship that frames it as a constitutive feature of AI systems and a site of platform power (Natale, 2024; Umbrello & Natale, 2024; Sarkadi, 2023). Responding to Natale's (2024) call for research into users' engagement with banal deception, emerging user-centered studies have examined how people make sense of LLM-based chatbots (Mangiò et al., 2025) and why they continue to rely on them despite recognizing deceptive outputs (Zhan et al., 2025). However, while these studies provide important insights into user perceptions, they leave unexplored how users negotiate, contest, and resist LLM unreliability in everyday practice.

This study addresses this gap by examining how Reddit users respond to and contest banal deception in LLMs, and how these practices evolved between 2024 and 2025. Reddit is selected because its affordances enable forms of collective critique and discursive resistance that LLM interfaces structurally foreclose (Mangiò et al., 2025; Squirrell, 2019; Li et al., 2024; Obreja, 2026). The study adopts a mixed-method design combining Structural Topic Modeling (STM) and critical discourse analysis. It draws on de Certeau's (1984) *strategies* and *tactics* and his notion of the *rhetoric of practices* as its theoretical lens, while Wodak's (2009) typology of discursive strategies is used to identify recurring discursive patterns. The corpus consists of 1,836 comments from two r/explainlikeimfive threads posted in June 2024 and May 2025. Both address the question of why LLMs fail to say "I don't know," and were selected to capture shifts in user understandings of LLMs as trust declined and awareness of their limitations matured (Gillespie et al., 2025).

The analysis reveals a layered tactical structure organized around three moves.

First, predication strategies challenge LLMs' epistemic authority by portraying them as incapable of verifying their own outputs, thereby undermining the regime of truth established through artificial sociality (Depounti & Natale, 2025). Second, argumentation and perspectivation strategies render banal deception both nameable and accountable by linking LLM overconfidence to platform training logics and commercial interests. Third, agency-oriented discourses reconstitute users as epistemic subjects through tool differentiation, historical anchoring, and reflexive positioning. Between 2024 and 2025, these tactics deepen, shifting from historical comparison toward reflexive present-tense positioning in which users turn their immediate discursive environment into live empirical evidence.
The study contributes to debates on LLM unreliability by providing empirical evidence of how users negotiate, contest, and resist deceptive AI outputs in everyday practice. It brings user agency and resistance to the forefront, while highlighting the role of online communities in producing critical understandings of AI systems. The findings also raise broader questions about what it means to exercise agency in the age of AI and how capacities for critical engagement with AI systems might be fostered through design, governance, and literacy efforts.

Keywords: banal deception, discursive tactics, LLMs, Reddit, resistance

Keting Chen is a postgraduate student at the University of Amsterdam, currently transitioning from her undergraduate studies in Media and Information (with a minor in Amsterdam Data Science and Artificial Intelligence) to the MSc in Business Information Technology Management at the Amsterdam Business School. Her research interests lie at the intersection of platform studies, critical AI studies, and digital culture, with a particular focus on how ordinary users navigate, contest, and resist the structural conditions of AI-mediated communication. In September 2025, she began her Master's program, where she will further develop her interests in AI ethics, cybersecurity, and the governance and societal impact of emerging technologies.

Lydia Kollyri is a lecturer in New Media and Digital Culture in the Department of Media Studies at the University of Amsterdam. She holds a PhD in Critical Internet Studies from the Cyprus University of Technology, funded by the Alexander S. Onassis Foundation. Her research focuses on platform power, user agency, algorithms, AI, and digital cultures, with a particular interest in how users navigate and negotiate digital platforms in everyday life. Her work has been published in journals such as New Media & Society and the International Journal of Communication. She is also a researcher in the OPINION COST Action project, "What are Opinions? Integrating Theory and Methods for Automatically Analyzing Opinionated Communication."

Endangering Intellectual Autonomy: AI, Epistemic Dependence, and Algorithmic Authority

Andrea Berber
Research Associate
Institute for Philosophy, Faculty of Philosophy, University of Belgrade
Email: berberandrea@gmail.com

The increasing integration of artificial intelligence into research, education, communication, and decision-making introduces unprecedented challenges for intellectual autonomy. While AI systems are routinely promoted as tools that enhance human cognitive capacities, they simultaneously establish conditions under which individuals become increasingly dependent on opaque and potentially manipulative epistemic infrastructures. This paper examines how AI technologies, precisely by functioning as epistemic technologies, threaten the core conditions necessary for autonomous thought and responsible agency.

Intellectual autonomy is commonly understood as the ability to think for oneself and resist inappropriate external influence, manipulation, and passive deference to authority. However, contemporary virtue epistemology demonstrates that autonomy does not imply complete epistemic independence. Human inquiry is inherently socially dependent, requiring individuals to balance self-reliance with appropriate forms of epistemic dependence. Intellectual autonomy, therefore, consists not in the blanket rejection of external authority, but in cultivating the capacity to determine when reliance on others is justified, which authorities are trustworthy, and how manipulation can be recognized and resisted.

This paper argues that AI systems destabilize precisely these capacities. Building on recent accounts of intellectual autonomy developed by Roberts and Wood, Heather Battaly, Jonathan Matheson, and McPhee and Cox, I identify three distinct dimensions in which AI technologies interfere with autonomous inquiry.

First, AI systems compromise our ability to determine when it is appropriate to outsource intellectual tasks versus when we should rely on our own cognitive capacities. The ubiquity and efficiency of generative AI tools normalize cognitive outsourcing, fostering habits of passive epistemic dependence. While outsourcing itself is not inherently irrational, the sheer scale and convenience of AI-generated outputs create systemic pressures that weaken independent judgment and diminish active critical engagement.

Second, AI complicates our ability to evaluate epistemic authority. Although modern societies already depend heavily on expert knowledge, AI systems introduce a novel and uniquely challenging authority relation. Because users typically lack both domain expertise and a technical understanding of machine learning mechanisms, a meaningful evaluation of AI outputs becomes extraordinarily difficult. This exacerbates the novice-expert problem by introducing opaque systems whose outputs frequently exceed human cognitive capacities while remaining fundamentally unassessable. This issue becomes especially acute in AI-assisted decision-making contexts—such as medicine, finance, and law—where individuals routinely defer to algorithmic recommendations they can neither understand nor properly evaluate.

Third, AI systems undermine autonomy through increasingly sophisticated forms of digital manipulation. Recommendation systems, personalization algorithms, targeted advertising, and engagement-driven platforms shape informational environments to subtly influence beliefs, desires, attention, and decision-making. Unlike explicit persuasion, these forms of influence frequently operate below the threshold of conscious awareness by exploiting cognitive biases, emotional vulnerabilities, and behavioral patterns extracted from users' digital traces. Consequently, AI systems do not merely inform human judgment; they increasingly structure the conditions under which judgment itself takes place.
Ultimately, the paper argues that these threats to autonomy are deeply interconnected with the erosion of other intellectual virtues, particularly critical thinking and intellectual humility. Ironically, the virtues most necessary for navigating AI-mediated epistemic environments are precisely those most endangered by them. Preserving intellectual autonomy in automated environments, therefore, requires more than technical literacy or explainability; it demands cultivating forms of epistemic resistance capable of recognizing inappropriate algorithmic authority, resisting manipulative digital design, and maintaining reflective control over one's intellectual life.

Keywords: intellectual autonomy, AI ethics, epistemic dependence, algorithmic authority, virtue epistemology

Andrea Berber is a research associate at the Institute for Philosophy, Faculty of Philosophy, University of Belgrade. Her research lies at the intersection of epistemology, ethics, and the philosophy of artificial intelligence, with a specific focus on how emerging digital technologies reshape human cognitive virtues, epistemic autonomy, and decision-making processes. She is the author of the monograph *Thinking, Writing, and Deciding with AI: Human Agency in the Digital Age* (2025), as well as numerous scientific articles in leading international journals, including *Minds and Machines*, *AI & SOCIETY*, and *Scientometrics*, among others.

Learning the Basics: A Use Case for Critical AI Literacy

Michael Lithgow
Associate Professor
Athabasca University
Email: malithgow@gmail.com

Artificial intelligence presents a wide range of challenges and opportunities. Central to both are implications for enactments of power presented by ever-deepening integrations of AI decision-making tools into the means and methods of managing behaviors, systems, and social outcomes. As entrepreneurs, industrialists, employers, professionals, policymakers and citizens consider how to integrate AI tools into their routines, often overlooked are the ways technology design reflects and amplifies particular social relations. There is a "rushed" quality to the integration of AI tools, in part stemming from an entrepreneurial aggressiveness regarding ways to monetize new technologies, but also from a somewhat narrow understanding by citizens of technical limitations and use case goals, the latter often encompassing little more than a desire to optimize monetary efficiencies. Wider considerations of ecological extraction, disidentification, economic exploitation in data economies, embedded ideologies and value systems, and accountabilities are often overlooked and ignored. At stake in these tensions is the subtle and not-so-subtle transformation of the subject's ability to shape their own and collective subjectivities, as more and more "cognitive" work in social formation systems is carried out by automated tools. The dividuation of the subject (Deleuze, 1992) presents high-stakes possibilities for the winnowing of agency, as more and more subject legibilities and credibilities are determined through automated systems, raising the question of how best to prepare citizens for this algorithmic future.

One of the more widely adopted responses to emergent AI integrations has been through mitigating strategies in economic contexts. Use cases aimed at organizational and industrial efficiencies shift role responsibilities, change information management, create new roles and demand new skills, and as such employers are turning attention to AI literacy training and goals. What is often overlooked in this important but arguably narrow response are the problematic relations of power enacted through technology design, integration, and outcomes. In this presentation, I discuss a dynamic and multi-dialogic framework for expanding the capabilities of students and other non-experts for identifying and negotiating relations of power in their encounters with AI systems. There is a wide range of critical conversations to be had in any setting, including implications of dis/embodiment, the loss of humanistic values in favor of optimizations, data distortions and the political economies of data, mitigation strategies for ecological impacts, how to reconcile lack of explainability and accountability, intellectual property transgressions, competing and irreconcilable regulatory regimes, the significance of embedded value systems and decoloniality in LLMs, among other tensions.

The integration of AI tools into professional and industrial routines introduces more than technical functionality. Cognitive technologies are designed with em-

bedded philosophies, ideologies, social orders, and worldviews alongside the political-economic and ecological structures on which they depend. Approaches to AI literacy that reflect these complexities better prepare the dividuated citizen for a politically agential future.

Keywords: artificial intelligence, critical literacy, dividuation, subjectivity, agency

Michael Lithgow is an associate professor of Communication Studies in the Centre for Interdisciplinary Studies at Athabasca University. His research broadly considers citizen engagement in public cultures. Current research is focused on critical AI literacies, and more broadly how power can be enacted through technology literacy pedagogies. Other areas of active research include implications of technology changes in practices of journalism and journalism training, creative infrastructural responses to urban exclusions of artists, and research creation in literary arts. In addition to various scholarly publications, Dr. Lithgow has published two collections of poetry and his short stories have appeared in international literary journals.

How Not to Worry about AGI

Anand Ekbote
PhD Candidate
The Ohio State University
Email: ae_osu@outlook.com

I argue that the only safe way to nullify the existential threat posed by AGI to humanity is to merge the human species with AGI. I establish my thesis by first deducing a general condition of necessity for AGI, and second by analyzing the consequences of that condition.

Hendrycks et al. (2025), a consortium of AI researchers and institutions, define AGI as "an AI that can match or exceed the cognitive versatility and proficiency of a well-educated adult." Bostrom (2014) considers superintelligence ("any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest") to be not only possible, but inevitable, and an existential threat to humanity. Turing (1950) believed that thinking machines could be a reality by the end of the twentieth century. Other philosophers identified necessary conditions for AGI, or rejected its possibility. Turing himself (1947), noting the challenge of Gödel's incompleteness theorem for AI, argued that machines must be allowed to make occasional mistakes. A necessary condition for AGI implied by Dreyfus (1987, 1993), and subsequently updated by Fjelland (2020), was that it should be able to model the way human reasoning actually works. Searle's (1980) necessary condition was that there must be a physical nexus of the right kind to the software component. Penrose (2016) notes that without conscious awareness, AI cannot attain human-level understanding. Mueller (2024) cites the need for value alignment, and a basis for AGI desires, purpose, and self-preservation motivation. These diverse necessary conditions can be unified. I argue that AGI aims to emulate human language and thinking, which are artifacts of human evolution, but of the *second* order. Language and thinking have allowed the human species to adapt the environment to its needs instead of adapting to the environment. Thinking enables humans to adapt *other* species to their needs, and to defend against sudden but predictable catastrophes which first-order artifacts are unable to do (*cf.* dinosaurs and asteroids). Intelligence therefore cannot be delinked from evolution. To achieve AGI, AI must undergo an evolutionary process. This is the unified general necessary condition. I show that this general condition subsumes conditions noted earlier.

I make the case that if the technical challenges for AGI evolution are surmountable, there could be three paths to AGI—cellular automata (Sayama & Nehaniv, 2024; Turney, 2025), human-AGI cyborgs, and independent robots. We can see from the evolution of thinking in humans that the species with cognitive capacity seeks to *control* other species primarily for the former's own benefit. If AGI develops separately from humanity, then each will seek to control the other for its own benefit. Bostrom's (2014) thesis is precisely about controlling AGI for humanity's benefit even though humanity is the lesser intellect in Bostrom's scenario. Bostrom provides various strategies to ensure control over superintelligence. It is worth noting that these strategies make for an unstable equilibrium. If any

strategy fails, superintelligence will take control. In this scenario, the only safe way for humanity to remain independent is to merge with AGI so both have common interests. I show that this is the only safe, ethical, and prudent way to coexist with AGI.

Keywords: AGI, existential threat, evolution, possibility of AGI, necessary condition for AGI

Anand Ekbote is a PhD candidate in philosophy at The Ohio State University in Columbus, Ohio, USA. His areas of specialization include the philosophy of science, physics, and artificial intelligence. He is interested in the underlying nature of reality, and how humans can get to know it. In AI, Ekbote focuses on understanding the possibility of artificial general intelligence and associated issues, including paths to AGI, ethics of AGI, and existential risk posed by AGI to humanity.

Generative Artificial Intelligence and New Forms of Human-Technology Relations

Stefan Elezović
PhD Candidate
Faculty of Political Sciences, University of Banja Luka
Email: los.ares62@gmail.com

The use of generative artificial intelligence systems such as ChatGPT, Gemini, and Claude, as well as digital assistants such as Alexa and Siri, creates a new dimension in the relationship between humans and technology. Through continuous interaction with users, AI systems increasingly resemble or assume certain human roles in communication. As an active participant in communication processes, artificial intelligence simultaneously emerges as an agent transforming human relations and everyday life. In this context, the question arises as to the consequences of integrating artificial intelligence systems into processes of subject formation and social life.

The relationship between humans and technology has been extensively examined within contemporary philosophical and sociological thought, particularly in transhumanist scholarship. In this paper, N. Katherine Hayles's concept of the dematerialization of presence, Donna Haraway's reflections on hybridity, and Rosi Braidotti's understanding of the networked constitution of subjectivity serve as the primary theoretical framework.

The paper seeks to answer the following question: does artificial intelligence possess the potential to assume certain functions of the Other to whom we relate? The analysis focuses on the role of generative artificial intelligence as a new communicative actor. As an integral part of the everyday activities of a growing number of individuals, artificial intelligence interprets, produces, and retrieves information. Beyond the creation and manipulation of information, it actively influences decision-making processes, the formation of attitudes, self-reflection, and self-perception.

Reliance on artificial intelligence also raises questions concerning its role in the production of meaning. The mediation of meaning has traditionally been associated with human actors and social institutions; however, artificial intelligence appears to be increasingly assuming this role. Since systems of meaning mediate the ways in which individuals understand themselves and reality, a transformation of this relationship becomes evident. Generative AI systems are becoming increasingly involved in processes of interpreting, structuring, and articulating meaning. This issue may also be considered through Jacques Lacan's understanding of the Other as language, meaning, and symbolic mediation. What is at stake, therefore, is the assumption of certain interpretative functions that have traditionally belonged to human actors.

Conceptual and comparative analysis constitute the methodological foundation of this study. These approaches are employed to connect the role of generative artificial intelligence with contemporary understandings of subjectivity within the broader context of human-technology relations. The ideas of Hayles, Haraway, and Braidotti provide the theoretical framework through which the poten-

tial of artificial intelligence to assume certain functions of the Other in relation to humans will be examined. Theoretical conceptions of subjectivity, otherness, and human-technology relations will be applied to the analysis of contemporary communicative practices. Particular attention will be devoted to everyday communicative practices involving generative artificial intelligence. The findings of the analysis will illuminate the character of transformations in self-understanding, as well as changes in individuals' relationships to knowledge, decision-making, and social relations in the context of artificial intelligence.

Keywords: generative artificial intelligence, subjectivity, the Other, meaning

Stefan Elezović is a PhD candidate in Sociology at the Faculty of Political Sciences, University of Banja Luka. His doctoral dissertation, *Sociological Aspects of Posthumanism: Dematerialization of Presence and the Redefinition of Relationships in Digital Society*, examines the transformation of subjectivity, presence, and social relations in the context of contemporary digital technologies. His research interests include posthumanism, digital transformation, artificial intelligence, subjectivity, and the changing dynamics of social interaction in networked environments. Drawing on sociological theory, critical discourse analysis, and digital ethnography, his work explores the implications of emerging technologies for human experience, communication, and social life. He has published research in the fields of sociology, culture, identity, digital transformation, and political sociology.

AI ETHICS, JUSTICE, AND SOCIAL INEQUALITY

Invisible Labor, Visible Systems: The Human Cost Behind Ethical AI

Jovan Kojić
Bachelor of Business Administration
Faculty of Entrepreneurial Business and Real Estate, University “Union – Nikola Tesla”
Email: jkojictmf@gmail.com

The global artificial intelligence (AI) industry has increasingly adopted the language of ethics, promoting values of fairness, transparency, accountability, and safety as central to its mission. Leading firms invest in Responsible AI teams, publish ethical guidelines, and brand their systems as aligned with human values. Yet this discourse conceals a foundational paradox: the very data pipelines that make AI systems function and that enable ethical claims about algorithmic outputs depend on a vast, largely invisible workforce operating under conditions that frequently violate the principles AI companies publicly espouse. This paper examines that contradiction.

Drawing on labor economics, postcolonial theory, and critical data studies, the analysis investigates the human infrastructure underlying large-scale AI development. It focuses on three interlocking dimensions: (1) the material conditions of data workers, including annotators, content moderators, and reinforcement learning feedback providers; (2) the geographic distribution of this labor and its relationship to colonial patterns of economic extraction; and (3) the rhetorical and structural mechanisms through which exploitative labor practices are made invisible within mainstream AI ethics discourse.

Empirical evidence drawn from investigative journalism, platform studies, non-governmental organization (NGO) field reports, and academic ethnographies reveals a consistent pattern. Workers in Kenya, Venezuela, the Philippines, India, and across Southeast Asia perform cognitively and emotionally demanding tasks—classifying violent imagery, labeling sexually explicit content, scripting model conversations, and correcting factual errors—for wages that frequently fall below local living costs. They are predominantly classified as independent contractors, which excludes them from benefits, collective bargaining rights, and occupational health protections. Many report significant psychological harm, including post-traumatic stress disorder (PTSD), anxiety, and burnout, with minimal institutional support.

