GO TO SESSION 1 GO TO SESSION 2

Panel 78. From 'Digital Technology' to 'AI': Emerging Challenges in the Making

Convenor:

Annapaola Ginammi, Politecnico di Milano

12 JUNE 2025 11.30 - 13.00 SESSION 1

ID 206 - AI as War Machine: Digital Ideology, Legal Surrender, and the Privatization of Everyday Practices

Joaquín Santuber, Johannes-Kepler-Universität

Marcos Chilet, Pontificia Universidad Católica de Chile

Kristina Tica, Universität für künstlerische und industrielle Gestaltung Linz

Pablo Hermansen, Pontificia Universidad Católica de Chile

Keywords: AI War Machine, Digital Ideology, Privatization of Everyday Practices, Decolonial Resistance, Ch'ixinakax utxiwa

Building on Deleuze and Guattari's notion of the war machine, this work frames artificial intelligence (AI) as a technology that exceeds state-centric legal boundaries, strategically reconfiguring the everyday. We argue that the European Union's AI Act not only surrenders to AI's deterritorializing logic but also furthers an ideology of inevitability, immateriality, and neutral progress. Legislation purporting to regulate AI effectively legitimizes its political power: rather than challenging the underpinnings of extractive data regimes, the law stabilizes illusions that veil the material and social costs of AI's proliferation.

We integrate Matteo Pasquinelli's Eye of the Master (2023) to highlight a crucial dimension of this war machine: AI emulates – and consequently privatizes – human practices accrued over millennia. Pasquinelli shows how AI operates as a platform that reifies centuries-old communal knowledge, transforming it into proprietary datasets and algorithms that privilege corporate interests. This privatization of everyday life extends previous colonial-enclosure logics, extracting value from intangible activities such as communication, creativity, and routine gestures. The result is a new form of dispossession, where data drawn from collectively evolved human practices is repackaged and locked behind corporate firewalls.

In parallel, the EU AI Act's "human-centric oversight" requirements hinge on the illusion that legal disclaimers or "stop buttons" can tame systems that are already structurally enmeshed with everyday human behaviour. Such provisions fail to address how AI's illusions of omniscience and neutrality actively forestall democratic accountability, while laboring bodies – particularly in the Global South – remain undercompensated and invisible. By permitting only superficial transparency, the law consolidates AI's hegemonic status and forestalls more radical critiques of how AI redefines social coordination.

Additionally, the ideology of the war machine obfuscates AI's dependence on exploitative supply chains and planetary extraction, from rare minerals to water-intensive server infrastructures. Leveraging the concept of Ch'ixinakax utxiwa (Rivera Cusicanqui, 2010), we examine how Latin American forms of resistance can unmask these "dark infrastructural" layers. Practices that embrace multiple, coexisting worlds and hybrid ontologies challenge the homogenizing thrust of AI, underscoring the need for collective, situated responses that depart from mere compliance with Northern regulations.

Ultimately, we ask how communities – especially those in marginalised contexts – might cultivate "new weapons" (Deleuze, 1992) of socio-technical resistance. We identify collaborative design approaches, data cooperatives, and subversive art interventions that expose AI's illusions and reclaim human practices from corporate appropriation. By situating these efforts within Pasquinelli's critique of AI's historicity and Rivera Cusicanqui's decolonial lens, our work offers a pointed reflection on the political potency of AI as war machine, the legal apparatuses that sanction it, and the possibilities for democratic, multispecies futures that refuse to surrender everyday life to privatised digital rule.

12 JUNE 2025 11.30 - 13.00 SESSION 1

ID 402 - Culture and AI: Moral considerations of Technology in Public Discourse

Vittoriana Loporcaro, Unisalento
Sergio Salvatore, Unisalento
Fiorella Battaglia, Unisalento
Matteo Reho, Università di Roma La Sapienza
Arianna Dini, Unisalento

Keywords: Cultural Representations, Digital Transition, Artificial Intelligence (AI), Semiotic Theory of Cultural Psychology (SCPT)

The present study investigates the cultural representations of the digital transition, focusing on the role of Artificial Intelligence (AI) and its implications for cultural production and consumption. The objective is to establish ethical and inclusive regulatory guidelines for AI, ensuring that its deployment does not reinforce social and cultural inequalities. The research is grounded in the Semiotic Cultural Psychology Theory (SCPT), which conceptualizes meaning-making as a function of latent symbolic universes (Salvatore S., 2020). Symbolic universes are generalised, affect-laden assumptions that shape collective sense-making and guide social action. These symbolic universes represent the most abstract level of social representations and serve as a framework for analysing the intersection between culture and policymaking.