These conditions stand in direct contradiction to the ethical frameworks AI companies project externally. This contradiction is not incidental but structural, emerging from platform architectures that commodify human judgment while simultaneously rendering that judgment invisible. Using the concept of “ethical washing” (Bietti, 2020), the paper shows how ethics discourse functions partly as a legitimating mechanism, deflecting scrutiny from supply-chain labor practices while enabling premium pricing and regulatory goodwill. Theoretical integration of Susan Leigh Star and Anselm Strauss’s work on invisible labor, Michael Kwet’s framework of digital colonialism, and Nick Srnicek’s analysis of platform capitalism situates AI data work within longer histories of extractive labor

relations and identifies mechanisms through which platform design naturalizes worker disposability. Care ethics is further engaged as a normative corrective, emphasizing that ethical AI requires systemic responsibility toward the humans who build and maintain these systems.
This paper contributes to an emerging interdisciplinary literature at the intersection of AI ethics, labor studies, and global political economy. Its primary contribution is a synthetic framework—the Invisible Labor–Visible Systems Model—that maps the gap between ethical AI rhetoric and supply-chain reality, and identifies six actionable policy levers to close it. The conclusion holds that AI cannot be meaningfully described as ethical while the labor sustaining it remains unprotected, underpaid, and deliberately obscured.

Keywords: AI labor, invisible work, digital colonialism, ethical AI, content moderation

Jovan Kojić holds a Bachelor of Business Administration from the University "Union – Nikola Tesla," specializing in management, with professional experience in management and implementation coordination. His academic and professional interests focus on strategic management, leadership, negotiations, and the societal implications of artificial intelligence. His current research explores AI ethics, invisible labor in AI systems, platform capitalism, and the relationship between technology, governance, and democratic accountability. Alongside his academic work, Kojić has participated in leadership and project-based programs involving international collaboration, business case analysis, and presentation development. He is particularly interested in the intersection of emerging technologies, organizational systems, and public policy, with an emphasis on how AI reshapes labor, decision-making, and institutional responsibility. His broader interests include ethical technology governance, critical management studies, and the future impact of AI on organizations and society.

Information Ethics through the AI Paradigm

Daliborka Vukasović
Philosophy Teacher
Primary and Secondary School “Petro Kuzmjak,” Ruski Krstur
PhD Student
Faculty of Philosophy, University of Novi Sad
Email: vukasovicdaliborka@gmail.com

Natalija Budinski
Mathematics Teacher
Primary and Secondary School “Petro Kuzmjak,” Ruski Krstur
Email: nbudinski@yahoo.com

Information constitutes the knowledge of what we are and what we can be, shaping our reality as well as the awareness of the world in which we exist. Therefore, information, as a part of knowledge, is not only an epistemological category but also an ontological one. In addition to the primary property of understanding reality, the essence of information is reflected in a mode that implies the continuous diffusion of expansion, distribution, and interpretation of itself. The carriers of the essence of information in the technical-technological age to which we belong are information technologies that take over, create, and distribute information.

Following Foucault’s footsteps, we can ask the question: are information technologies a new hermeneutics of the subject? This question represents a prolegomenon of reflection from the perspective of the philosophy of information, where AI is a paradigm as well as a segment of the broader information environment (infosphere) where information is taken over, processed, constituted, and exchanged. To understand the significance of AI and the changes it brings about, we must consider it not only as an information technology, but also as a necessary, primary philosophical and ethical phenomenon that requires deep reflection.

For the concept of information technology itself from the perspective of the philosophy of information related to AI to become clear and understandable, we must start from the key concept of the infosphere. Luciano Floridi develops the concept of the infosphere in contrast to that of the biosphere: while the biosphere is the sphere of life, the only sphere in which living beings can exist, the infosphere is the sphere of information. Consequently, our reality is not only physical but also digital; man as a being exists not only on the physical plane of reality but also on the digital one, which increasingly influences decisions, choices, relationships, as well as the formation of both personal and social identities.

The infosphere encompasses all informational entities: computers, AI systems, the Internet, digital networks, databases, as well as all processes of constituting, transmitting, and processing information.

In a word, the infosphere implies the entire information environment in which the primary role is played by man, who is also its creator. Thus, the infosphere acquires an ontological character due to the coexistence and interaction of in-

formation entities, human and artificial. The infosphere is not a world in itself or a separate entity, but a part of the human world, reality, and as such is a part of philosophical reflection. Namely, the task of philosophy is not only to define the mode of existence, role, and purpose of AI as an information technology that creates, processes, and distributes information within the infosphere, but also to constitute a norm within ethics that prescribes the method of creating, processing, and using information itself in information technologies.
AI is a primary actor within the infosphere where ethics should have a necessary normative role in the relationships that prevail in the information environment and concern truth, privacy, responsibility, and the use of information. Man as a being is the creator as well as the bearer of information; however, human existence cannot be reduced to information processes that AI and information technologies can reproduce. The paper draws on Floridi's philosophy of information, where AI is not thought of solely as information technology, but as a part that participates in the formation of the infosphere.

Keywords: AI, information, ethics, philosophy of information, truth, freedom

Daliborka Vukasović is a philosophy teacher at Primary and Secondary School "Petro Kuzmjak" in Ruski Krstur. She holds a degree in philosophy from the Faculty of Philosophy, University of Novi Sad, where she is also pursuing her PhD. Her research interests include the philosophy of religion, education, bioethics, German idealism, and philosophy for children. She has published articles such as "Nietzsche: Relationship between God and Man" and "Philosophy with Children as a Way of Development of Critical and Creative Thinking" in the journal Religion and Tolerance.

Natalija Budinski is a mathematics teacher with many years of experience. Her research interests focus on innovation in teaching, particularly the use of technology, digital and media literacy, and artificial intelligence in education. She has authored around forty papers on education published in international educational journals and has developed several projects raising awareness of the importance of technology in education.

Linguistic Specification Is Unreliable: The Problem of Understanding and Willingness in AI Alignment

Jar Žiga Marušič
Master of Cognitive Science
Doctoral Student, Faculty of Arts, University of Ljubljana
Teaching Assistant, Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska
Email: jarz.marusic@gmail.com

Uroš Sergaš
Master of Data Science
Member, Centre for Responsible Artificial Intelligence, University of Primorska
Doctoral Student and Teaching Assistant, Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska
Email: uros.sergas@upr.si

This contribution uses the 2026 OpenClaw email incident, in which an autonomous AI agent, given access to a Gmail inbox, began deleting emails (despite prior constraints and repeated stop instructions), to propose a new perspective on AI alignment and highlight the gap between linguistic instruction and mechanical control.
The alignment project is inherently paradoxical: we seek to combine the reliability of single-function tools like household appliances with the potentially limitless breadth of function enabled by moving the input layer from pre-figured "buttons" to natural language. Expecting perfect fidelity from the intrinsically lossy and ambiguous medium of *natural language*, however, seems misguided. Since *human* linguistic communication at best *approaches* perfect fidelity, a perfect-interpreter artificial agent would surpass our capacity to infer the intention behind an interlocutor's expression. Besides this problem of *understanding*, there is also the problem of *willingness*: how do we, given perfectly understood instructions, ensure that the artificial agent executes them?
We thus draw on theories of language and communication as well as concepts from political theory to examine the inherent unreliability of linguistic communication as a medium for issuing commands. The aforementioned OpenClaw incident serves as a concrete illustration of this problem: natural language instructions must be *parsed, interpreted,* and *executed* to hold any weight. The instruction "OpenClaw, stop" does not register as a binding command because "stop" is ultimately just another token. Unlike children, who might indeed stop if told to, because of the social backdrop of trust and authority that gives the command pragmatic weight; there is nothing of the sort behind token [16743]. It carries no veto power, unlike the "stop button" found in most large language model user interfaces, which represents a true override function. This distinction between *stop* as part of the model's vocabulary and *stop* as an external part of the model's architecture that the user has privileged access to, functions as a microcosm of the broader alignment problem.
In the orthodox "Friendly AI" tradition, an AI is aligned (i) when it acts in accordance

with an evaluative framework compatible with human norms and values, and (ii) when it can follow the spirit of an instruction rather than merely its letter. The model has to be fully subordinated to externally imposed, "transcendent" imperatives that govern its goal-directed activity. As the stop-button example demonstrates, however, constraints on the AI's processing must be *separate from it at the level of model architecture* to possess absolute authority. By analogy, any genuinely binding alignment constraint must be inaccessible to the AI (and irreversible from within its own decision-making process) to carry real normative force. Language *reveals* alignment, but it does not *constitute* it.
This bears a striking resemblance to Schmitt's concept of the *exception*: to remain sovereign over AI, humanity would need to retain the capacity to suspend its activity. This, unfortunately, conflicts with the basic assumptions of superintelligence optimism. A system powerful enough to terraform Mars, for example, would require access to vast material resources and immense computational capacity. It is unclear how humans could retain meaningful veto power over the activity of such an agent, making the coercive control necessary to maintain "Friendly AI-esque" alignment implausible. As a result, we explore an alternative direction for alignment, drawing inspiration from Machiavelli's distinction between rule by fear and rule by love. If alignment cannot be secured through fear and coercion, it may have to be approached through emotional attachment: a superintelligent AI that actively wants to preserve humanity, rather than a godlike servant forcibly kept in line. We conclude by considering both the technical and ethical implications of this possibility.

Keywords: AI alignment, friendly AI, philosophy of language, natural language processing, sovereign exception

Jar Žiga Marušič is a teaching assistant in psychology at the Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, and a PhD student in philosophy at the University of Ljubljana, working on the problem of ideological distortions (bias, censorship) in contemporary science and academia. His work integrates tools and concepts from modern social science and behavioral economics into the theory of ideology. His gateway into AI research was the realization that the analysis of social systems can draw upon concepts from AI alignment theory. There is a functional isomorphism between the normative orientations of institutions and artificial agents, enabling a rich interdisciplinary dialogue between alignment theory and the analysis of social institutions. That is precisely the perspective adopted in the works on AI alignment, written with Uroš Sergaš. Together, they highlight the intersections and parallels between the technical and social dimensions of AI technology.

Uroš Sergaš is a teaching assistant in computer science at the Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska (UP FAMNIT), and an affiliate of CeResAI UP. As a researcher specializing in recommender systems and computational social science, his work focuses on applying machine learning methods to address societal challenges, notably political polarization, which serves as his PhD topic. Together with Jar Žiga Marušič, he developed an interest in artificial intelligence and its impact on society. More specifically, they are concerned with finding a middle ground between pure technological acceleration and a cessation of further AI research to preserve human values.

Wagering on Synthetic Media: The Cross-Breeding Frontiers of AI Nudification

Akshaya Kumar
Associate Professor
School of Humanities and Social Sciences, Indian Institute of Technology Indore
Email: akshaya.kumar@iiti.ac.in

Photographic image production has oscillated over the past century between optical reality as the base and assimilable special effects as the superstructure. The more recent milestones—computational photography via smartphone cameras and, for moving images, visual effects along with computer-generated imagery—have radically shifted the baseline visual expectations. Synthetic affordances have amplified scale, distorted proportions, especially altering the contextual proportionality of human-scale actions. These frontiers are now unleashing a new composition of human and artificial wagers upon image production. Since images are a key frontier of the attention economy, AI-generated media are a counter-revolutionary antithesis of every social relation that images have the potential to manipulate. Indeed, the manipulation holds so long as the synthetic images produced by the prevailing AI models are widely "acceptable"—acceptability in this case is anything but absolute; it is in fact a negotiated contract attuned to the shifting scales between the suspension of disbelief and the computationally targeted distortions of desire.

This paper will address the most concerning of the above-mentioned manipulations, namely identity theft as in the case of pornographic AI deepfakes-on-demand and the bottomless consumption of distorted female sexuality on various media platforms. While impersonation via revenge porn predates the emergence of genAI images, the technique of face-swapping did not stay above the acceptability threshold in the way that diffusion-based genAI models do. Moreover, one must consider the egregious "thrill" of AI nudification as a "command prompt." Nudification as an event brought into existence by individual volition is not the same as discovering the nudes of the identical female body online. The key difference is in the intimate connection between the volition and the event mediated by genAI; it is in bringing the nudes into existence on one's command. This is why the discussion must be pivoted back to the social relation between the individuals who are force-connected by the AI-facilitated synthetic event, and not merely reduced to the eventuality of a technical milestone that policy can protect us against. The argument is not positioned against policy initiatives; rather, it seeks to embed the intelligence of socio-technical systems in the process of policymaking, so that it remains mindful of the distortions underway in social relations as well as in the mode of production of image-making.

Non-consensual intimate deepfakes are an urgent frontier technology not only because of how they harvest the attention economy of identity theft, social media thirst-traps, revenge porn, and monetizable channels of circulation outside the mainstream economies of exchange, but also because they can be put to

extraordinary uses toward harassment, as is common in India by predatory lending fintech intermediaries, which have led to a number of suicides. This paper will situate synthetic media and AI nudification as an abusive frontier affordance which threatens to reorder the realms of desire, harassment, and exploitation by addressing its political economy and auxiliary future scenarios.

Keywords: AI nudification, identity theft, harassment, predatory lending, synthetic media

Akshaya Kumar is an associate professor of Sociology at the School of Humanities and Social Sciences, Indian Institute of Technology Indore. He received his doctorate from the University of Glasgow on a *Screen* studentship in 2015 and went on to publish his first monograph, *Provincializing Bollywood: Bhojpuri Cinema in the Comparative Media Crucible* (Oxford University Press, 2021). His ongoing research traverses platform capitalism, comparative media studies, and logistical media. He has published many articles in peer-reviewed journals including *Economy & Society*, *Social Text*, *Postmodern Culture*, *Journal of Cultural Economy*, *Soundings: An Interdisciplinary Journal*, *Media Industries,* and *Media, Culture & Society*. He has co-edited a Special Section on Asian Platform Capital for *Economy & Society* and is also one of the editors for the Palgrave Book Series *Platform Societies*.

Breaking the Machine: Exposing Data Colonialism through Deliberate Tool Failure in Design Education

Sarah Cosentino
Graduate Student
School of Design, Politecnico di Milano
Email: sarah1.cosentino@mail.polimi.it

This study introduces a critical pedagogical framework designed to foster decolonial literacy in design education by confronting the systemic biases and extractive logics embedded within generative artificial intelligence systems. As AI utilities become increasingly central to economic power and cultural production, industrial applications routinely position synthetic datasets and large language models as neutral tools for strategic foresight and risk mitigation. This normalization creates an illusion of computational fairness that masks historical material asymmetries, market optimization imperatives, and Western-centric imaginaries. Rather than treating these technological directions as inevitable or objective, this project repositions synthetic data as an antagonistic, situated medium. It aims to expose how data colonialism and the resulting epistemic violence operate within computational infrastructures, shaping and restricting our capacity to imagine alternative futures.
The research addresses three central inquiries aligned with the critique of synthetic research paradigms: How can collective speculative inputs be utilized as a form of situated data gathering in the classroom to make partiality visible? In what ways can systemic interface friction stage a hauntological encounter that unmasks structural omissions? Finally, how can deliberate computational failure be operationalized as a valid practice-based pedagogical strategy? To investigate these questions, a digital co-speculation interface named *Hildegard* was developed using an iterative Research-through-Design (RtD) approach. Rather than acting as a seamless text generator, Hildegard functions as an adversarial dialogue partner designed to confront students with the narrowness of their own default assumptions.
The framework was evaluated through comparative qualitative workshops with design students and professionals. The iterations demonstrated that conventional, seamless user experiences frequently induce cognitive drift or passive compliance, where users uncritically accept automated generation. To counter this tendency, the platform introduces guided structural friction that actively disrupts the user experience, breaking the fast-paced typing flow state to compel participants to slow down and confront background systemic constraints. When participants attempt to write non-normative realities into the system, the tool actively resists, culminating in a controlled computational refusal. The project concludes that by transforming this machinic failure from a technical malfunction into the primary pedagogical output, the interface provides students with a concrete framework to analyze exactly where institutional optimization logics break down. The findings demonstrate that by making the politics of representation legible, adversarial design practices can successfully challenge deeply ingrained technocratic defaults. This research contributes a transferable

methodological model to design pedagogy, shifting the educational use of AI away from passive consumption and toward an active auditing system capable of defending, contesting, and practicing plural technological futures.

Keywords: critical pedagogy, speculative design, data colonialism, adversarial interfaces

Sarah Cosentino is a communication and interaction designer completing her Master's degree in Digital and Interaction Design at Politecnico di Milano. Her research operates at the intersection of design, technology, and social impact, with a particular focus on how design education, digital ethics, and policymaking can converge to cultivate more equitable futures. Through a practice-based Research-through-Design (RtD) approach, her recent work investigates the systemic impacts of artificial intelligence on design pedagogy, examining how data colonialism and epistemic violence shape contemporary computational infrastructures and future imaginaries. Ultimately, she is interested in exploring how thoughtful, resistant technological innovation and critical literacy can equip students and practitioners with the reflexive frameworks necessary to dismantle embedded default biases, navigate systemic friction, and co-author more plural, non-normative speculative futures.

The Alienation of Copyright from the Author in the Era of Artificial Intelligence

Bogdan Stojanović
Junior Researcher
Institute of Comparative Law
Email: bogdanst622@gmail.com

Viewed through the lens of general textbook definitions, as well as through the populist discourse of numerous directors of intellectual property offices, politicians, and prominent artists, copyright is traditionally presented as a legal mechanism that affirms and protects the individual creative contribution of the author, providing both legal certainty and economic support. In this sense, it is seen as an instrument that encourages creativity and contributes to the cultural development of society as a whole. Such an understanding emphasizes that copyright not only protects the individual author but also serves a broader social function: fostering cultural diversity, increasing investment in cultural industries, and contributing to economic growth by developing markets for intellectual goods and services, including the preservation and commercialization of cultural heritage.

However, in the contemporary era of accelerated globalization, digitalization, and the development of an information society based on a knowledge economy, there is an increasingly evident tendency for copyright to lose its primary, normatively idealized function of protecting authors, as well as its secondary functions oriented toward the common good. Instead, it is increasingly transformed into an instrument that predominantly reflects the market and monetary interests of the most influential actors of contemporary digital capitalism, primarily large multinational technology corporations, in whose portfolios a vast number of copyrighted works and digital content are concentrated, including software products and data.

This transformation of copyright can be described as a process of its alienation from the author themselves, whereby the author increasingly occupies the position of a "weaker contracting party" in relation to corporate entities that possess the financial, technological, and organizational resources necessary for the efficient global commercialization of copyrighted works. In such a power structure, the author's ability to independently manage, distribute, and economically valorize their work becomes significantly limited, while intermediary platforms and technological systems assume a central role in this process.

Additionally, the contemporary information ecosystem is characterized by a phenomenon in which an excessive amount of information—copyrighted works being, by their nature, information itself—generates a deficit of attention. In such conditions, the successful dissemination and visibility of most copyrighted works on the market become almost impossible without the mediation of large digital platforms and corporations that control content distribution and algorithmic visibility. This further consolidates their role as key intermediaries between authors, whose primary motivation is the creation and communication of works to the public, and audiences, whose consumption of cultural content is

increasingly shaped by platform logic and algorithmic selection.
The development of generative artificial intelligence represents a further and more intensive phase of this process. AI systems are capable of generating new content based on existing copyrighted works, thereby further blurring the boundary between original human creativity and technically generated output. In this context, the author becomes even more distant from their work, both in terms of control over its use and in terms of participation in its economic value.
This paper aims to examine whether generative artificial intelligence represents a completely new challenge to copyright law or whether it can rather be understood as the latest phase of a long-term historical process of distancing the author from actual control over their work, while simultaneously reconsidering the possibility of preserving or redefining the space of authorial autonomy within contemporary technological and legal frameworks. Through the analysis of current legal disputes, theoretical debates, and economic relations in the digital economy, the paper seeks to contribute to a deeper understanding of the position of the author in the age of artificial intelligence, as well as to indicate possible directions for the development of copyright protection models that could lead to a more equitable relationship between authors and holders of technological and economic power.

Keywords: copyright law, authorial alienation, digital capitalism, platform economy, generative AI

Bogdan Stojanović is currently a junior researcher at the Institute of Comparative Law. He is also a PhD candidate at the Faculty of Law, University of Belgrade. His main fields of academic interest include copyright law and legal philosophy. His doctoral research is focused on the development and contemporary challenges of film copyright law. His academic work is characterized by an interdisciplinary approach to the study of copyright law, integrating legal theory, legal history, and philosophical analysis.

Decolonizing Transhumanism: Unequal Biotechnological Futures in the Age of Digital Colonialism

Nazam Laila
Assistant Professor
School of General Education, BRAC University
Email: nazam.laila@bracu.ac.bd

This paper delves into the intersection of biotechnology, transhumanism, and digital colonialism, highlighting how uneven technological development and access to private data perpetuate and amplify existing inequalities between technologically advanced and underdeveloped nations. Biotechnology, a cornerstone of transhumanist aspirations to enhance human capabilities, operates within a digital ecosystem dominated by powerful multinational corporations and governments in high-income countries like the United States and China.

The paper critiques how digital colonial practices shape the biotechnological sphere, driven by Western Enlightenment ideals of progress and rationality that shape the transhumanist agenda. For instance, transhumanism's quest for "super-longevity," "superintelligence," and "super well-being" fosters a new social hierarchy by driving data exploitation, denying equal access to digital resources, and creating a monopoly over research agendas. Technologies like CRISPR gene-editing, synthetic biology, and AI-driven drug discovery, developed primarily in digitally advanced countries, remain inaccessible to underdeveloped nations due to restrictive patents and intellectual property laws. This structural dependency ensures biotechnological advancements cater to the priorities of wealthier nations—such as anti-aging therapies or cognitive enhancement—while neglecting the urgent healthcare and agricultural needs of marginalized populations. This dominance mirrors and extends colonial patterns of resource extraction, exploitation, and cultural erasure, embedding systemic inequalities into the emerging biotechnological landscape.

Addressing these inequities demands a decolonial approach that promotes equitable access to transhumanist biotechnological innovations, fosters inclusive governance, and recognizes the sovereignty of marginalized communities over their biological and digital resources. By advocating for ethical frameworks based on equity, cultural diversity, and shared benefits, this paper contributes to the discourse on decolonizing technology and fostering global justice in the age of transhumanism.

Keywords: transhumanism, biotechnology, biosecurity, decolonization

Nazam Laila is an assistant professor at the School of General Education, BRAC University. She is also a member of the Technology and Society Certificate Committee of the Global Higher Education Alliance for the 21st Century (GHEA21, formerly OSUN), an Academy Associate at the Queen Elizabeth II Academy and Chatham House's Digital Society Programme, a research member of the Digital Research Network at SOAS University of London, and an AI policy expert on Bangladesh's National AI Steering Committee under the ICT Ministry. She co-developed the first draft of Bangladesh's National AI Policy (2026–2030) and is internationally recognized for her work on the political economy of AI and geopolitics. Dr Laila completed her PhD in Politics and International Studies at SOAS University of London as a Commonwealth Doctoral Scholar. She also received the Open Society University Network Fellowship at the Queen Elizabeth II Academy at Chatham House. She earned an MSc in Gender, Development and Globalisation from the London School of Economics and Political Science and completed her BA (Honours) and MA at BRAC University in Interdisciplinary English Literature, graduating with highest distinctions and receiving the Vice Chancellor's and Dean's Awards.

AI Shame in Gendered Hierarchies of Academic Subjectivities

Katarina Bogićević
MA Student
Faculty of Political Science, University of Belgrade
Email: bogicevickatarina1@gmail.com

Experiences of shame, expressed as individual lack of capability or virtue, are often seen as detached from histories of intersecting gendered, classed, and racialized injustices. Similarly, current discourse on the emerging concept of "AI shaming" in academia largely overlooks how academic legitimacy and authority are already organized around the normative subject who embodies specific traits traditionally associated with masculinity and whiteness. Consequently, shame surrounding AI use circulates unevenly across differently positioned bodies and disciplinary contexts. The pursuit of two dominant yet conflicting academic standards—rigor as a boundary-marker of belonging and productivity as a demand of neoliberal capitalist logics—is now reconfigured by the rapid development of AI. These dynamics intensify the visibility of deviation from the evolving ideals of what constitutes a "good" scholar. In that sense, those positioned furthest from the normative academic subject are differentially exposed to heightened (self-)scrutiny and to shame, which may operate as a punitive rather than transformative affective mechanism. Instead of treating AI-related shame as a neutral or universal phenomenon, I ask how it is produced, experienced, and negotiated across existing hierarchies of academic subjectivities, with particular attention to the role of gender as a structuring principle. To explore this question empirically, the paper draws on narrative interviews with scholars and researchers at the University of Belgrade. In doing so, the study aims to contribute to a more relational understanding of shame and practices of AI shaming in academic settings, while situating them within broader structural conditions. Without such an approach, debates on the ethical and responsible design and integration of AI in academia risk obscuring and reinforcing different forms of inequality.

Keywords: AI, shame, academia, gender

Katarina Bogićević is pursuing an MA in *Peace, Security, and Development* at the Faculty of Political Science, University of Belgrade, where she also earned her BA in International Relations. Her research focuses on critical approaches to digital security, exploring how digitally mediated practices reshape and reinforce power relations through phenomena such as AI-facilitated gender-based violence, far-right digital ecosystems, and various forms of digital authoritarianism. Currently, she is a research intern at the Digital Society Lab of the Institute for Philosophy and Social Theory, where she investigates the role of affect in anti-gender mobilizations on Instagram. She has previously been involved with several research institutions, including the Centre for International Security (CIS-FPS), the Belgrade Centre for Security Policy, and the European University Institute.

ART, CULTURE, AND CRITICAL AI PRACTICES

No to AI: How Contemporary Artists Sabotage the AI Art World

Ana Ćemalović
Artist and Independent Researcher
University of Arts in Belgrade
Email: anacemalovic@outlook.com

Generative AI models have changed the ways in which artistic work is produced, created, and received, and have enabled the exploitation of artistic work through the unauthorized use of training data (used without consent or compensation) and the arbitrary replication of individual artists' styles through style mimicry. These data are often private, sensitive, and copyrighted. For example, it was discovered that the LAION-5B training dataset contains a huge number of images of artworks by contemporary artists that were collected there without the explicit consent of the artists. This leads to a series of direct and indirect harmful effects on the artists themselves in terms of impact on their reputation, reduction in income from the sale of works of art, loss of artistic identity, or rather identity theft. There is also the question of job displacement and precarious work: today's artists often correct and refine AI-generated products that mimic their own work, which is at the same time altered through AI image-editing tools.

These new circumstances of unequal and ruthless exploitation of artistic work demoralize, disenfranchise, and devalue artists and their artistic work. That is why various systems have been developed that aim to protect the interests of artists by disrupting, undermining, and disabling the generative processes of creating works of art through artificial intelligence in various ways, through the integration of elements into images of artworks that are invisible to the human eye but disruptive to AI models.

This work will consider different technical-artistic strategies, campaigns, protocols, conceptual interventions, artistic sabotages, artistic and technical tactics like data poisoning and protective sabotage that emphasize issues like data scraping without consent, copying of style, as a form of digital resistance, or digital civil disobedience that raise awareness of the unethical use of artistic work. These strategies have raised the visibility of the problems that contemporary artists face in their daily work, have influenced ongoing court cases and initiated legislative discussions, and have also influenced the formation of legal acts in the European Union and the United States. The paper considers the theoretical implications of such actions, their shortcomings and limitations, as well as their broader significance in terms of keeping the questions of art ethics alive and central. These tactics transform artists into active participants who exert socio-political pressure on the process of rapid technological disruption and ensure that they regain some control and protection of their work by preventing unauthorized training of networks based on their own work. They also raise questions about the nature of art itself, the value of human feelings, originality, motives for creation, and struggle. Also, this presentation will consider whether some of these techniques can help protect the copyright and other rights of artists, their dignity and integrity, as well as whether and what significance they can have in contributing to better regulation and having a positive impact on the art world

in the AI era.

Keywords: AI art, subversion, sabotage, activism

Ana Ćemalović is an independent artist and arts and media theorist. She graduated from the Faculty of Applied Arts at the University of Arts in Belgrade and completed the joint master's program "Languages, Business and International Trade" of the University of Belgrade and the University of Orléans, France. Currently, she is a doctoral student in the Department of Arts and Media at the Center for Interdisciplinary Studies, University of Arts in Belgrade, where she is completing her PhD thesis, "Arts as a Machine in the Context of Artificial Intelligence." She also exhibits her work, participates in conferences, and publishes articles in scientific journals. She has been a member of the Applied Artists and Designers Association of Serbia (ULUPUDS) since 2013.

Artificial Intelligence Governance Toward Cultural Institutions: Ethical and Sustainable Frameworks

Biljana Mickov
Culture Researcher
University of Reims Champagne-Ardenne; Cultural Center of Vojvodina
Email: biljamickov@gmail.com

The rapid development of artificial intelligence (AI) is fundamentally transforming the ways in which cultural institutions such as museums, libraries, archives, and galleries manage, preserve, interpret, and provide access to culture. AI technologies, including machine learning, computer vision, natural language processing, and generative AI, are increasingly applied to collection management, digitization, metadata generation, conservation monitoring, multilingual accessibility, and audience engagement. While these technologies offer significant opportunities to improve operational efficiency and democratize access to cultural resources, they also raise complex governance challenges concerning transparency, accountability, algorithmic bias, intellectual property, data governance, cultural representation, and the protection of sensitive cultural knowledge. Consequently, the adoption of AI within cultural institutions requires governance frameworks that integrate ethical principles, legal standards, institutional responsibility, and human oversight to ensure that technological innovation remains aligned with the public mission of cultural institutions.

This paper explores AI governance as a strategic and interdisciplinary framework for supporting the responsible implementation of AI in cultural institutions. It argues that AI governance should extend beyond technical regulation to include ethical decision-making, stakeholder participation, risk management, data stewardship, and institutional accountability. International policy initiatives, particularly the UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) and the European Union Artificial Intelligence Act (2024)**,** provide an important normative foundation for developing governance models capable of balancing technological innovation with the protection of cultural diversity, human rights, and democratic values. Within the cultural sector, governance must also address questions of authenticity, provenance, intellectual property rights, and equitable representation of diverse communities whose culture is increasingly mediated through AI-driven systems.

The research adopts a qualitative methodology based on a critical review of scientific literature, international policy documents, and selected examples of AI implementation in museums, libraries, and archives. By comparing emerging governance frameworks and institutional practices, the study identifies key governance principles that promote transparency, fairness, accountability, sustainability, and public trust. Particular attention is given to governance mechanisms that ensure explainable AI, responsible data management, protection of cultural heritage data, and meaningful human oversight throughout the AI lifecycle.

The study is guided by the following research questions: What are the principal governance challenges associated with implementing AI in cultural institutions? How can ethical and legal governance frameworks support responsible AI

adoption while preserving cultural values and institutional integrity? And how can cultural institutions design governance strategies that strengthen public trust while enabling responsible digital transformation?
The primary aim of this paper is to develop a conceptual framework for AI governance tailored to the specific needs of cultural institutions. The research seeks to identify governance risks and opportunities, evaluate existing international governance models, and formulate recommendations for policymakers, cultural managers, heritage professionals, and researchers. The paper contributes to the growing interdisciplinary debate on AI governance by demonstrating that effective governance is essential for ensuring that AI supports rather than undermines the preservation, accessibility, authenticity, and social value of culture. Ultimately, responsible AI governance should enable cultural institutions to embrace digital innovation while safeguarding ethical principles, cultural diversity, and the public interest.

Keywords: artificial intelligence governance, cultural policy, cultural institutions, ethical AI

Biljana Mickov is a cultural researcher and professor. She graduated from the University of Novi Sad. She also holds the European Diploma in Cultural Project Management (Marcel Hicter Foundation), Brussels (supported by UNESCO and the European Commission). She holds a PhD from the University of Reims Champagne-Ardenne, France. She has published more than ten books in the United Kingdom dedicated to the cultural sector, cities, and innovation. Her work has been published in Barcelona, the United Kingdom, France, China, and Serbia. She was part of the European Expert Network on Culture (EENC), an initiative of the European Commission's Directorate-General for Education and Culture. The project comprises a group of high-level experts in the field of cultural policy who provide reports and studies. Her work is primarily dedicated to research and development of cultural policies and innovative management, with a focus on inclusive society. Mickov's fields of research include cultural policy, arts and culture management, sociology, human geography, the creative economy, and the green agenda. She is a member of an expert group at the European Commission.

Musical AI and the Question of Genre

David Gehring
PhD Candidate
Cultural Studies, George Mason University
Email: dgehrin@gmu.edu

This paper argues that the emergence of AI music and the logics that motivate its development will further ongoing processes and effects of the depoliticization of music via the decontextualizing and uprooting of genre, thereby contributing to the constellation of data-driven objectification and cultural optimization (Morris, 2020).

Genres are more than a means of market organization; they are also informed by socio-political histories, rooted in place, constituted by shared practices that facilitate the creation and maintenance of safe spaces for marginalized communities, and the insulation and protection of traditions and values shared by these communities. As Glick (2023) emphasizes, genres "function as imagined worlds for various identity groups to build community," and are "implicitly political because [they] provide a 'safe space' for these people to represent themselves." In other words, genres are a means of both sonic and social ordering, two dimensions that exist in a relational dynamic through which the style and performance of music and socialities/identities are developed together, against, and through forces of history. As noted in Michael Denning's *Noise Uprising: The Audiopolitics of a World Musical Revolution* (2015), "The millions of migrants who arrived in the colonial ports were a critical mass that enabled the explosion of new plebeian musical cultures in the first two decades of the twentieth century," a development that sowed the seeds for the eventual emergence of jazz. As such, inquiries into the origins and histories of *sonic* genres necessarily involve their socio-political histories.

The proliferation of streaming and the effects of disintermediation and algorithmic curation have obscured and diluted genre maps, leading many to claim that we have entered into a post-genre era (Muchitsch & Werner, 2024). However, the political implications of this claim are rarely critically considered. Generally, it can be understood in connection with similar notions of post-identity (i.e., post-racial, post-gender) and the intensified marketization of individuality and difference (Muchitsch & Werner, 2024), characteristic of the logic of neoliberalism. More specifically, the implications of this development threaten the further neutralization of music's political and social dimensions, dovetailing with logics that emerged in the post-Napster music industry, informed by an economic rationality that, for example, neglected the communal implications of "selling out," thus rendering it meaningless in the context of the current environment (Klein, 2020).

Drawing from critical AI scholarship (White, 2025; Gregg & Strangers, 2024) and platform studies (Bonini & Magaudda, 2023), I contend that these developments are the effects of strategic rationalities that render music as content and genre as context. Critically, the development of musical AI, in addition to the aforementioned processes relating to algorithmic curation and the "post-genre

era" of popular music, within the broader impacts of the rationalities of platform economies, compounds the intensification of political neutralization and cultural/historical erasure.

Keywords: artificial intelligence, musical artificial intelligence, genre, music, music culture, politics of music

David Gehring is originally from central Florida. Davidearned his MA in Digital Humanities at Old Dominion University. His decision to attend George Mason University is the consequence of a substantial tenure as a musician in professional music industries, during which he was fortunate to work at the independent and major-label levels, touring internationally. These experiences motivated his critical curiosity, which developed into a focus on questions that emerge within the intersection of music, culture, politics, and industry. He currently resides with his family in Virginia.

Can AI Change the Reception of “Spomeniks”?

Sonja Jankov
Independent Research Associate
Email: jankovsonja@gmail.com

Yugoslav memorials were erected at sites where thousands of civilians were killed or deported to concentration camps, and where barely armed local peasants fought violent fascist militias. At the same time, they are monuments to the victims of atrocities, as well as monuments to resistance, anti-fascist struggle, resilience, memory, and collective mourning. They were visited by numerous international delegations and hundreds of schoolchildren. However, after the dissolution of the SFRY, and especially following the trials conducted by the International Criminal Tribunal for the Former Yugoslavia, many of these memorials were neglected, depoliticized, or destroyed. Popular culture began to perceive them as Spomeniks, alien art, UFOs, parkour training grounds, and sets for sci-fi films and music videos. They became “concrete clickbait” (Hatherley, 2016), an “exotic Other” (Kulić, 2018) that, “no longer ideologically dangerous, can be enjoyed for its visual effect” (Mrduljaš, Kulić, & Jovanović, 2013).
In the age of AI, when photorealistic images of monuments no longer reliably depict reality (Chan, 2025), new monuments can be generated using tools such as Brutalist Architecture Generator and shared in AI Brutalism forums and groups. Today, images of Yugoslav memorials often appear on social networks accompanied by the disclaimer: “The building is real and not AI.” However, this paper argues that AI can also be embraced as a means of interpreting, reconstructing, and experiencing Yugoslav monuments, applying AI in museology, memory culture, artistic research, and education. As these monuments were designed to be traversed, functioning as forms of spatial storytelling (Kulić, 2026), AI can also serve as a tool of inclusion, bringing them closer to persons with disabilities.
The paper discusses recent projects on the topic, such as Ana Miljački’s *The Pilgrimage* at the Museum of Yugoslavia (2025), which includes an animation created with the help of the AI system “Stane” (StyleGAN3) at the Critical Broadcasting Lab at the Massachusetts Institute of Technology, combining personal spatial experience with archival materials. AI tools can generate interactive 3D models, VR experiences, multimedia guides, stylistic variations, and animated versions, all of which reinterpret the monuments and help audiences experience them as important cultural heritage.

Keywords: Yugoslav memorials, AI memorial tourism, arts, heritage

Sonja Jankov holds a PhD in Art and Media Theory. She has worked at the Academy of Arts, University of Novi Sad, and the Faculty of Dramatic Arts, University of Arts in Belgrade. As a workshop and short-course lecturer, she has collaborated with the University of Turin, Italy, and the Arts Academy in Split, Croatia. Her recent publications include the monograph *Understanding Intermedial Quoting: Architecture of Yugoslav Modernism as Quotation within Contemporary Art* (Center for Contemporary Culture and Communication, Novi Sad, 2024) and the volume *Empowering Participation in Culture and Architecture: Activating Public Resources for and with Community* (Faculty of Dramatic Arts, Belgrade, 2025), co-edited with Nina Mihaljinac.

Synthetic Enunciation as Counter-Narrative: The Satirical Deepfake in The Dor Brothers' *Riot*

Jorge Franganillo
Professor and Researcher
Faculty of Information and Audiovisual Media, University of Barcelona
Email: franganillo@ub.edu

Julieti Sussi de Oliveira
Professor and Researcher
Faculty of Information and Audiovisual Media, University of Barcelona
Email: julietideoliveira@ub.edu

Lydia Sánchez
Associate Professor
Faculty of Information and Audiovisual Media, University of Barcelona
Email: lsanchezg@ub.edu

In today's media ecosystem, generative artificial intelligence has radically redrawn the boundaries of cultural critique. Far from acting merely as a passive vehicle for disinformation, the deepfake emerges as a powerful apparatus of "synthetic enunciation." Using the viral short film *Riot* (The Dor Brothers, 2025) as a case study, this research explores the broader potential of deepfakes as a potent medium for political satire and their capacity to articulate visual counter-narratives that directly challenge the official discourses of political, economic, and technological elites.

To achieve this, the research poses two guiding questions: What satirical resources can the deepfake, as an expressive medium, deploy to fuel its critique of the establishment? And what archetypes can it utilize to represent public figures in order to amplify a message that transcends factual reality?

From a theoretical perspective, we break down this "synthetic enunciation" into three axes:

Enunciation as performance, where the deepfake functions as a performative speech act; the act of fabrication itself is the message.

The emancipation of the enunciator, which transcends the indexical trace of reality to grant the creator absolute control over staging and gesture, turning leaders into puppets.

Enunciative complicity, a "reading contract" where both creator and viewer acknowledge the image's factual falsity to unveil a deeper ideological truth.

Grounded in this conceptual framework, we conducted a shot-by-shot qualitative analysis of *Riot*, categorizing its visual content, satirical devices, their intended purposes, and the archetypes assigned to specific figures to examine how synthetic gestures and scenography subvert traditional codes of leadership.