Methodologically, the study employs the Automated Co-Occurrence Analysis for Semantic Mapping (ACASM), an approach that integrates text mining techniques and lexical correspondence analysis (LCoA) to identify underlying semantic structures in media discourse. The corpus consists of newspaper articles and reports addressing AI and its cultural implications. The analytical procedure follows three key phases:

Corpus Selection: Relevant textual materials are selected based on thematic pertinence, prioritizing sources discussing the digital transition in a socio-cultural context.

Semantic Analysis: Textual data are processed using T-Lab software, which identifies recurring thematic clusters and semantic polarities.

Qualitative Interpretation: Quantitative findings are contextualised through the SCPT framework to identify dominant symbolic universes.

The results highlight two competing narratives framing the digital transition: AI as a driver of cultural democratization versus AI as a threat of cultural homogenization. These representations are anchored in contrasting symbolic universes: the "Ordered Universe," which, as mentioned above, associates AI with technological progress and societal trust, and the "Niche of Belonging," which emphasizes vulnerability and perceived threats.

The proposed methodology enables a comprehensive understanding of how media shape and disseminate shared meanings concerning AI, influencing both public perception and policymaking decisions. The media's power to shape shared meanings of AI reinforces the necessity of regulatory frameworks that integrate cultural complexity into the governance of AI.

A key contribution of this study is the identification of interconnections between media representations and collective sense-making processes. Prior research has demonstrated that media discourse not only reflects but also, actively constructs, social reality – thereby - influencing the ways in which AI is understood and integrated into society.

The relationship between symbolic universes and policymaking thus emerges as a critical factor in ensuring that AI deployment does not exacerbate existing inequalities. By bridging cultural analysis with technological governance, this study provides a robust framework for developing ethical guidelines that promote inclusivity and cultural sustainability in the digital era.

12 JUNE 2025 11.30 - 13.00 SESSION 1

ID 419 - How are Large Language Models transforming medical self-diagnosis?

Lara Dal Molin, University of Edinburgh

Keywords: Large Language Models, self-diagnosis, algorithmic transformations

On November 30th, 2022, the global release of OpenAI's ChatGPT permanently shifted human interactions with Artificial Intelligence (AI) algorithms. Previously experimental artefacts within a small network of 'expert' developers and users, Large Language Models (LLMs) are currently at the centre stage of human-computer interaction – with ChatGPT alone figuring between 200 and 300 million weekly active users worldwide (Backlinko, 2025; Reuters, 2024). ChatGPT is one of many LLMs available through online user interfaces: other popular models include OpenAI's o1, Google's Gemini, Anthropic's Claude and Microsoft's Copilot. In their most common form, LLMs generate humanlike text through combining statistical probability distributions with AI algorithms called Transformers, which mimic human attention by processing a word while focussing on the surrounding context in the input text (Vaswani et al., 2017). Technology companies advertise LLMs as productivity tools, supporting a variety of tasks including text summarisation, computer code generation and debugging, general brainstorming, web browsing and data analysis. However, the advanced capabilities of LLMs combined with their pervasiveness in terms of access elicit a complex social and ethical picture. Both popular and scholarly accounts suggest that users worldwide substantially engage with LLMs for purposes other than productivity, including health mental support and companionship (Metz, 2020; Siddals, Torous and Coxon, 2024).

In the context of "re-ordering care", this contribution considers the emerging role of LLMs in medical self-diagnosis. Balasubramanian and Dakshit (2024) suggest that the use of LLMs for this purpose naturally follows – and might supersede – engaging with online resources and applications, such as symptom checkers. Hence, the trustworthiness and reliability of LLMs in producing medical diagnoses should be assessed. On the one hand, disciplinary literature across both the medical field and Natural Language Processing is persistently exploring the applicability of LLMs to the diagnostic practice, in an attempt to automate, accelerate and optimise access to medical knowledge (Nazi and Peng, 2024). Here, optimistic accounts report LLMs 'outperforming' medical and mental health professionals in specific tasks, such as accurately identifying the symptoms of obsessive-compulsive disorder (Kim et al., 2024). On the other hand, critical perspectives emphasise the pitfalls of these approaches, and warn against the biases and misinformation that LLMs might perpetuate (Barnard et al., 2023; Ziaei and Schmidgall, 2023). Vulnerable and minoritised communities are especially susceptible to these risks. For instance, Chang et al. (2024) found that four LLMs generate inappropriate – meaning unsafe, biased and inaccurate – responses to prompts where LGBTQIA+ identity is relevant to clinical care. As part of this panel, this contribution will situate the emerging use of LLMs for self-diagnosis as an epistemic and normative transformation in medical knowledge and practice.