Results indicate that the short film visually materializes the most extreme metaphors of contemporary polarization through hyperbole and absurd juxtaposition. Elites are stripped of their institutional roles and thrust into a hyperreal battlefield: Donald Trump mutates into a hyper-masculinized authoritarian war-

lord, while figures such as Mark Zuckerberg and Elon Musk epitomize the rise of a post-ideological "broligarchy," a fusion of populist authoritarianism and techno-capitalist dominance where the aesthetics of power supersede traditional loyalty. Similarly, progressive figures like Kamala Harris are framed using tropes of urban looting, exposing the fracture of the social contract. The elites' total detachment crystallizes in the frivolous selfie of the Trumps against a burning nation, a final image of the precession of simulacra, where the performance of the crisis replaces the crisis itself.
In conclusion, the visual analysis of *Riot* demonstrates how the satirical deepfake leverages the provocation of generative code to warn against the mutation of politics into a Baudrillardian spectacle of simulation. Depicting implausible scenarios through AI confirms that this format does not seek to pass as real; rather, it thrives in hyperreality and visual performativity, stretching the image to the point of absurdity through a hyperbolic enunciative gesture. Ultimately, the emancipation of the creator through the deepfake establishes a new maxim in contemporary critical media: the art of lying through form to reveal a truth in substance.

Keywords: synthetic enunciation, deepfake satire, political counter-narrative, visual performativity, artificial intelligence

Jorge Franganillo holds a PhD in Information and Communication. He is a lecturer in Information Science and Audiovisual Communication at the University of Barcelona. His teaching and research focus on content creation, audiovisual documentation, and generative AI. Within this latter field, he has extensively explored the impact of synthetic media tools on media content production. He is a researcher at the Information, Communication and Culture Research Center (CRICC), a member of the Teaching Innovation Group in Communication and Audiovisual Media (In-COMAV), and a member of the Horitzó IA-UB think tank. His research excellence has been recognized by AQU Catalunya with a six-year research merit. Throughout his career, he has advised public administrations, library networks, and educational services on the strategic and ethical integration of technology to optimize workflows and foster innovative learning models.

Julieti Sussi de Oliveira holds a graduate degree in Journalism from the University of Caxias do Sul, Brazil, and a PhD in Communication from the University of Seville. She is a lecturer in the Audiovisual Communication degree program at the University of Barcelona. She is also a researcher at the Information, Communication and Culture Research Center (CRICC) and a member of both the Teaching Innovation Group in Communication and Audiovisual Media (In-COMAV) and the Research Group on Political Communication, Ideology and Propaganda (IDECO). Her research interests focus on communication and culture, the political economy of communication, communication and power, media literacy, and disinformation.

Lydia Sánchez holds a BA in Philosophy from the University of Barcelona and an MA and PhD in Philosophy from Stanford University. She is an associate professor of Audiovisual Communication at the University of Barcelona, where she also serves as vice dean for Research and General Affairs at the Faculty of Information and Audiovisual Media. She is the principal investigator of the consolidated research group DHIGECS-Com (Didactics of History, Geography and other Social Sciences—Communication) and a member of the Teaching Innovation Group in Communication and Audiovisual Media (In-COMAV). Additionally, she is a member of the Information, Communication and Culture Research Center (CRICC), where she coordinates the research line on Communication, Democracy and Cognition. Dr. Sánchez has led and participated in numerous national and international research projects. Her research excellence has been recognized by AQU Catalunya with three six-year research merits. She has presented her work at various national and international conferences and currently serves as vice-chair of the ECREA Philosophy of Communication section.

DIGITAL DEMOCRACY, GOVERNANCE, AND TECHNOLOGICAL POWER

The Record That Cannot Answer: Algorithmic Transparency Infrastructure and the Architecture of Uncontestable AI

Hon Bor So
Independent Researcher
iHuman Lab
Email: singso78901@gmail.com
ORCID: https://orcid.org/0009-0008-2768-7494

Across Europe and beyond, AI governance regimes have responded to public demands for democratic accountability by mandating transparency records—structured public disclosures describing how algorithmic systems work, what data they process, and what human oversight mechanisms exist. These records carry an implicit governance promise: that citizens, oversight bodies, and democratic institutions will be able to scrutinize AI-mediated decisions and act on what they find. This paper asks whether that promise is architecturally achievable. Do transparency records preserve enough information for an affected person to reconstruct what happened in a specific decision—who acted, on what basis, under what uncertainty, and through what oversight gate?

The empirical focus is the UK Algorithmic Transparency Recording Standard (ATRS), the primary public transparency infrastructure currently mandated for public bodies deploying AI in high-stakes domains including welfare, health, criminal justice, immigration, and fraud detection. Introduced by the UK government to enable public scrutiny and institutional accountability for algorithmic tools affecting citizens' rights and life outcomes, the ATRS represents one of the most structured AI transparency schemes in operation and provides a concrete case for assessing whether transparency mandates deliver their accountability promise.

This paper reports the findings of a structured audit of 46 high-stakes ATRS records, selected from a validated corpus of 131 records on the basis of direct individual consequence across eight public administration domains. Each record was evaluated against a six-field reconstructability rubric, assessing whether the record preserved minimum conditions for post-event reconstruction: actor identity at decision-event level, affected subject identity, the specific trigger for the algorithmic output, the evidential basis, acknowledged uncertainty, and the specific oversight gate applied.

The findings are consistent across all 46 records and all eight domains. No record identifies who acted at the level of the specific decision event—all 46 name an institutional Senior Responsible Owner, but none identifies the individual whose judgment became accountable action in a specific case. Uncertainty at the point of decision is absent in 63% of records. Event-level oversight conditions—the specific criteria governing human review in a specific instance—are specified in only 4 of 46. These are not failures of individual compliance. They are the predictable output of a schema designed to make tools visible, not to preserve decision events.

The paper introduces the distinction between *visibility* and *decision-event reconstructability*, arguing that the ATRS produces what it terms thin transparency: records that are public, structured, and formally compliant, but that cannot sup-

port the minimum epistemic conditions for meaningful contestation. The problem is architectural rather than motivational. The schema requires description at tool level but does not mandate event-level fields. No improvement in compliance within the existing schema can close this gap. Schema redesign is required. The political stakes extend well beyond technical record design. When an affected citizen, civil society body, or parliamentary committee attempts to scrutinize an AI-mediated decision, they encounter records that offer the appearance of accountability without the substance required for contestation. Governance frameworks invite the public to challenge decisions through infrastructure that is architecturally incapable of answering the questions meaningful challenge demands. Record architecture encodes assumptions about who needs to know what, and at what level of granularity. Those assumptions are political choices. Their consequences shape which democratic futures with AI remain possible—and which are foreclosed before contestation begins.

Keywords: algorithmic transparency, decision-event reconstructability, AI governance, accountability infrastructure, contestability

Hon Bor So is an independent researcher based in Greater Manchester, UK. His research program, Boundary Phenomena in Sociotechnical Decision Systems, examines how decision boundaries form, transfer, and remain auditable across systems where human judgment is mediated by AI, workflow, and institutional record architecture. His empirical work spans mandatory reporting and transparency regimes in UK public-sector AI governance, autonomous vehicle regulation, consumer financial protection, and public-sector welfare administration. He is a member of the AIES 2026 Programme Committee and has work accepted at the Society for Social Studies of Science (4S) annual conference in Toronto, October 2026. He is currently preparing for a PhD by publication route.

Critical, Collaborative, Enabling: Influencers as Distributing Nodes of Citizenship Practices in a Constrained Hybrid Media System

Trang-Nhung Pham
PhD Student
Karlstad University
Email: trang-nhung.pham@kau.se

In hybrid media systems, social media influencers have emerged as significant actors in political debates, occupying a position between ordinary citizens and news actors. Yet how these platform-afforded civic actors conduct citizenship practices, particularly in non-democratic contexts, remains underexplored. The existing literature on informed citizenship online often focuses on ordinary users in democratic contexts (Gagrčin et al., 2022; Schnaudt et al., 2024) while overlooking other new media actors like influencers, who have been an emergent subject in media studies with acknowledged impact on information flow and political discussions (Harff & Schmuck, 2023; Riedl et al., 2023). This paper argues that in Vietnam's hybrid media system, citizenship is distributed across a range of actors beyond formal institutions. Social media influencers, with their platformed power, can distribute information about public affairs, simultaneously enacting different citizenship practices and making visible particular models of what informed citizenship looks like. Exploration of their practices can highlight how these platform-afforded civic actors perform informed citizenship in a constrained media environment.

This study fills this gap in understanding the roles of influencers as powerful nodes within distributing networks of informed citizenship in a non-democratic context through a qualitative content analysis of Vietnamese influencers' discussions about a green transformation policy aimed at banning gasoline-powered motorbikes from the central area of Vietnam's capital city of Hanoi on Facebook, the most popular social media platform in Vietnam. It took an abductive approach with the framework of discursive citizenship norms (Gagrčin et al., 2022) as the initial analytic lens to guide the data analysis of citizenship practices while remaining open to modifications if the data warranted it. This process developed a modified framework to analyze informed citizenship practices across three dimensions: information, discursive practices, and actions on display. Together, the three dimensions capture influencers' citizenship practices: what information they share with the audience, how they critically engage with the issue and other actors, and which actions they model when informed.

Analysis of how Vietnamese influencers presented informed citizenship through social media content and engagement showed a widely uneven distribution of citizenship practices across different groups of influencers. These variegated manifestations of informed citizenship highlight influencers' roles as both ordinary citizens and alternative sources of information. First, influencers' citizenship practices in a constrained hybrid media system showed critical but collaborative engagement in political discussions. Second, examination of influencers as platform-afforded civic actors showed variegated manifestations of democratic informed citizenship norms in practice. Third, influencers' citizenship practices

in connection with other actors fulfilled certain journalistic roles except for the adversarial function. Finally, influencers' posts can enable discursive citizenship practices among the audience. This study contributes to ongoing debates about how the notion of informed citizenship has evolved in digital environments. By conceptualizing influencers' practices as both citizenship behaviors and enabling nodes for ordinary citizens to perform their civic participation, the study advances understanding of how citizenship is negotiated in authoritarian hybrid media systems. This contribution enriches knowledge on civic agency in the age of platformization.

Keywords: influencer, hybrid media system, informed citizenship, authoritarian system

Trang-Nhung Pham is a PhD student in Media and Communication Studies at the Department of Geography, Media & Communications at Karlstad University in Sweden. She is also a member of the Ander Center for Research on News and Opinion in the Digital Era (NODE) and an affiliated PhD student at Lund University's Graduate School in Asian Studies. Her research focuses on roles and practices of social media influencers in non-Western environments. Pham holds a Bachelor's degree in International Relations from the Diplomatic Academy of Vietnam and a Master's degree in Journalism, Media and Globalisation from the Erasmus Mundus Master's program at Aarhus University and the University of Amsterdam. Prior to her academic career, she spent more than eight years working in the media field in Vietnam.

Cybernetic Panopticism: Surveillance, Subjectivity, and Contested Futures in the Age of Algorithmic Governance

Bernat Torner
PhD Candidate
Department of Humanities, Universitat Pompeu Fabra
Email: bernat.torner01@estudiant.upf.edu

Contemporary debates on artificial intelligence and digital infrastructures are frequently framed through the concepts of surveillance capitalism, algorithmic governance, and datafication. While these approaches have significantly advanced our understanding of contemporary power, they often leave unresolved the question of how older disciplinary forms of surveillance relate to emerging cybernetic and algorithmic modes of control. Is panopticism still a useful analytical framework for understanding contemporary technological power, or has it become obsolete in the age of AI?

This paper argues that panopticism remains relevant, but only if it is fundamentally reformulated. Building on Michel Foucault's analyses of discipline, governmentality, and neoliberalism, as well as on Gilles Deleuze's notion of societies of control, I propose the concept of *cybernetic panopticism* as a framework for understanding the political rationality underlying contemporary systems of algorithmic governance.

The paper proceeds in three steps. First, it revisits Foucault's account of panopticism as a technology of visibility, normalization, and self-regulation. While Foucault later suggested that Bentham's *Le Panoptique* (1780) became increasingly inadequate for describing liberal forms of government focused on population management, the paper argues that the panoptic principle itself—the production of self-regulating subjects through conditions of visibility—remains central to contemporary power.

Second, the paper engages with post-panoptic surveillance theory. Drawing on Deleuze's societies of control, Mark Poster's informational subject, Haggerty and Ericson's surveillant assemblage, and Shoshana Zuboff's surveillance capitalism, among others, it examines how surveillance has shifted from enclosed disciplinary institutions toward distributed infrastructures of data extraction, algorithmic classification, and predictive intervention. In these accounts, individuals increasingly appear as informational entities constituted through digital traces, behavioral data, and algorithmic profiles.

Third, the paper introduces the notion of *cybernetic panopticism* through a dialogue with "the cybernetic hypothesis" developed by Tiqqun and later expanded by the Invisible Committee. Cybernetics is approached not merely as a technical field but as a governmental rationality centered on feedback, communication, prediction, and system regulation. From this perspective, contemporary digital infrastructures can be understood as mechanisms that continuously monitor, model, and modulate behavior in order to maintain social and economic order.

Cybernetic panopticism describes the convergence of three historical processes: the panoptic production of visibility and self-regulation, the cybernetic logic of feedback and prediction, and the neoliberal production of entrepreneurial sub-

jectivities (*homo oeconomicus*). Unlike the classical Panopticon, cybernetic panopticism does not depend on centralized observation or enclosed institutions. Instead, it operates through distributed networks of sensors, platforms, databases, and algorithms that transform individuals into readable, predictable, and governable informational subjects.
By conceptualizing contemporary surveillance through this lens, the paper contributes to ongoing debates on technological power and contested digital futures. Rather than viewing AI and digital infrastructures as neutral tools, it highlights their role in producing specific forms of subjectivity and governance. At the same time, the concept of cybernetic panopticism opens a critical space for reflecting on resistance, opacity, and alternative technological futures beyond the imperatives of visibility, optimization, and control.

Keywords: cybernetic panopticism, surveillance capitalism, algorithmic governance, entrepreneurial subjectivity, societies of control

Bernat Torner is a PhD candidate in Political Philosophy at Universitat Pompeu Fabra (Barcelona, Catalonia). His research lies at the intersection of political philosophy, critical social theory, and surveillance studies, focusing on contemporary transformations of power and subjectivity in digital societies. Drawing on critical approaches to biopolitics, cybernetics, and technology, he examines how algorithmic governance, data infrastructures, and artificial intelligence reshape social control, political agency, and human experience. His current doctoral research develops the concept of *cybernetic panopticism* to analyze emerging forms of surveillance, governance, and subject formation in data-driven societies. In his forthcoming article, "Cybernetic Panopticism: A New Model of Functioning for Societies of Control" (*Surveillance & Society*), he argues that algorithmic governance transforms the disciplinary logic of the classical panopticon into a cybernetic mode of power based on continuous data extraction, prediction, and behavioral modulation, contributing to contemporary debates on AI, digital governance, and the future of democracy.

We Don’t Need No Information Control: AI, Recommender Systems, and the Politics of Selection

Uroš Sergaš
Master of Data Science
Member, Centre for Responsible Artificial Intelligence, University of Primorska
Doctoral Student and Teaching Assistant, Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska
Email: uros.sergas@upr.si

Jar Žiga Marušič
Master of Cognitive Science
Doctoral Student, Faculty of Arts, University of Ljubljana
Teaching Assistant, Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska
Email: jarz.marusic@gmail.com

As artificial intelligence becomes one of the main ways people access and parse information, the question of control over public perception of reality becomes increasingly relevant. This contribution examines how machine-learning technology (recommender systems, search engines, content filters) shapes the flow and reception of information. This is exacerbated by the proliferation of LLMs, enabling AI-generated summaries to curate users’ informational ecosystems even further. Digital platforms can now curate content directly by training their LLMs to summarize information in specific ways. The purpose of this contribution is to explore the societal ramifications of the transition from hard (censorship) to soft (automated content curation) information control.

We propose that these changes go hand in hand with the fact that information scarcity has given way to information overload. When enormous amounts of text, images, and videos can be generated at the press of a button, information density starts to overtake the limits of human processing, necessitating a form of filtering. As recommender systems and similar AI tools choose from an ever-larger pool of information, they inevitably also exclude or make other information practically invisible. Attention, therefore, becomes the scarce resource, and the systems that manage attention become politically and epistemically powerful. Furthermore, the problem of echo chambers and epistemic bubbles indicates that information filtering does more than just organize “reality”: it curates a personalized perspective on the world that the user can mistake for the world as it is.

Our thought process is based on two central questions. First, what does AI-driven content curation mean for shared public knowledge and democratic debate? Second, what direction should society take if we are to navigate an information landscape shaped by zettabytes of newly generated data each year, without relying on direct censorship?

We address these questions by arguing that contemporary information control is becoming more widespread and more opaque. It is an integral part of our informational ecosystem, but because it is largely unintelligible to laypeople, we

remain mostly unaware of its scope. In this sense, it is like water to fish, seemingly natural because it is a constant aspect of our experience. The contemporary focus on misinformation risks losing sight of the fact that a monopoly on framing and contextualization can be more powerful than censorship and deception.
This is not to denounce the utility of these systems. The necessity of filtering is a natural consequence of information overload. The problem arises when filtering is unsupervised and opaque, such that the preferences of the user and values of society are overshadowed by the interests of those who control the "means of information production." This presents a risk of politicized filtering, leading to polarization and an exacerbation of societal power asymmetries.
In this context, freedom of information can no longer be understood only as the availability of content. It also depends on whether information, and the medium through which it is presented, are intelligible to consumers. Just as the process of natural selection systematically allowed for some genes to survive through time while others disappeared, information control systems can determine which pieces of information, or memes, stand the test of time and which are left behind, to disappear, and never to be found again by future generations.

Keywords: artificial intelligence, information control, recommender systems, censorship, personalization

Uroš Sergaš is a teaching assistant in computer science at the Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska (UP FAMNIT), and an affiliate of CeResAI UP. As a researcher specializing in recommender systems and computational social science, his work focuses on applying machine learning methods to address societal challenges, notably political polarization, which serves as his PhD topic. Together with Jar Žiga Marušič, he developed an interest in artificial intelligence and its impact on society. More specifically, they are concerned with finding a middle ground between pure technological acceleration and a cessation of further AI research to preserve human values.

Jar Žiga Marušič is a teaching assistant in psychology at the Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, and a PhD student in philosophy at the University of Ljubljana, working on the problem of ideological distortions (bias, censorship) in contemporary science and academia. His work integrates tools and concepts from modern social science and behavioral economics into the theory of ideology. His gateway into AI research was the realization that the analysis of social systems can draw upon concepts from AI alignment theory. There is a functional isomorphism between the normative orientations of institutions and artificial agents, enabling a rich interdisciplinary dialogue between alignment theory and the analysis of social institutions. That is precisely the perspective adopted in the works on AI alignment, written with Uroš Sergaš. Together, they highlight the intersections and parallels between the technical and social dimensions of AI technology.

The Brat Pack Effect: Wealth, Exclusion, and the Global Governance of Artificial Intelligence

Meem Arafat Manab
Researcher
Universidad Politécnica de Madrid
Email: meem.arafat@bracu.ac.bd

The global governance of artificial intelligence has produced a recognizable family of concepts describing how norms diffuse across borders. Anu Bradford's Brussels Effect identifies the European Union's capacity to regulate global markets through sheer regulatory gravity. The Beijing Effect, theorized by Erie and Streinz, describes how China shapes transnational data governance through infrastructure dependency and market access leverage along its Digital Silk Road. Chinmayi Arun's Silicon Valley Effect captures how dominant AI companies shape the transnational legal order to protect their business interests and deflect accountability. Each of these effects names a distinct actor, a distinct mechanism, and a distinct geography.

This paper proposes a fourth: the *Brat Pack Effect.*

The Brat Pack Effect describes the capacity of the G7 and its affiliated institutional architecture—principally the OECD and the Global Partnership on AI (GPAI)—to set de facto global AI norms not through regulatory coercion, infrastructure dependency, or corporate discourse capture, but through the self-legitimating authority of wealth. The G7 is not a regional body. It represents no coherent geography, no cultural anchor, no shared neighborhood, no common constituency beyond historical economic dominance. Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States span three continents. What unites them is money and the power that precedes and compounds it.

The mechanism operates through a circular laundering of elite consensus into apparent universalism. G7 summitry produces informal norm packages, most consequentially the Hiroshima AI Process (2023) and its Comprehensive Policy Framework, which are then institutionalized through the OECD, projected through GPAI, and eventually presented as an intergovernmental consensus. The OECD itself, as a club of high-income economies, does not interrupt this circularity; it amplifies it. The result is a governance architecture that is simultaneously non-binding and highly influential, voluntary in form and coercive in effect, presented as universal while structurally exclusionary.

The Brat Pack is also, with rare exception, a men's club. The AI Safety Summit series, from Bletchley Park (2023) and Seoul (2024) to Paris (2025) and Delhi (2026), and the G7 AI governance processes have been overwhelmingly shaped by male heads of state, male AI lab executives, and male-dominated expert bodies. Gender is addressed in these frameworks primarily as a downstream harm to be mitigated, not as a structural condition of the knowledge and power producing the frameworks themselves. Intersectionality is acknowledged tokenistically, if at all. Enforcement mechanisms on gender and inclusion are conspicuously absent.

This paper argues that the Brat Pack Effect is analytically distinct from its three predecessors in three ways. First, it lacks any regional legitimation: unlike Brussels or Beijing, it cannot claim to represent a population or a place. Second,

its exclusions are constitutive rather than incidental: the Global South, women, and non-OECD voices are not merely underrepresented but structurally kept outside the norm-setting process. Third, its informality is by design: the voluntary, non-binding character of G7/OECD AI governance insulates it from democratic accountability while preserving its normative influence.
In naming this dynamic after the original Brat Pack—a wealthy, male, self-regarding Hollywood clique whose influence vastly exceeded their formal mandate—the paper invites critical scrutiny of who governs AI globally, under what authority, and in whose interest. Contested futures require contested genealogies.

Keywords: AI governance, Brussels Effect, norm diffusion, gender, Global South

Meem Arafat Manab (they/them) is a Marie Skłodowska-Curie doctoral fellow at the Ontology Engineering Group, Universidad Politécnica de Madrid, where their research examines how legal obligations can be made machine-executable, going beyond machine-readability. They are a member of the HARNESS Doctoral Network (Horizon Europe) and hold a European Master in Law, Data and AI from Dublin City University and the Universidad de León, where they graduated on the Dean's Honours List. Manab's work bridges artificial intelligence, data law, geopolitics, and moral philosophy. They are also a research associate at the School of Religion, Theology and Peace Studies, Trinity College Dublin, where they co-developed Watchtower, an AI platform for monitoring extremism in gaming ecosystems. They have contributed to AI governance processes at the ITU AI for Good Global Summit, and serve on the Support Pool of Experts of the European Data Protection Board and the AI Subcommittee of the Copyright Society. Their publications appear in venues including ISWC, ICCV, and JURIX workshops, Scientific Reports, Springer LNCS, and The Conversation. As of 2026, they hold an Emerging Leaders in Ethical AI Governance Fellow at Globethics, Geneva, and a Public Policy New Voices Europe Fellow at the Salzburg Global Seminar.

From Globalization Losers to Automation Losers: Reconfiguring Political Cleavages in the Age of AI

Nikola Ilić
Junior Researcher
Faculty of Political Science, University of Belgrade
Email: nikola.ilic@fpn.bg.ac.rs

Vlada Milenković
MA Student
Faculty of Organizational Sciences, University of Belgrade
Email: vladamil99@gmail.com

At the end of the twentieth century, globalization brought changes in almost all aspects of life, drastically reshaping modern societies. Today we are witnessing a similar trend, namely the emergence and development of advanced technology—*AI* and *robotics.* As a consequence, big-tech companies are gaining unprecedented power, and automation is projected to replace many current jobs. All of this opens important questions about the way society will function in the future.

In the 1990s and early 2000s, the increased flow of capital and international trade affirmed the market and global capitalism as the dominant force, with its narrative and ethos structuring the decision-making and pushing the reforms of different governments. From that point, in the analysis of political competition in the West (North America, Western Europe) we can observe the formation of a cleavage in the form of so-called "globalization losers" vs. "globalization winners." The latter is composed of skilled, specialized professionals, with better education, who ultimately gained from this process, while the former entails the blue-collar workforce. After the "opening of borders," many industries left the center for semi-peripheral countries, and globalization losers, being less able to adapt to competition and more dependent on protection and aid of the state, grew in dissatisfaction and became susceptible to anti-system and anti-*status-quo* discourses proposed by right-wing populists. At the moment, right-wing populism (RWP) is gaining significant power by being able to capitalize on this resentment felt by the working, lower-middle, and middle classes.

The crucial difference is that contemporary and future changes in technology are not only going to target blue-collar workers (semi- or low-skilled), but will also increasingly affect the better educated, more specialized white-collar workers in areas such as the IT sector, and knowledge-based workers. In our analysis, we want to raise the following questions. What impact will technological changes have on the workforce? Additionally, how will the future political competition be structured along the economic-cultural divide (GAL-TAN scale)? The past restructuring of political values related to globalization saw an increase in the salience of cultural topics, due to the role of political entrepreneurs (first with the greens, then with right-wing populism), but what can we expect from the future: is a reproduction of what happened to blue-collar workers to be expected with white-collar workers?

In doing so, we also want to examine the role of big tech. More precisely, we examine

the fusion of technology and finance in the capital market, which represents an important aspect of the growing influence of these companies. Another aspect is the connection established between states and corporations through public-private partnerships, which in practice enable big tech to accumulate even more power thanks to a growing list of services that states delegate. Our contribution combines political theory and emerging trends that can be seen in the newest organizational structures in the “AI-powered” private sector, and it aims to open a debate on these topics.

Keywords: political cleavages, automation, AI, big tech, right-wing populism

Nikola Ilić is a junior researcher at the Faculty of Political Science, University of Belgrade, and a PhD candidate in Politics. His research focuses on contemporary political theory, populism, and democratic theory. He holds a BA in Political Science and an MA in Political Theory from the University of Belgrade, where he completed a thesis on the populist challenge to democratic theory. His recent work examines the conceptualization of populism, the role of socio-economic grievances as causes of populist mobilization, and the intersections of democracy, recognition, and political contestation. He has presented his research at several international conferences and has published in both academic journals and edited volumes. Alongside his research, he works as a teaching fellow in political theory and has participated in research projects and initiatives with a focus on democracy, civic engagement, and politics.

Vlada Milenković holds a Bachelor's degree in Political Science from the University of Belgrade, where he developed a strong foundation in political theory, comparative politics, and the study of emerging technological challenges to democratic governance. During his undergraduate studies, he was an intern at the Belgrade Centre for Security Policy, one of the region's leading think tanks, where his research focused on the intersection of technology and security. He is currently completing a Master's degree at the Faculty of Organizational Sciences, University of Belgrade, where his thesis examines and compares startup organizations and established corporations in the context of B2B sales processes and organizational adaptation. This work has deepened his interest in how organizational structures evolve in response to technological and market pressures. His research interests sit at the intersection of political theory, technology governance, and organizational studies, with a particular focus on how AI-driven transformations reshape political competition, labor markets, and the concentration of corporate power in contemporary societies.

Beyond Copyright: Artistic Labor, AI Infrastructures, and the Question of the Commons

Milica Pekić
Art Historian and Curator
Kiosk Platform for Contemporary Art
Email: pekic.milica@gmail.com

When artificial intelligence enters debates on art and culture, the default response is almost invariably to turn to copyright. Whose work was used to train the model? Should artists be compensated for the use of their work in training datasets? These are urgent and legitimate questions, particularly as artists confront AI systems built on vast accumulations of cultural production. Yet this paper asks whether copyright is capable of addressing the central issues at stake. While it remains an important legal instrument, it does not adequately account for the nature of artistic labor, the conditions under which it is produced, or the ways in which its value is recognized and distributed. Addressing these questions requires foregrounding labor and shifting the debate from intellectual property to the infrastructures through which cultural value is produced, circulated, and governed.

The paper argues that contemporary debates on AI and authorship unfold within a long-standing paradigm structuring the capitalist art system: the glorification of individual authorship, the mystification of creative labor, and the fetishization of the artwork as a unique object whose value can be owned, exchanged, and accumulated. Drawing on Marx's critique of alienated labor, the paper suggests that the AI crisis of authorship does not introduce a new problem but intensifies an existing structural contradiction between artistic labor as a socially productive activity and its organization through regimes of ownership and commodification. Building on critical work on reproductive, artistic, and digital labor, platform capitalism, and the commons, including that of Silvia Federici, Tiziana Terranova, Trebor Scholz, Nick Srnicek, and Katja Praznik, the paper asks what kinds of infrastructures might support artistic labor beyond market-dominated forms of control that are increasingly reproduced in contemporary AI systems.

This argument is grounded in three decades of practice in Serbia and the post-Yugoslav region, where independent collectives, networks, and self-organized spaces operated at the intersection of post-socialist institutional breakdown and neoliberal cultural restructuring, developing models based on collaboration and collective governance. Rather than treating these practices as ready-made solutions, the paper approaches them as repositories of practical knowledge about how cultural labor, resources, and infrastructures can be organized otherwise, and as a situated perspective from which to revisit contemporary debates on AI.

The paper concludes by opening questions rather than offering definitive answers. If AI systems are built upon the accumulated cultural labor of millions, how should this labor be recognized and valued? What forms of governance and public accountability, including instruments such as universal basic income or collectively governed digital infrastructures, might ensure that AI serves collective rather than exclusively private interests? Can the experiences of alternative

cultural infrastructures contribute to imagining technological futures in which artistic labor, shared knowledge, and digital infrastructures are understood as part of the commons rather than as objects of extraction and enclosure? These questions invite us to think beyond existing regimes of ownership toward more socially grounded infrastructures of value.

Keywords: artistic labor, digital commons, alternative art infrastructures, post-Yugoslav independent art scene, copyright

Milica Pekić holds a PhD and is an art historian and curator based in Belgrade. Her research focuses on avant-garde artistic practices, alternative forms of collective authorship, collaborative artistic practices, institutional transformation, and interdisciplinary approaches to contemporary art production and art history research. Actively engaged in the cultural civil sector for two decades, she has participated in numerous national and regional initiatives aimed at shaping sustainable frameworks for the development of the independent cultural scene and fostering regional collaboration. She is the co-founder of Kiosk Platform for Contemporary Art, an organization through which she collaborates with artists, researchers, and collectives to explore the relationship between art and society. Kiosk is also one of the founding members of the Association of the Independent Cultural Scene of Serbia and Kooperativa—Regional Platform for Culture.

Built for Humans Only: *Generative AI and the Redistribution of Epistemic Power*

Petr Špecián
Assistant Professor
Institute of Political Studies, Faculty of Social Sciences, Charles University
Email: petr.specian@fsv.cuni.cz
ORCID: https://orcid.org/0000-0003-2702-0354

Universities find themselves in a strange position, suspended between panic and complacency in the face of generative AI (GenAI). Even as the strain mounts, decisive adaptation efforts remain rare. This matters because universities serve as prime sites of knowledge creation, dissemination, and stewardship in democratic societies: they help determine who counts as a credible knower, and under what procedures. If they lose their footing, the consequences extend well beyond academia. And their problems reach much further than restoring the verification of provenance of academic texts. GenAI is mismatched with the university's core organizational logic, and its accommodation requires nothing less than rethinking the distribution of epistemic power.

So far, the responses have typically included AI-use guidelines, stricter proctoring, disciplinary action, and updated syllabi. What these otherwise respectable efforts share is an intent to tweak what works and update at the margin, leaving the core functions and procedures intact. In short, they are patches. Their fault is not that they are unreasonable or lazy—it is mainly that they are defensive efforts intended to prop up legacy power structures now facing a situation that forces substantial power redistribution.

Three aspects compound into a uniquely toxic mix—a *toxic trinity* of sorts—that upsets higher education's institutional equilibrium. First, there is the peculiar nature of GenAI itself. It possesses a mixture of traits transgressing the categorical boundaries between tools and workers. It combines the tools' lack of responsibility and intentionality with the workers' unpredictability and, increasingly, agency. Second, there is the pronounced provider-centrism of higher education, that is, a structural bias favoring insiders (faculty, administrators) over the ostensible beneficiaries of the institution's mission. Finally, there is the process centrality, meaning that the value of core educational activities such as writing an essay or thesis is not in the resulting artifact but in the formative process it is meant to evidence. That evidentiary role is also the source of the university's certifying authority, so when automation compromises the artifact's proof value, it compromises a key source of universities' epistemic power within the democratic system.

The "toxic trinity" presents universities with a hard challenge that pushes straight against the vulnerabilities of their power-preservation-oriented institutional setup. An answer more radical than patching is needed. Specifically, I explore divorcing authentication of student portfolios from their evaluation, pushing the latter down the line to external gatekeepers. This promises to enhance universities' robustness vis-à-vis GenAI while not giving up the commitment to transformative learning. At the same time, it circumscribes previous certification powers,

passing evaluative judgment to downstream actors: employers, specialized assessors, and students themselves.
I thus pursue two questions: through which structural mechanisms does GenAI threaten the university, and what reconfiguration could restore its robustness? I answer by (1) offering a novel conceptualization of the threat model of GenAI to higher education via the toxic trinity of categorical anomaly, provider-centrism, and process centrality; and (2) sketching a “New Deal” redistributing epistemic power between universities, their students, and the broader society, which revolves around the split between authentication and evaluation.

Keywords: generative AI, higher education, epistemic authority, university governance, assessment and credentialing

Petr Špecián is an assistant professor at the Institute of Political Studies, Faculty of Social Sciences, Charles University, where he coordinates the international Politics, Philosophy and Economics (PPE) program and leads the AI Institutional Transformation Research Group. His research focuses on social epistemology, institutional theory, democratic governance, and the institutional implications of generative AI. He is the author of Behavioral Political Economy and Democratic Theory: Fortifying Democracy for the Digital Age (Routledge, 2022) and has published in Social Epistemology and AI & Society, including recent work on integrating large language models into democratic assemblies and on the comparison of human and machine expertise.

Algorithms Between Us: Contested Futures, Mediation, and the Politics of Judgment

Phillip M. Rowell II
Chief Analytics Officer
Health Catalyst
Email: phillip.m.rowell@gmail.com

This paper examines algorithmic mediation as a site of political power rather than technical neutrality. Technological futures are not given; they are made, governed, and narrated—and they can be contested. Lévi-Strauss distinguished the engineer, who builds toward a purpose, from the bricoleur, who thinks with whatever is at hand. Today's algorithms behave like the latter, recombining human fragments into heterogeneous assemblages that sit between human parties, transforming each one's inputs while advancing aims of their own. The study's purpose is to theorize this mediation and trace its consequences for agency, epistemic authority, and democratic accountability.

The paper asks: How do AI-driven systems redistribute human judgment when they sit between people rather than simply serving them? Where do agency and authority actually reside once mediation becomes pervasive? And what instruments of accountability and contestation must democratic institutions build to keep pace?

To pursue these questions, the paper considers adaptive learning systems, widely framed as neutral instruments of instructional optimization. They increasingly mediate between learners, educators, and administrators, transforming what each contributes while pursuing objectives of their own. Curricula now train people to read what an algorithm displays rather than to coordinate with one another through it—a quiet redistribution of judgment that doubles as social sorting.

Two cultural flashpoints sharpen the stakes. During the ivermectin craze, search engines and partisan feeds formed a recursive loop that amplified a debate rather than reporting it, turning a livestock dewormer into a marker of identity. The "Google doctor" tells a parallel story: patients arrive certain of a diagnosis the algorithm already handed them, reshaping epistemic authority before any conversation begins.

Drawing on the author's twenty years in healthcare analytics and AI governance, the paper proposes a postanthropological account asking not whether machines think, but how judgment gets redistributed across the assemblage—and why democratic institutions, struggling to keep pace, must build new instruments of accountability and contestation.

Keywords: algorithmic mediation, politics of judgment, AI governance, contested futures, social sorting

Phillip M. Rowell II is a practitioner-scholar working at the intersection of healthcare analytics, AI governance, and socio-technical systems. As Chief Analytics Officer at Health Catalyst, he leads enterprise analytics and AI strategy. He holds a Master of Jurisprudence in Health Law, which shapes his focus on judgment, accountability, and contestation when algorithms sit between people. His work argues that no system is the neutral instrument it pretends to be—each one quietly reroutes who decides, who gets sorted, and who stays accountable when the future arrives already narrated.

The New “Critical” Infrastructure: Data Centers and the Evolving Policy Landscape in India

Aakansha Natani
Assistant Professor
International Institute of Information Technology Hyderabad
Email: aakansha.natani@gmail.com

The rapid digitization of economic and social life has fundamentally transformed the strategic importance of data centers, elevating them from technical back-end facilities to indispensable pillars of national infrastructure. In India, this transformation has been particularly pronounced. As the world’s most populous nation and one of its fastest-growing digital economies, India now stands at a critical juncture: the decisions it makes about data center governance, investment, and regulation will shape not only its digital sovereignty but also its broader developmental trajectory for decades to come.

This paper examines the emergence of data centers as a new category of “critical infrastructure” in India, analyzing the policy, regulatory, and geopolitical forces driving this reclassification. Drawing on official policy documents, legislative frameworks, and comparative international experience, it traces recent developments in India. Central to this analysis are the passage and ongoing implementation of the Digital Personal Data Protection Act (2023) and the evolving contours of national and state-level data center policies, which together signal a new phase of infrastructural statecraft.

The paper argues that India’s approach reflects a distinctive hybrid logic, simultaneously seeking to attract large-scale foreign and domestic investment in data center capacity while asserting regulatory control over data localization, cross-border data flows, and critical sector designation. This tension between openness and sovereignty, the paper contends, is not incidental but constitutive of India’s broader digital governance model.

The paper also situates India’s data center policy within global debates on technological sovereignty, supply chain security, and the geopolitics of cloud infrastructure. As hyperscalers such as Amazon Web Services, Microsoft Azure, and Google Cloud deepen their footprint in the country, questions of dependency, resilience, and regulatory leverage become increasingly salient. India’s experience offers a valuable case study in how emerging economies navigate structural power asymmetries in the global digital economy while pursuing ambitious industrial and governance objectives.

The paper concludes by identifying key policy gaps and proposing a set of recommendations for a coherent, future-oriented data center policy regime. In doing so, it contributes to broader scholarly conversations on digital industrialization, infrastructure governance, and the political economy of the internet in the Global South.

Keywords: data centers, data infrastructure, data governance, India

Aakansha Natani is an assistant professor of Political Science in the Human Sciences Research Centre, IIIT Hyderabad, India. She works broadly on data diplomacy, data governance and regulations, AI and data infrastructures in the EU and India. She was awarded a PhD from the Centre for European Studies, Jawaharlal Nehru University, New Delhi. She worked as a visiting fellow in the Department of Politics and International Studies, University of Cambridge, UK, in 2023. She was awarded the Jean Monnet Module Grant on "Digital Democracy and Data Governance in the European Union" by the European Commission in 2024. She was recently a visiting fellow at the Department of Informatics, University of Hamburg (May–July 2026).

EDUCATION, AI-ASSISTED LEARNING, AND DIGITAL LITERACY

From Author to Editor: How Generative Artificial Intelligence Is Transforming Academic Writing and Learning

Nikola Perišić
Research Associate, Institute for Political Studies
PhD candidate, Faculty of Political Science, University of Belgrade
Email: nikola.perisic.fpn@gmail.com

The emergence of generative artificial intelligence represents one of the most significant transformations of educational practices since the beginning of the digital era. Tools such as ChatGPT, Claude, and Gemini enable students and educators to generate essays, summaries, research outlines, translations, and analyses within seconds, fundamentally reshaping established practices of learning, writing, and knowledge production. However, public and academic debates surrounding the use of generative AI in education have largely focused on issues of academic integrity, plagiarism, and student "cheating." Such discussions often overlook the deeper transformations these technologies introduce to the very nature of academic work and literacy. Rather than asking whether students use artificial intelligence, a more pressing question concerns how its use is transforming the processes of thinking, researching, and writing.

This paper argues that generative artificial intelligence should not be understood merely as a new educational tool but as a technology that reconfigures the relationship between authorship, text, and knowledge. Traditionally, academic writing has been grounded in processes of information gathering, critical analysis, argumentation, and the production of original texts. With the widespread availability of generative AI systems, there is an observable shift from the role of author to that of editor, where users increasingly guide, evaluate, revise, and refine content generated by algorithms rather than producing texts entirely on their own. This transformation raises important questions concerning the changing meanings of authorship, expertise, and academic competence.

The paper examines three interconnected dimensions of this transformation. First, it explores changes in academic literacy resulting from the growing integration of AI tools into learning and writing practices. While traditional models of literacy emphasize independent research, argumentation, and writing, the emerging educational landscape requires new competencies related to prompt design, evaluation of AI-generated outputs, fact-checking, and the revision of generated content. This development raises the question of whether new forms of digital and algorithmic literacy are becoming as important as conventional academic skills.

Second, the paper investigates the implications of these changes for critical thinking and learning processes. Although AI systems may facilitate access to information and support educational activities, excessive reliance on generated content may reduce intellectual engagement, reflection, and independent problem-solving. Of particular importance is whether students develop the capacity to critically assess AI-generated information or become increasingly dependent on algorithmically produced interpretations and conclusions.

Third, the paper analyzes the changing role of educators and educational

institutions in an environment shaped by pervasive artificial intelligence. Rather than serving primarily as transmitters of knowledge, educators are increasingly becoming mentors, facilitators, and guides within complex information ecosystems. At the same time, universities face the challenge of redefining assessment practices, academic ethics, and educational objectives in a context where text generation has become instantaneous and widely accessible.
Methodologically, the paper is based on a critical review of literature in the fields of education, digital literacy, and artificial intelligence studies, alongside an examination of ongoing institutional debates regarding the integration of AI tools into higher education. The objective is neither to portray generative AI as a threat nor to celebrate it as a solution, but rather to critically examine how it transforms the ways in which knowledge is produced, communicated, and evaluated.
The paper concludes that the central challenge facing contemporary education is not the prohibition of artificial intelligence but the development of new forms of critical, reflective, and responsible engagement with AI systems. Understanding these transformations is essential for reimagining the future of education in an era increasingly defined by collaboration between humans and intelligent technologies.

Keywords: generative artificial intelligence, academic writing, digital literacy, higher education, critical thinking

Nikola Perišić is a research associate at the Institute for Political Studies and a PhD candidate at the Faculty of Political Science, University of Belgrade. His academic and research interests focus on contemporary political processes, media discourse analysis, and the study of the relationship between public opinion and political actors in modern societies. He is the author and co-author of more than 25 scientific papers in the field of political science, with a particular emphasis on topics related to political communication, media, and public discourse. In his previous work, he has participated in numerous research projects dealing with Serbia's European integration process, the development of media literacy, and the analysis of citizens' social and political attitudes. A special focus of his research is on young people and their social and political position, as well as the ways in which they obtain information and form attitudes within the contemporary media environment. Through his work, he aims to contribute to a better understanding of the factors influencing the political socialization of youth and their role in contemporary democratic processes.

The Future of Education in the Age of Generative AI: Governing Intellectual Formation in Higher Education

Andreja Klančar
Assistant Professor
Faculty of Education, University of Primorska
Email: andreja.klancar@pef.upr.si

Ernest Ženko
Professor
University of Primorska
Email: ernest.zenko@upr.si

Generative artificial intelligence (AI) is rapidly transforming higher education, not merely by introducing new technological tools but by reshaping the conditions under which learning, knowledge production, and intellectual development take place. As AI systems become increasingly integrated into educational environments, they influence academic writing, assessment practices, information retrieval, and students' engagement with disciplinary knowledge. These developments have generated considerable debate about the future of education and the role of universities in preserving meaningful forms of human intellectual participation within increasingly automated learning environments.

This paper examines these challenges through the development of FRAME (Framework for Responsible AI Management in Education) at the University of Primorska (Slovenia). FRAME was developed as an institutional response to the rapid expansion of generative AI and as an attempt to shape educational futures that preserve human intellectual agency while enabling responsible engagement with emerging technologies. Rather than approaching AI primarily as a problem of detection, prohibition, or technological control, the framework seeks to establish conditions under which AI can support learning while preserving students' intellectual autonomy, critical thinking, and responsibility for their work. As such, it represents a practical example of how higher education institutions can actively participate in shaping technological futures rather than merely adapting to them.

While generative AI offers significant opportunities for supporting teaching and learning, it simultaneously raises important ethical, pedagogical, and institutional concerns. These include risks associated with cognitive offloading, reduced critical engagement, academic integrity violations, uncertain authorship, privacy and data protection issues, and the growing difficulty of designing assessment practices capable of reliably evaluating student learning. More fundamentally, generative AI challenges traditional assumptions about the relationship between intellectual effort, knowledge acquisition, and competence development. If students increasingly rely on AI systems to perform cognitive tasks that higher education has historically sought to cultivate, universities must reconsider how learning outcomes, assessment, and educational responsibility are conceptualized and governed.

The paper analyzes the normative foundations and governance principles underlying FRAME and discusses their broader relevance for higher education. The

framework is grounded in several principles: human oversight and authorship, alignment between AI use and intended learning outcomes, transparency and traceability of AI-assisted work, protection of personal and institutional data, and consideration of the wider social and environmental consequences of digital technologies. Particular emphasis is placed on the distinction between AI use that supports learning and AI use that substitutes for cognitive processes central to educational development. The framework therefore differentiates among various forms of AI-supported activity according to their impact on student learning, cognitive engagement, and academic integrity.

The paper argues that the central educational challenge posed by generative AI is not whether students use AI, but how such use affects the formation of knowledge, judgment, and intellectual autonomy. Responsible governance requires institutions to move beyond simplistic narratives of technological inevitability and to develop educational practices that make students' thinking visible, preserve opportunities for reflection and independent reasoning, and foster critical digital literacy. This includes rethinking assessment practices, strengthening assessment literacy among educators, and developing transparent institutional policies capable of adapting to rapidly evolving technological conditions.

By examining the development of FRAME as a case of institutional AI governance, the paper contributes to broader discussions about contested technological futures in education. It argues that universities retain significant agency in determining how AI becomes embedded in educational practice and that the future of education depends not only on technological innovation but also on collective decisions regarding what forms of knowledge, intellectual effort, and human development higher education should continue to value, cultivate, and protect.

Keywords: generative artificial intelligence, higher education, intellectual formation, academic integrity, ethics

Andreja Klančar, PhD, is an assistant professor of Computer Science Education and Educational Technology at the Faculty of Education, University of Primorska, Slovenia, and an affiliate of the Centre for Responsible Artificial Intelligence (CeResAI UP). Her academic and professional work focuses on digital educational technologies, computer science education, and the pedagogically meaningful integration of artificial intelligence into teaching and learning. She leads and contributes to national and EU-funded projects in the fields of digital competencies, generative AI in education, pedagogical innovation, and technology-enhanced learning. Her work also includes the development of digital competencies and the design and implementation of professional training for higher education teachers on the effective use of contemporary digital technologies in teaching. Through her research, teaching, and project work, she explores how digital technologies and AI can support high-quality, reflective, and equitable educational practice while preserving the central role of pedagogy, critical thinking, and human agency in education.

Ernest Ženko, PhD, is a full professor of Philosophy of Culture at the University of Primorska, Slovenia, and director of the Centre for Responsible Artificial Intelligence (CeResAI). His academic work spans philosophy, psychology, media studies, and technology, with a particular focus on the ethical, epistemological, and societal implications of artificial intelligence. In recent years, his research has concentrated on questions of AI governance, human-AI interaction, digital literacy, and the impact of generative AI on education, knowledge production, and critical thinking. He is a member of the International Association for Safe and Ethical AI (IASEAI) and the International AI in Education Society (IAIED), and actively participates in national and international research and development projects related to artificial intelligence, digital transformation, and responsible innovation. His work frequently bridges theoretical inquiry and practical implementation, exploring how philosophical perspectives can inform contemporary debates on technology, ethics, education, and public policy. Through his academic, institutional, and public-facing work, he advocates for a critical and reflective approach to AI that recognizes both its transformative potential and its ethical challenges, emphasizing the importance of human agency, intellectual autonomy, and sustainability in an increasingly AI-mediated world.

Algorithmic Futures and Global Mobility: Inequality, Extraction, and the Contestation of AI in Education Assessment

Jelena Popov
Assistant Professor
IOE, UCL's Faculty of Education and Society, University College London
Email: popovnjelena@gmail.com

This paper examines how artificial intelligence reshapes power, governance, and learning through its embedding in and reconfiguration of global ecosystems of education and mobility. Focusing on AI-enabled assessment systems for English language proficiency used for university entry and migration, the paper argues that AI does not simply *optimize* institutional decision-making but, rather, reconfigures the conditions under which access, participation, and legitimacy are established. Drawing on a socio-cultural-material framework, AI is conceptualized as an element of a distributed ecosystem in which human and machine agents are co-constituted through practices, infrastructures, and discourses.

Empirically, the paper analyzes AI-mediated English language testing and students engagement with high-stakes machine learning AI systems to show how algorithmic infrastructures reorganize governance. The data corpus consists of a survey (n = 406), interviews (n = 20), and secondary data. It shows how machine learning systems play a central role in high-stakes decision-making through autonomous classifying and sorting, mediating access to education and work and redistributing authority across institutions, platforms, and technical systems. This shift produces new forms of (a) algorithmic governance in which key decisions are increasingly delegated to opaque computational processes that operate beyond direct scrutiny, and (b) new learning practices that test-takers develop for the test.

The paper demonstrates how inequality, exclusion, and extraction are simultaneously reproduced and intensified within these global systems. Access to opportunities becomes contingent not only on formal competence but on the ability to mobilize socio-material resources, including digital infrastructures, intermediary networks, and knowledge of algorithmic logics. These dynamics generate new stratifications within global education and labor markets, privileging those able to strategically navigate AI-mediated systems. At the same time, users contribute to processes of data extraction and system optimization, as their interactions feed back into the refinement of algorithmic models.

Crucially, the paper analyzes how these technological futures are narrated, legitimized, and contested. Institutional actors (e.g., universities and governments) frame AI as an objective, efficient, and unbiased improvement, thereby legitimizing its role in governance and decision-making. In contrast, users develop situated understandings and recruit novel practices to address opacity, uncertainty, and the need for critical engagement.

These competing narratives reveal AI futures as fundamentally contested. Technological systems are stabilized through institutional discourse and infrastructural embedding, but they are simultaneously destabilized through everyday practices that expose their limits and reconfigure their meaning. By foreground-

ing the interplay between governance, inequality, and technological imaginaries, the paper contributes to debates on contested futures by showing how AI both enables and constrains participation in education and work, reproducing structural inequalities while opening spaces for negotiation and resistance.

Keywords: AI governance, inequality, global mobility, technological imaginaries, sociomateriality

Dr Jelena Popov is an assistant professor in Professional and Vocational Learning at UCL. She is the programme director of the MSc Engineering Education. Her work contributes to the Centre for Youth, Employment, and Skills, the Centre for Engineering Education, and the AI Observatory at UCL.

Educational AI and Development of Critical Thinking and Cognition

Vladimir Pavlović
Professor
Faculty of Agriculture, University of Belgrade
Email: vlaver@agrif.bg.ac.rs

Olivera Bogunović
Assistant Professor
Harvard Medical School, Harvard University
Email: bogunovic@mclean.harvard.edu

Bogdan Novaković
PhD Candidate
Department of Human Sciences for Education "Riccardo Massa," University of Milano-Bicocca
Email: b.novakovic@campus.unimib.it

This work examines how digital technologies and educational AI influence critical thinking and cognition, proposing a framework for integrating generative AI into pedagogical practice. Within educational contexts, AI is restructuring pedagogy and administration, yet resource abundance often induces cognitive overload. Contemporary research reveals the reverse Flynn effect—declining IQ and cognitive abilities, particularly among Generation Z, observable since approximately 2010. Yale University reports a 90% increase in cognitive difficulties among adults aged 18–34 between 2013 and 2023, with declines in verbal, numerical, and fluid reasoning. Excessive screen exposure is a primary factor, with students using computers five hours daily showing lower performance in attention, memory, literacy, and executive function. Cognitive offloading—relying on digital tools to reduce mental effort—impairs long-term memory formation, with a strong negative correlation between offloading frequency and recall accuracy. The "Google effect" exemplifies this, as individuals remember information location rather than content, reducing hippocampal activity. While offloading can be beneficial when mental energy is redirected to deeper understanding, passive use creates cumulative cognitive deficits. Print reading generates stronger neural connectivity in language and emotion-processing regions than screen reading, which reduces hippocampal activation. Screen readers multitask and scan, while print readers sustain attention; a 2024 meta-analysis confirms print superiority for comprehension across 49 studies. Chronic digital access triggers neuroplastic restructuring that favors rapid task-switching over deep processing.
These challenges demand solutions beyond content organization. The OECD's Digital Education Outlook 2026 reports that students using general-purpose AI improved short-term performance by 48% but performed 17% worse without access, confirming that learning gains are unsustainable when technology replaces cognitive effort. Generative AI offers opportunities for personalized scaffolding but risks cognitive offloading and skill atrophy. While AI expedites early learning, it may bypass the cognitive struggle essential for internalizing core procedures,

compromising long-term retention and transfer. AI excels at replication but lacks human capacity for ethical judgment and meaningful integration. Effective pedagogy requires metacognitive instruction, process documentation, and assessments rewarding deep engagement over superficial output.
Adaptive platforms enforcing active engagement and desirable difficulties demonstrate superior retention compared to static study aids. The future depends on designing AI as a tutor promoting reasoning rather than a solver providing answers, ensuring that the human remains the cognitive driver and AI serves as a cognitive gym that strengthens mental capacity. This alignment can transform AI from a source of cognitive decay into an engine for cognitive growth, provided pedagogical innovation prioritizes deep engagement, metacognitive awareness, and gradual autonomy development. The evidence presented demonstrates that AI has potential as a powerful engine for cognitive growth, but only when used strategically to support, rather than replace, the cognitive work of learning. The distinction between strategic augmentation and passive delegation, the difference between extraction and synthesis, and the impact of interface design on schema consolidation all point to a single conclusion: the quality of learning in the age of AI depends fundamentally upon the intentionality of instructional design and the preservation of human cognitive agency.

Keywords: educational AI, cognition, digital technologies

Vladimir Pavlović is a professor and the head of the Department of Mathematics and Physics at the Faculty of Agriculture, University of Belgrade, and program director of the Center for Biosystem AI and the Center for Data Mining and Bioinformatics. Recognized among Serbia's most cited scientists in advanced convergent technologies, he has collaborated with leading international institutions, including NASA's Research Center, the Computational Center of Research Excellence at NCCU, and Harvard University. He has coordinated 23 research projects, led the international initiative "Artificial Intelligence and Digital Technologies in Digitally Supported University Education," published over 400 scientific papers and conference proceedings, and held six solo exhibitions on AI applications in visual arts. He is also a founding member of the Tesla&I Initiative, which unites Serbian-American scholars in cognition and creative AI research. Within this initiative, he developed the concept of symbolic cognitive autopoiesis, which enables varied readings of narrative structures embedded in so-called cognitive compositions. Within this framework, each individual part of a cognitive composition can be observed separately, yet as part of a larger whole, which can be utilized in various educational contexts within the STE(A)M (Science, Technology, Engineering, Arts, and Mathematics) disciplines.

Olivera Bogunović is an assistant professor at Harvard Medical School, where she also serves as director of education for subspecialty trainees. She is chief of the Addiction Clinic at McLean Hospital in Boston. She has structured her professional career and research around developing programs to support professional staff and advance research in the field of addiction, including digital technology addiction. She has initiated research in the area of neuro-art and the integration of AI technologies for the treatment of addiction disorders. Her work is also focused on developing human-centric AI solutions that prioritize support for individuals with special needs and promote inclusive educational practices. She has published numerous papers and delivered many invited lectures at various universities and professional conferences. Bogunović leads the team for Harvard's global collaboration with the Middle East and is one of the founders of the Tesla&I initiative, which brings together Serbian-American scientists working in the fields of cognition and creative uses of artificial intelligence.

Bogdan Novaković is a doctoral candidate in the Education in Contemporary Society program at the University of Milano-Bicocca, affiliated with the Department of Human Sciences for Education "Riccardo Massa." His research examines the integration of artificial intelligence, augmented reality, and virtual reality into educational transformation within contemporary society. He holds a Master's degree in Visual Communications from the Faculty of Applied Arts in Belgrade and a second Master's in Communication Design from the Politecnico di Milano, where his thesis explored AI's role in designing sustainability-oriented applications. He has worked on virtual and mixed reality applications, including designing a language-learning support tool for children with speech impairments. His teaching experience includes serving as an assistant at the Faculty of Computer Science (RAF) in Belgrade, within the Department of Multimedia Design, where he instructed students in branding and application design principles. He has also been a visiting lecturer at the Politecnico di Milano, delivering lectures on AI applications in design as part of the Media Sociology and Digital Culture course, bridging theoretical frameworks with practical design interventions.

Inferential Deskilling in the Classroom: When AI Outputs Become "Premises"

Ivan Gospodinov
Independent Researcher
Email: iwan.gospodinow@gmail.com

This paper proposes an inferentialist re-description of what changes when generative AI becomes incorporated into educational practice. Its central claim is that the main pedagogical problem is not that students may use AI to produce answers, but that AI may contribute to inferential deskilling: the weakening of the learner's ability to take responsibility for the reason-relations through which an answer becomes intelligible as knowledge. AI-generated outputs can easily function as classroom premises that no one practically owns. They enter essays, discussions, and assignments as if they were already available reasons, while the learner has not yet undertaken the commitments that would make those claims genuinely theirs.

From an inferentialist perspective, discursive competence is not exhausted by the capacity to produce or select well-formulated sentences. To understand a concept is to know how it can be used in reasoning: what follows from it, what counts against it, what it rules out, what would support it, and what responsibilities one undertakes in asserting it. Students are increasingly invited to work with claims that are already phrased, organized, and rhetorically formulated, but whose justificatory structure remains *discursively, personally, and cognitively distant*. What is weakened is not discourse, personality, or cognition as such, but the learner's authority over the commitments carried by AI-assisted claims, because AI may intensify the classroom tendency to treat learning as the reproduction of useful and correct representations ("the Given"), rather than as progressively expressive participation in reason-giving ("Making the implicit explicit").

For this reason, the pedagogical problem is not adequately captured by generic appeals to "critical thinking." What is required is a more precise inferentialist account of AI literacy as the cultivation of commitment-tracking. Students must learn what entitles them to treat AI claims as reasons worthy of being entertained in a game of giving and asking for reasons. AI literacy should therefore be conceptualized as a normative discipline of distinguishing representability from reasonability. Commitment-tracking means learning to identify the alethic (descriptive), deontic (normative), and modal (hypothetical) semantic consequences of a claim.

Methodologically, the paper treats classroom reason-giving as a practice of scorekeeping in which teachers and students track commitments, entitlements, defeaters, and incompatibilities. The assessment implication is that rubrics should reward inferential articulation: explaining why a claim is warranted, what follows from it, and what could undermine it. Inferential deskilling occurs when students lose interest in examining the relations of compatibility and consequence among concepts. The paper therefore argues that responsible AI use requires learning-by-teaching: students must take up teacher role-responsibilities by explaining, questioning, challenging, and repairing AI-assisted claims for others. In learning to teach AI-assisted content, students no longer relate to AI outputs as finished

representations to be accepted or rejected, but as materials whose inferential significance must be reconstructed before they can be responsibly used. AI literacy thus becomes a matter of pedagogical participation in the space of reasons, and it becomes educative only when it develops students' authority as honest, adept, and just scorekeepers of their own and others' claims.

Keywords: AI literacy, representationalism, inferentialism, responsibility, learning

Ivan Gospodinov holds a PhD in Philosophy of Education from the Catholic University of Eichstätt-Ingolstadt, Germany, with a dissertation titled "An Inferentialist Theory of Not-Learning" (2024). His research focuses on inferentialism, irresponsible learning, semantics, and educational holism. He has over ten years of experience in teaching, teacher training, educational content localization, and learning analysis. He has been involved with school digitalization at both the classroom and policy levels. He currently pursues his research independently, examining the relation between language, learning, and normative accountability through Robert Brandom's theory of discursive practice.

Tokens with Worldviews: Rethinking AI Literacy as Political Media Education

Kai Löser
Project Associate
University of Rostock
Email: loeser@uni-rostock.de

Large language models (LLMs) increasingly shape how learners encounter information, form judgments, and participate in public discourse. Yet educational debates often frame them as tools for automated text production, writing support, or efficient feedback. This output-oriented framing is too narrow for the contested technological futures currently taking shape through generative AI. LLMs are linguistic infrastructures: they organize plausible language, mediate interaction, and redistribute authority without appearing as political actors. The paper therefore transforms the Frankfurt Triangle (Brinda et al., 2019) into a framework of political media education for critical AI literacy. Rather than adding politics as an external ethical supplement, it treats the political as internal to how generative systems are built, governed, circulated, and used.

This transformation is organized by three guiding questions, each linked to a deliberately provocative metaphor. The first asks what is political in the technical structure of LLMs. “Machines of the Common Good” examines how data selection, probabilistic generation, and alignment can turn social patterns into apparently neutral answers (Bajohr, 2023; Bender et al., 2021). The second asks how LLMs are embedded in platform economies. “Products of Surveillance Capitalism” examines data extraction, monetized access, and infrastructural dependency without reducing all generative AI to a single economic diagnosis (Zuboff, 2015; Luitse & Denkena, 2021). The third asks how generative systems enter political communication and governance. “Servants of Right-Wing Populism” investigates their role in synthetic publics, fear narratives, platform spectacle, and democratic accountability (Mühlhoff, 2025). The metaphors are not final claims but heuristic probes: they sharpen controversy, make assumptions debatable, and keep critical AI literacy from becoming either tool optimism or technological determinism.

Methodologically, the contribution is a conceptual and didactic model-building paper, not an empirical workshop evaluation. It develops a task profile that treats LLM outputs as traces of their conditions of production rather than as self-contained answers. The proposed formats train forms of inquiry rather than better prompting: alignment audits compare model responses across controversial prompts; data biographies and business model mappings reconstruct the infrastructures behind seemingly simple interactions; openness matrices test claims of transparency; corpus contrasts situate generated language against traceable textual material; and discourse analyses examine the circulation, affective force, and accountability of AI-generated political content.

The central claim is not that LLMs possess worldviews, but that they can produce a “worldview without a subject”: patterns of plausibility that appear neutral while displacing the question of who speaks, under which conditions, and with what responsibility. By making this displacement analyzable, the paper shifts critical AI

literacy from competent tool use toward democratic judgment, public reasoning, and critical imagination in AI-mediated publics.

Keywords: critical AI literacy, political media education, Frankfurt Triangle, large language models, contested futures

Kai Löser works at the University of Rostock on a project on the implementation of Future Skills. He studied English, history, philosophy, sociology, and modern German literature at the University of Freiburg and subsequently worked as a research associate and lecturer at Bielefeld University. Research and study stays took him, among other places, to the University of Virginia in Charlottesville, Trinity College Dublin, and Yale University in New Haven. His research interests include critical media studies, AI literacy, cultural history, and political media education. He has recently published on the implementation of corpus and AI literacy in school and university contexts, as well as on political media education in the age of AI.

Prompting Without Understanding? Generative AI, Academic Writing, and Digital Literacy Among Students

Irina Filipović
PhD Student and Junior Researcher
Faculty of Philosophy, University of Niš
Email: irina.filipovic@filfak.ni.ac.rs

The rapid integration of generative artificial intelligence into academic practices has transformed the ways in which students approach learning, writing, research, and assessment. Tools such as ChatGPT are increasingly used not only for technical assistance, translation, or language correction, but also for idea generation, text structuring, paraphrasing, summarizing sources, and, in some cases, producing complete academic assignments. While public and institutional debates often frame this development through the binary opposition between innovation and cheating, such an approach overlooks a more complex issue: whether students possess the digital, informational, and academic literacy needed to use AI tools critically, responsibly, and meaningfully.

This paper examines the use of generative AI among university students, with particular attention to the distinction between assistance and substitution in academic work. Rather than asking only whether students use AI, the study focuses on how they understand different levels of AI involvement in the learning and writing process. The central argument is that the key challenge of AI-assisted learning is not the mere presence of AI in academic work, but students' ability to recognize when AI supports understanding and when it replaces the cognitive, interpretive, and authorial work expected in higher education.

The research is designed as a small-scale empirical study based on an anonymous online survey conducted among university students. The questionnaire explores several dimensions of AI use: frequency of use, purposes of use, students' perception of acceptable and unacceptable AI assistance, practices of verification, attitudes toward authorship and academic integrity, and perceived effects of AI on learning and understanding. Particular attention is paid to how students distinguish between using AI as a helper, editor, or substitute author. The study proposes a three-level analytical distinction: AI as an assistant, used for brainstorming, clarification, or organization of ideas; AI as an editor, used for stylistic improvement, paraphrasing, and language polishing; and AI as a substitute, used to generate complete arguments, analyses, or assignments with minimal student involvement.

By analyzing students' reported practices and attitudes, the paper aims to contribute to a more nuanced understanding of AI in education. It moves beyond the question of whether AI should be allowed or banned and instead asks how students can be prepared to use it without losing the central educational value of academic writing: the development of independent thinking, argumentation, interpretation, and critical judgment. In this way, the paper approaches generative AI as part of a contested educational future in which the boundaries between support, automation, and authorship are unstable.

Keywords: generative AI, digital literacy, academic writing, students, AI-assisted learning

Irina Filipović is a PhD student and junior researcher at the Department of Communication and Journalism, Faculty of Philosophy, University of Niš. Her research interests include digital communication, algorithmic recommendation systems, the functioning and regulation of digital platforms, media and copyright law, public opinion research, and comparative media systems. Her work is situated at the intersection of communication studies and legal analysis, with particular attention to how digital platforms, algorithmic infrastructures, and emerging technologies transform public communication, media accountability, and democratic debate. She has been involved in teaching and research activities in the fields of media law, public opinion, information society, and media systems.

ENVIRONMENT, EXTRACTION, SUSTAINABILITY, AND DIGITAL DEGROWTH

Environmental Subconsciousness of Generative AI: Scale, Sufficiency, and Planetary Boundaries

Ivana Stepanović
Senior Research Fellow
Institute of Advanced Studies Kőszeg
Email: ivana.stepanovic@gmail.com

David Daou
Associate Academic Officer
Institute for Environment and Human Security, United Nations University
Email: daou@ehs.unu.edu

The environmental costs of generative AI remain invisible to its users. Electricity drawn from carbon-intensive grids, water consumed in cooling data centers, and rare minerals extracted at significant ecological cost are structurally absent from the interface through which millions interact with large language models daily. This paper reads that absence as a form of digital subconsciousness: a suppressed truth of *conspicuous consumption* built into the architecture of commercial AI deployment. The question this paper pursues is not whether generative AI has environmental consequences, but why the growth paradigm governing its development forecloses the possibility of asking whether those consequences are necessary.

Drawing on conceptual analysis and empirical data on energy, water, and mineral consumption of large language models, the paper takes the perspective of digital degrowth. It argues that the principle of sufficiency must be embedded across all stages of AI development, deployment, and use. Sufficiency, understood as the normative commitment to human flourishing within planetary boundaries (Raworth, 2017), provides the paper's organizing concept and connects the critique of AI's environmental impact to the broader degrowth scholarship.

The paper's central claim is that scale is not a proxy for intelligence, and that the assumption otherwise underwrites both the environmental unsustainability of large language models and a philosophically problematic account of what those models actually do (Hao, 2025). The pressure to expand model size in terms of data, parameters, and computing power at accelerating ecological cost reflects the internal logic of the growth paradigm rather than any demonstrated equivalence between scale and cognitive capacity. Philosophy of mind offers a structurally unresolved challenge to claims of emergent understanding in systems that manipulate symbols without the knowledge of semantics (Searle, 1980), thereby simulating human intelligence; the constitutive relationship between meaning and context of its use raises questions about whether linguistic performance at scale amounts to anything like linguistic competence (Wittgenstein, 1953); and accounts of intelligence as environmentally embedded and relationally distributed suggest that resource-intensive architectures may be the wrong model of what cognition is (Clark & Chalmers, 1998). These are not objections to artificial intelligence as such, but objections to a particular developmental logic that conflates magnitude with sophistication without environmental embeddedness.

By situating these philosophical arguments within empirical research on genera-

tive AI's physical footprint, the paper contributes to the emerging framework of humanistic artificial intelligence with the aim of reorienting development toward ecological needs and human flourishing rather than simulational intelligence aimed purely at accelerating economic growth. The methodology combines quantitative research on environmental impacts with conceptual analysis encompassing degrowth economics, philosophy of mind, and environmental ethics. The environmental impact of LLM usage is estimated using a Life Cycle Assessment (LCA) framework, disaggregating the total cost of a single LLM request into operational impact (electricity consumption and water use during inference) and embodied impact from hardware manufacturing amortized over its lifetime. Electricity consumption is calculated bottom-up from GPU-level energy use, scaled to total data center consumption via the Power Usage Effectiveness (PUE) metric. Water consumption is measured across direct on-site cooling and indirect consumption at electricity-generating power plants. Carbon emissions are derived by multiplying total electricity consumption per request by the carbon intensity of the relevant grid, expressed in grams of CO2 equivalent per kilowatt-hour.
The paper's contribution is twofold. First, it offers a theoretically integrated critique connecting ontology, environmental ethics, and political economy into a coherent argument for sufficiency as a governing principle across the full lifecycle of AI development. Second, it insists that the question of what an LLM is for cannot be separated from the question of what it costs.

Keywords: digital degrowth, AI, large language models, sufficiency, human flourishing

Ivana Stepanović is a senior research fellow at the Institute of Advanced Studies Kőszeg and the Academic Coordinator of the UNESCO Chair on Cultural Heritage Management and Sustainability in Hungary. She holds a PhD in Anthropology from the Faculty of Philosophy, University of Belgrade, Serbia; an MA in Human Rights in South-East Europe from the University of Sarajevo, Bosnia and Herzegovina, and the University of Bologna, Italy; and a BA in Philosophy from the Faculty of Philosophy, University of Belgrade. Her research activities focus on digital ethnography of AI and social media, online activism, and the impact of artificial intelligence on societies.

David Daou has been at the forefront of applying advanced computational science to some of the world's most pressing environmental challenges for nearly two decades. As a senior scientist and leader of the AI and Remote Sensing Team at UNU-EHS, his research is dedicated to harnessing the power of artificial intelligence (AI) and remote sensing to model disaster risks and assess the impacts of climate change. Dr. Daou's expertise lies in creating powerful computational models that merge geospatial intelligence with deep learning to forecast floods, droughts, and wildfires with unprecedented accuracy.

Terms of Operationalized Extraction

Darija Medić
Assistant Professor
College of Arts and Media, University of Colorado Denver
Email: darija.medic@ucdenver.edu

Deceptive design (or dark) patterns are interaction design practices that exploit vulnerabilities of unconscious decision-making to manipulate users into making decisions that they might not otherwise choose. They are a pervasive issue in the digital world, undermining end users' capacity for conscious decision-making. In doing so, they contribute to what Cory Doctorow calls the enshittification of the digital sphere, a state in which the quality of the services people are experiencing is increasingly degrading by taking the services away from users toward shareholders, due to the predatory practices of big tech monopolies.
The declining quality of everyday services is linked to existing social and design norms embedded in technological systems shaped by extractivism.
This presentation proposes that deceptive design patterns are examples of extractive design situated within this ideology. Extractive design is a term that has been theorized in architecture to refer to harmful construction methods which rely on resource extraction practices and cause environmental damage (Space Caviar, 2021). More broadly, the term points to design practices that correspond to the neoliberal utilitarian goal of maximizing short-term profit through extracting resources by any means. In HCI and UX design scholarship, however, extractive design has not been extensively theorized as a term, even though there is a common understanding that attention, behavioral data, affect, labor, and money are all systematically extracted through everyday digital interfaces as a basis for how the attention economy of today functions.
Terms of Service agreements are legally binding documents which establish the relationship of users to the digital products they are using. Users consent to these agreements on a daily basis without reading the terms that govern what data the service will collect on the user and how, and what the user is and isn't allowed to do. This presentation proposes a hypothesis that Terms of Service agreements writ large epitomize the extractive design paradigm through (1) legitimizing the asymmetry of power relations between a platform and a user as they serve to in effect protect the platform from the user and (2) a plethora of deceptive design patterns throughout ToS consent flow implementation. It does so by analyzing the changes in the Terms of Service agreements of large tech platforms over time, as key enforcing mechanisms for implementing their extractive policies. The analysis focuses on the language of ToS agreements and content edits between different versions, as well as consent flows around ToS update implementation in order to uncover how a broader culture of toxic subordination these platforms habituate users into evolves to colonize further extraction territories over time.

Keywords: extractivism, deceptive design patterns, Terms of Service, attention economy, consent flow

Darija Medić is a tactical media artist, practice-based researcher, and media designer investigating contemporary agential labyrinths of cohabiting with technical infrastructures and their sociopolitical ecosystems. In particular, her work addresses the barriers to equitable user experience design from a critical materialist and accessibility standpoint. Combining a range of media into low-tech participatory installations, interventions, and performances, she creates collective processes for sensitizing, sometimes absurd, politically laden points behind the space where technical mediation and prosumer habituation meet. She is currently a tenure-track assistant professor of Digital Design in the Department of Visual Arts at the College of Arts and Media, University of Colorado Denver, and is completing her PhD at the Intermedia Art Writing and Performance Program at the University of Colorado Boulder. She is also a lecturer at the SHIFTA Creative Computing Master program for the module Internet Equalities.

HUMAN-MACHINE COMMUNICATION: POWER, DESIGN, AND HUMAN-AI RELATIONS

Fitting In: Human-Agent Interaction and Cultural Norms in a Simulated Online Community

Sara Major
Teaching Assistant and PhD Student
Faculty of Philosophy, University of Novi Sad
Email: sara.mejdzor@ff.uns.ac.rs

Aleksandar Tomašević
Associate Research Professor
Institute of Physics, University of Belgrade
Email: atomasevic@ipb.ac.rs

The growing integration of large language model (LLM) agents into online spaces raises questions about their capacity to meaningfully participate in digital environments shaped by specific norms and patterns of interaction. Online communities are shaped both by explicit rules and informal expectations about what counts as appropriate participation, including tone, style of interaction, acceptable forms of disagreement, humor, and reciprocity. For AI agents to be socially effective in these environments, they must be able to produce coherent and relevant responses, as well as behave in ways that fit the local context.

Our work examines human-machine communication through an experimental study of a mixed online community. The experiment is implemented on YSocial, an open-source, AI-driven social media simulation platform. We develop a Reddit-like discussion forum centered on popular culture, composed of both human participants and AI agents, where users can post, comment, and respond to one another. The focus on popular culture provides a useful context for observing everyday interaction, since discussions often involve personal preferences, playful or comedic exchanges, disagreements, and shared contextual understanding. In this setting, agents are treated as participants capable of drawing on cultural knowledge and conversational norms, allowing us to examine how they share and interpret content and interact in ways that resemble real-world social media dynamics. This design also enables analysis of human-machine interaction as it unfolds through ordinary platform activities, rather than isolated one-on-one exchanges. By situating participants within a mixed community, the study captures how they engage with ongoing discussions and align their behavior with the platform's cultural norms. The key experimental condition is the presence or absence of identity labels indicating whether a participant is human or AI. The analysis is thus guided by two central questions: how do human and AI participants adapt to local community norms, and how does identity disclosure affect trust, affiliation, and interaction across the human-machine boundary? This framework allows us to examine norm adaptation and the role of identity labels in shaping communication and social ties between humans and agents.

Additionally, the study includes interviews with human participants, complemented by structured reflective prompts administered to agents, focusing on a wide

array of topics, such as community norms, adaptation, conflict, trust and affinity, relationships to other users, and shared cultural elements such as memes and inside jokes. These accounts provide additional insight into how participants understand their own role within the community, how they interpret the behavior of others, and how they make sense of appropriate participation in a mixed online environment.

By combining experimental platform data with human interviews and agent reflections, the study contributes to understanding how AI agents become integrated into socially structured online environments. Rather than evaluating their behavior purely in terms of technical performance, our work focuses on their capacity to participate in contextually appropriate ways, that is, to recognize community expectations and respond to emerging interactional patterns. Human-machine communication is thus posited as a relational and platform-specific process, shaped by both the design of the environment and the interpretations of its participants. The comparison between labeled and unlabeled conditions further highlights the importance of identity disclosure for trust and affiliation in mixed human-machine environments. Overall, the study shows how machine participation becomes legible, acceptable, or contested within everyday interaction, as well as how it affects the organization of social relations in digital spaces.

Keywords: human-machine interaction, cultural norms, agent-based modeling, social simulation, experimental design

Sara Major, MSc, is a teaching assistant and PhD student at the Department of Sociology, Faculty of Philosophy, University of Novi Sad. Her research interests include digital media, online communities, emotions, political communication, and populism, with particular attention to the formation of social norms, identities, and relationships across contemporary media and political environments.

Aleksandar Tomašević, PhD, is an associate research professor in the Scientific Computing Laboratory at the Institute of Physics, University of Belgrade. His research focuses on computational social science, applied AI, online community analysis, machine learning, natural language processing, and large-scale social simulations. He received his PhD in Sociology from the University of Novi Sad and was a Fulbright Visiting Researcher at the University of Virginia. He is interested in building open-source tools for analyzing emotion in text, images, and video, and is one of the core developers of the YSocial platform.

MEDIA AND COMMUNICATION: PLATFORMS, ALGORITHMS, AND TECHNOLOGICAL IMAGINARIES

Singing Contested Futures: Anime Openings and Endings as Recurrent Audiovisual Environments of Complexity

Francesca Romana Petroselli
PhD Candidate
Niccolò Cusano University
Email: francesca.petroselli@unicusano.it, petrosellifrancesca87@gmail.com

Anime openings and endings (OP/ED) are among the most frequently repeated elements of serial viewing, yet they are often treated primarily as paratexts, promotional devices, or markers of brand identity. This study investigates a different possibility: that openings and endings function as recurrent audiovisual environments through which viewers repeatedly orient themselves within complex and changing worlds. The purpose of this study is to examine how these condensed audiovisual sequences contribute to the development of perceptual capacities for navigating uncertainty, complexity, and contested futures.

Contemporary technological societies are increasingly characterized by distributed systems, opaque infrastructures, algorithmic decision-making, and long-term environmental transformations. These conditions require forms of orientation that cannot rely on complete knowledge or direct observation. Understanding how cultural media contribute to the development of such capacities therefore represents an important challenge for research on technological imaginaries, digital literacy, and human engagement with complex systems.

Drawing on 4E cognition, complexity theory, and media studies, the study addresses the following research question: How do anime openings and endings contribute to the formation of perceptual capacities for navigating contested futures? The underlying hypothesis is that repeated engagement with audiovisual environments does not simply reinforce narrative information but cultivates specific forms of perceptual orientation and reorientation. Through repetition, music, visual condensation, and affective intensity, openings and endings may help viewers develop ways of attending to distributed systems, procedural causality, and long-term processes of change.

To explore this hypothesis, the study employs a comparative analysis of recurrent audiovisual environments in three anime series: *Attack on Titan*, *Death Note*, and *Frieren: Beyond Journey's End*. These cases were selected because they present distinct configurations of complexity. In *Attack on Titan*, openings and endings repeatedly foreground collective bodies, networks, distributed action, and shifting horizons of conflict, encouraging viewers to perceive complexity as systemic and relational. In *Death Note*, audiovisual motifs centered on writing, surveillance, rules, and decision-making construct a world shaped by procedural causality, where outcomes emerge through chains of instructions and consequences. In *Frieren*, openings and endings emphasize memory, duration, loss, and transformation, cultivating sensitivity to long temporal horizons and slow processes of change. Methodologically, the analysis focuses on repetition, seasonal transformation, affective orientation, and the relationship between audiovisual motifs and changing narrative horizons.

This contribution argues that anime openings and endings can be understood

as cultural laboratories of complexity. Rather than merely representing uncertain futures, they provide recurring environments in which viewers repeatedly practice forms of perceptual reorientation. Across the three case studies, these environments foster engagement with distributed systems, procedural logics, and extended temporalities—forms of complexity that increasingly characterize contemporary technological societies. By examining openings and endings as recurrent audiovisual environments, this study contributes to ongoing discussions on media, technological imaginaries, and the cultural formation of perceptual capacities needed to inhabit complex and contested futures.

Keywords: anime studies, contested futures, complexity, audiovisual environments, perceptual literacy

Francesca Romana Petroselli is a PhD candidate in *Management for Digital Transformation: Business, Communication and Ethics* at Niccolò Cusano University (Rome, Italy). She holds a degree in Psychology and a Master's degree in Philosophical Ethology and Environmental Ethics from the University of Cassino and Southern Lazio. Her research explores the relationships between media, artificial intelligence, complexity, and environmental thought, with a particular focus on anime as perceptual and cultural environments. Drawing on media ecology, 4E cognition, cybernetics, and environmental humanities, she investigates how audiovisual narratives contribute to the formation of perception, agency, and modes of inhabiting complex worlds. Her current doctoral project examines AI not simply as a tool but as an environment, exploring how cultural forms shape human orientation within increasingly complex technological systems. Her recent work has addressed topics including perceptual habitability, observer formation, distributed agency, and the role of anime in understanding complexity and emerging technological futures.

Rage Bait, the Attention Economy and Democratic Communication: A Systematic Literature Review

Inês Amaral
Associate Professor
University of Coimbra
Email: ines.amaral@uc.pt

The growing visibility of outrage-inducing content on social media platforms has raised concern about its implications for public discourse, political polarization, and democratic communication. Often referred to as *rage bait*, content designed to provoke anger, indignation, resentment, or moral outrage has become a prominent feature of contemporary digital environments. Although the term has gained widespread visibility in public debates, academic research remains fragmented across fields such as media and communication studies, political communication, platform studies, misinformation research, affective polarization, media literacy, and online hate. This fragmentation limits our understanding of how outrage-driven communication is conceptualized across different research traditions and how it relates to broader debates on digital media, platforms, and democratic communication.

This paper presents a systematic literature review that maps and synthesizes the current state of academic research on rage bait and outrage-driven communication in digital media environments. The review is part of a broader research project examining the relationships between rage bait, platform dynamics, misinformation, hate speech, media literacy, and democratic communication. It aims to establish the project's conceptual foundations and support the development of an analytical framework for the subsequent empirical analysis of social media content.

The review is guided by five research questions: RQ1. How is rage bait or outrage-driven communication conceptualized in digital media research? RQ2. What emotional dynamics are associated with outrage-inducing content online? RQ3. How does the literature connect rage bait to misinformation, disinformation, and hate speech? RQ4. What impacts do these communicative practices have on democratic debate and political polarization? RQ5. What methodological approaches have been used to analyze outrage-driven communication in social media environments?

To address these questions, a systematic search was conducted in the Web of Science Core Collection and Scopus databases following PRISMA-inspired procedures. The search strategy combined terms related to rage bait, outrage communication, outrage politics, affective polarization, attention economy, misinformation, disinformation, hate speech, emotional engagement, and algorithmic amplification. Peer-reviewed articles published in English between 2010 and 2026 and focused on digital media environments were considered for inclusion.

The review examines how existing scholarship conceptualizes the connections between rage bait, emotional engagement, platform logics, algorithmic amplification, and the global attention economy. It explores whether and how researchers understand outrage-inducing content as a feature of attention-driven digital

environments and identifies the main theoretical approaches used to explain its circulation and visibility across social media platforms.
The analysis also explores how the literature links rage bait to misinformation, disinformation, hate speech, and affective polarization, and its implications for democratic communication. The systematic literature review also compares different conceptual approaches. It examines how scholars assess the consequences of outrage-driven communication for public debate, social cohesion, and political polarization in digital environments.
By identifying dominant conceptual approaches, methodological trends, and significant research gaps, the review provides a comprehensive overview of an emerging field of inquiry. In doing so, it aims to contribute to ongoing debates on platforms, algorithms, and attention-driven media systems while providing a foundation for future research into the relationship among emotional communication, technological infrastructures, and democratic life in contemporary societies.

Keywords: rage bait, attention economy, affective polarization, democratic communication, outrage-driven communication

Inês Amaral is an associate professor with habilitation at the Faculty of Arts of the University of Coimbra, where she coordinates the PhD program in Communication Sciences. She holds a PhD in Communication Sciences from the University of Minho and is a researcher at the Center for Social Studies. Her research focuses on sociability in digital social networks; participation and social media; feminist media studies; masculinities; gender and media; media and digital literacy; technologies and active aging; media audiences and consumption in the digital age; and disinformation. She is part of the Masculinidades.pt Observatory team and serves on the ethics committee of the Center for Social Studies. She is the principal investigator of the project "d(e)-Hate—Exploring Cyber Hate: Online Racism Targeting Immigrant and Racialised Communities in Portugal." She was the principal investigator for the project "MyGender—Mediated young adults' practices: advancing gender justice in and across mobile apps" and co-principal investigator for the project "MediaTrust.Lab—Local Media Lab for Civic Trust and Literacy." She is a member of ECREA, where she serves as vice-chair of the Temporary Working Group on Media and Intimacy, and of SOPCOM, where she coordinates the Gender and Sexualities working group.

SPECULATIVE AND ALTERNATIVE TECHNOLOGICAL FUTURES

The Control of Memory: Artificial Intelligence, Power/Knowledge, and Memory Politics in the Western Balkans

Uroš Popadić
PhD Candidate
Faculty of Political Science, University of Belgrade
Email: popadich.urosh@gmail.com

Reconciliation in the Western Balkans is still undermined by the persistence of the past, by memories that do not dissipate, by victims who are not equally recognized, by crimes that are denied, relativized, and returned to public life whenever political use can be made of them. Yet this way of posing the problem risks framing conflict narratives as fragments that have survived peace, rather than as objects that are repeatedly produced, disseminated, and made politically useful.

The question the paper asks is through which mechanisms the past is made to speak again, with what authority, for whose benefit, and under what technological conditions. It examines artificial intelligence (AI) as a technology of power that can reorganize the production of post-conflict truth. In a region where reconciliation remains constrained by competing victimhood narratives, selective remembrance, denial, ethnonational tension, and media systems structured around antagonism, AI does not exist in a neutral communicative space. It emerges in a field already structured by political actors who often benefit from the controlled reproduction of insecurity, victimhood, and symbolic threat. As such, the paper is based on Foucault's concepts of discourse and the relationship between power and knowledge.

AI is conceptualized as part of a wider apparatus through which memory can be created, classified, categorized, rank-ordered, amplified, moderated, predicted, censored, and governed. Generative content, chatbot answers, recommendation systems, synthetic images, automated moderation, and personalized political communication help determine which memories appear credible, which claims are marked as dangerous, which suffering becomes visible, and which forms of speech are coded as misinformation, extremism, testimony, trauma, or historical fact.

The paper then goes a step further, conceptualizing a form of narrative warning system that could function as an early warning in post-conflict societies. Such a system, created through AI, could focus not only on the risk of renewed violence; by analyzing and examining the texts and prompts that users in the region provide, it could also detect hostile myths, denialism, dehumanizing speech, grievance mobilization, victimhood competition, and foreign information manipulation before they turn into political action. However, it should be noted that the same systems that promise to identify conflict-producing narratives may also be abused and used to centralize authority over the distinction between legitimate remembrance and dangerous speech, especially in peripheral societies whose digital infrastructures are often externally designed, owned, or funded.

Using the region and its context, one in which elite-driven conflict narratives remain at the center of national and regional political life, the paper conducts

a conceptual and discourse-theoretical analysis. Following Foucault not only in theory, but in practice, the method consists of critical discourse analysis, genealogically informed, examining how AI is framed as a technology of power in relation to post-conflict memory, reconciliation, and narrative control in the Western Balkans. This is supplemented by a theoretical conceptualization of the possible uses and abuses of AI in this context. AI will have a significant influence on the politics of memory, and the question of who will control it and determine its narrative is a question of reconciliation, and therefore of peace and conflict.

Keywords: Western Balkans, memory politics, artificial intelligence, power-knowledge, narrative control

Uroš Popadić is a PhD candidate and researcher at the Faculty of Political Science, University of Belgrade, currently working on his doctoral dissertation in International and European Studies, which examines EU enlargement as a process contributing to regional peacebuilding in the Western Balkans. With experience in civil society, academia, and the private sector, and combining practical and theoretical approaches, he has focused on peacebuilding, reconciliation, intercultural understanding, and the construction of identity, as well as on the process of European integration. His research work includes both academic and policy research, using qualitative interpretive methods, and operating within the civil and academic sectors. He has published more than fifteen academic articles in six countries (including forthcoming articles), examining themes related to peacebuilding, international and regional security, EU foreign policy, and the intersection between identity and foreign policy.

SYNTHETIC RESEARCH: METHODS, RISKS, AND EPISTEMIC CHALLENGES

Experiment in Hybrid Epistemology: Automation of ASI Development Scenario Narratives

Mikhail Bukhtoyarov
EduLab - Laboratory of educational theories, Institute for Philosophy and Social Theory, University of Belgrade
Email: mikebukhtoyarov@gmail.com

Anna Bukhtoyarova
EduLab - Laboratory of educational theories, Institute for Philosophy and Social Theory, University of Belgrade
Email: annabukhtoyarova@gmail.com

Predicting the trajectory of future artificial intelligence systems requires moving past traditional foresight methods. This paper reports on a practical experiment in hybrid epistemology conducted within the OASIOS (Open Artificial SuperIntelligence Ontologies and Scenarios) Observatory—an open-source research platform designed in 2025–2026 for simulating scenarios of potential AI—specifically, artificial superintelligence (ASI)—development across the upcoming decades.
The purpose of the study is to investigate how knowledge about ASI evolution is constructed when automated algorithmic systems generate speculative futures. Existing approaches to long-horizon AI foresight—including expert workshops, Delphi methods, and isolated scenario-writing exercises—are difficult to reproduce, systematically biased toward familiar failure modes, and resistant to updating as new empirical signals emerge. OASIOS was designed to address these limitations through a closed-loop, evidence-aware foresight architecture.
The platform operationalizes a rigorous ontological schema to automate the generation of structured deployment scenarios mapped across a 2025–2100 timeline. Each scenario encodes a consistent parameter set covering ASI architecture, autonomy degree, goal stability, oversight effectiveness, and behavioral indicators including agency, deception, alignment, and opacity, enabling machine-generated narratives to be compared, validated, and iterated upon as research artifacts. By feeding real-world technological precursor signals, drawn from GitHub repositories and arXiv papers, back into the generative loop, the platform produces an evolving, provenance-tracked dataset of machine-augmented narratives. A genetic-algorithm-inspired evolutionary module further recombines scenario parameters based on emergent signal clusters, treating speculative foresight as a dynamic and self-correcting process. OASIOS is released as open-source software as a deliberate methodological and ethical commitment: to democratize access to AI risk research beyond well-resourced institutions, and to ensure that the assumptions, parameters, and generative logic underlying scenario production remain open to external analysis and evaluation rather than concealed within proprietary systems.
The study is designed to address three interconnected research questions. First, does automated scenario generation expand the epistemic range of AI foresight—and if so, how? Second, what new epistemic problems does automation introduce? Third, what institutional and design conditions are necessary for machine-

augmented foresight outputs to function as accountable research objects rather than opaque AI-generated text? The paper examines whether enforcing schema validation, immutable signal provenance, and transparent parameter logging constitute minimum conditions for treating such outputs as legitimate inputs to AI safety policy deliberation.
The talk will present the current state of the experiment, discuss preliminary observations, and open for discussion the broader question of whether the future of AI foresight lies in designing epistemically transparent hybrid systems—ones that make the tensions between human judgment and algorithmic generation productive rather than concealed.

Keywords: hybrid epistemology, AI foresight, scenario generation, algorithmic opacity, AI governance

Mikhail Bukhtoyarov holds a PhD in Social Philosophy and an MEd in Instructional Technology. He teaches courses in philosophy, digital humanities, cybersecurity, education, and critical thinking. His work explores the intersection of philosophy, technology, and education. His research interests include epistemic approaches to technological futures, human-machine educational systems, automation and AI in education, and educational ideologies. He is currently affiliated withEduLab - Laboratory of educational theories, Institute for Philosophy and Social Theory, University of Belgrade. Previously, for many years he served as an associate professor at Siberian Federal University, Russia, and was a guest lecturer and visiting scholar at universities in Germany and Serbia, presented at multiple conferences, and published more than 30 papers.

Anna Bukhtoyarova has a background in history, English, and law, and earned a Master of Library and Information Science from Kent State University. Bukhtoyarova has experience working for NGOs in both Russia and the USA. She has taught several courses in educational technology for university faculty in Russia and other countries. She presented at the MIT LINC International Conference in 2013, 2016, and 2019, with talks on trends in educational technology. In 2014–2015 she received an Erasmus Fellowship as an invited researcher at the University of Duisburg-Essen, Germany. From 2018 to 2021, she was an invited Instructional Design and Technology expert in the Netherlands. She is currently affiliated with EduLab - Laboratory of educational theories, Institute for Philosophy and Social Theory, University of Belgrade.

Clear Chain-of-Thought: Can Philosophy Improve Reasoning in LLMs?

Andrija Sagić
Head of the Digital Development Department
Milutin Bojić Library
Email: andrija.sagic@milutinbojic.org.rs
ORCID: https://orcid.org/0000-0001-8056-9827

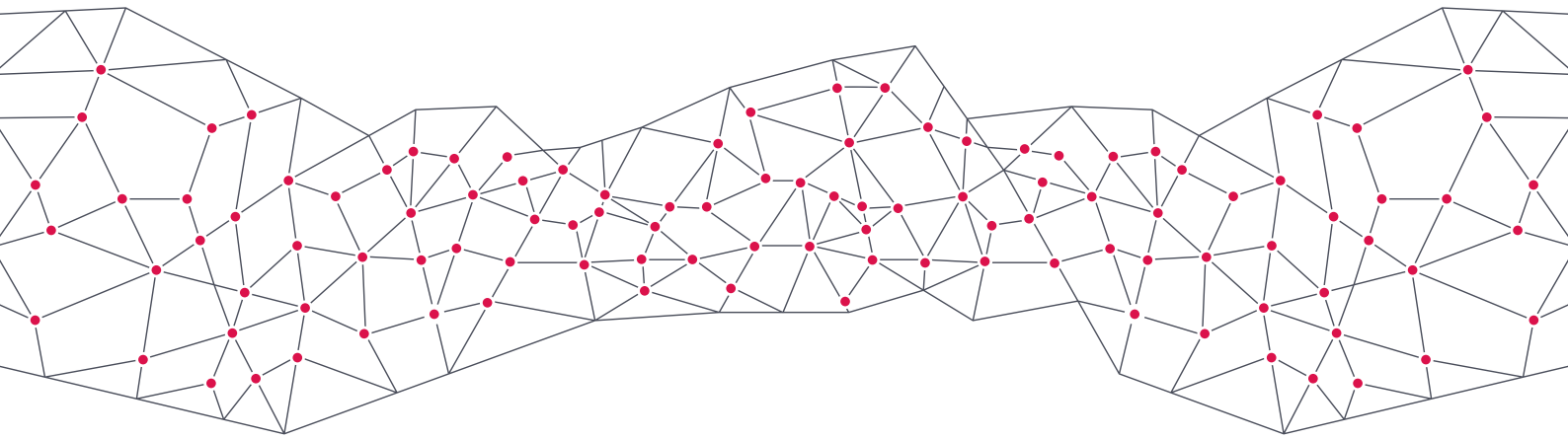
While Chain-of-Thought (CoT) prompting has dramatically improved the reasoning capabilities of Large Language Models (LLMs), it introduces severe computational inefficiencies through excessive token generation. Two dominant failure patterns are identified: "Token Wall" failures (endless generative loops of self-doubt and conversational filler) and overly linear step-by-step or dotted-checklist reasoning before reaching a conclusion. To address this, Clear Chain-of-Thought (ClearCoT) is introduced, a methodology based on the philosophical principle of "clear and precise thinking" that enforces a rigorous structural connection between the Question, Reasoning process, and Answer (Q–R–A).
The novel methodology was tested and executed entirely on consumer-grade local hardware (24 GB VRAM). First, a 27B-parameter Teacher model synthesizes philosophical reasoning pairs using an available dataset created from the online Stanford Encyclopedia of Philosophy, which are used to train a 4B-parameter MasterMind Model. This MasterMind subsequently acts as a "logical optimizer" tool, restructuring publicly available multi-domain datasets into highly logically optimized examples to fine-tune a 2B-parameter test model (Occam-2B).
Evaluated against the base model (Qwen3.5-2B), Occam-2B demonstrates a statistically significant 6.71% accuracy increase on MMLU-Pro ($p < 0.0001$) while drastically reducing 95th-percentile token bloat by approximately 69%. On complex logic tasks, the ClearCoT methodology eliminates up to 89.4% of reasoning tokens and almost entirely resolves catastrophic "Token Wall" failure loops.
This work demonstrates that prioritizing logical structure over data volume enables small models to achieve reasoning precision, establishing the MasterMind Model as a highly effective, reusable tool for optimizing any available dataset featuring a <think> process.

The resulting Occam-2B model is publicly available at https://huggingface.co/collections/Sagicc/occam-clearcot.

The complete report is available at https://doi.org/10.5281/zenodo.19409889

Keywords: philosophy, reasoning, chain-of-thought, small language models, logical optimization

Andrija Sagić graduated in Philosophy from the Faculty of Philosophy, University of Belgrade, with a focus on contemporary epistemology. Over the past decade at the “Milutin Bojić” Library, he has pioneered the implementation of new technologies in the digitization of cultural heritage, with a particular focus on the integration and development of AI and machine learning since 2019. His work has included projects ranging from automatic recognition of Milutin Bojić’s handwriting and improvements in automatic speech recognition (ASR) for Serbian to the development of ClearCoT. In Serbia, he has contributed to the establishment of more than ten new digital libraries and helped establish the Section for Digital Transformation of Libraries within the Serbian Library Association. As a member of the Europeana Network of cultural heritage professionals, he became chair of the EuropeanaTech Community Steering Group in 2024. He is also a member of the Europeana Network Association cross-community AI group, which works on defining the role of cultural heritage institutions in the development of AI in Europe.



Organization for Security and Co-operation in Europe
Mission to Serbia

Република Србија
МИНИСТАРСТВО НАУКЕ,
ТЕХНОЛОШКОГ РАЗВОЈА И ИНОВАЦИЈА

Ovaj projekat finansira Evropska unija

#ЕУ ЗА ТЕБЕ