As a nascent topic of scholarly inquiry, it will attempt to map its current sociotechnical landscape through considering the following questions. Does using LLMs for self-diagnosis significantly depart from consulting online resources through search engines? Why and how would patients trust LLMs with diagnostic practices over traditional medical institutions? What are the practical and ethical challenges that researchers may encounter when attempting to study these phenomena?

12 JUNE 2025 11.30 - 13.00 SESSION 1

ID 462 - Exploring the Moral Acceptability of AI in Professional Settings: The Role of Human Product Characteristics in Shaping Perceptions

Tiffany Morisseau, Université Gustave Eiffel
Giusy Cirillo, Strane Innovation
Julien Cestac, Université Gustave Eiffel

Keywords: AI, Workplace trust, Creativity, Transparency, Perceived Fairness

The spread of generative AI tools presents new challenges, as people gradually learn to integrate them into their daily tasks. This is particularly evident in professional contexts, where the adoption of these systems has revolutionised the way work is performed, speeding up processes and often leading to more productive outcomes. In many cases, AI tools enable better results by allowing professionals to focus on tasks requiring thinking and strategy, while the AI handles calculations and repetitive work. However, the widespread adoption of these tools also presents risks to the originality of human-created products, raising two key concerns: 1) the potential for increased inequality and plagiarism, and 2) the erosion of trust within the workplace. This is especially problematic because, in many instances, institutions have yet to implement ethical guidelines or official rules that ensure AI is used transparently and with clear boundaries. To underscore the importance of preserving fairness and equity in the workplace, we have developed two experimental protocols aimed at measuring individuals' perceptions of generative AI tools in professional contexts. In two scenario-based experiments featuring realistic AI use scenarios, we seek to assess how manipulating specific features of human-created products influence the perceived fairness and acceptability of AI use.

The first experiment focuses on two key factors: transparency and creativity. Participants are asked to evaluate AI use based on 1) whether the agent is transparent about using AI and 2) the nature of the work – whether it involves creativity (e.g., artistic tasks) or is more repetitive and mechanical. We measure participants' moral judgments regarding AI use, the agent using AI, and their confidence in the agent's skills. We hypothesize that greater transparency in AI use will lead to higher moral acceptability, while more creative tasks will be seen as less acceptable for AI involvement. The second experiment examines the moral acceptability of AI in recruitment, specifically its use in evaluating and generating CVs and cover letters. We hypothesize that AI-generated or -evaluated documents will be rated more negatively than those created without AI. Additionally, we expect a stronger preference for human-authored or -evaluated cover letters over CVs, as cover letters are viewed as more creative and subjective. The results of these two experiments will be discussed to highlight the ethical implications for AI's use in professional environments, with a focus on preserving fairness, creativity, and trust in workplace practices.

More generally, our work offers a novel approach to understand the potential negative impact of AI on interpersonal trust in professional and economic settings. We believe it is crucial to first consider individuals' moral attitudes toward AI in the workplace, and then reflect on how to develop an ethical framework for its use, as individuals are the building blocks of a (professional) community. This framework should be transparently designed, and approved by all members of a community or institution, fostering cooperation and trust – fundamental elements of any successful organization.

12 JUNE 2025 11.30 - 13.00 SESSION 1

ID 480 - Corporate Influence Over Open Sociotechnical Systems: How Linux Was Transformed From a Subversive Technology Into the Centrepiece of Digital Capitalism

Davide Carpano, New York University

Keywords: Sociotechnical imaginaries, open source, material political economy

In this paper I analyse corporate influence over free and open source software (FOSS) through a case study of IBM's adoption of the Linux operating system between 1998 and 2003. Through a combination of archival research and oral history interviews I examine how IBM played a role in reshaping Linux from a project focused on liberating computer users from proprietary software controlled by large corporations to one primarily oriented toward the needs of large-scale enterprise systems. This project builds on scholarship in economic sociology and science and technology studies that analyses the role that imagined futures (or fictional expectations) play in the economy and in shaping trajectories of technological development (Beckert 2016; Jasanoff and Kim 2015). While many studies that use this framework focus primarily on the rhetorical and ideological aspects of imaginaries, my approach seeks to identify how imaginaries performatively shape the material world. To do this, I adapt Donald MacKenzie's (2021) approach to the study of sociotechnical systems called material political economy. This approach asks us to make sense of how the material world is ordered through analysis of the political conditions that lead to a particular ordering, and the economic consequences of that ordering. Using this method I trace IBM's interactions with the Linux community to show how the company mobilised its economic and human resources in order to gain influence over the OS's development trajectory and instigate changes to critical subsystems of the OS.

I argue that IBM's understanding of Linux was shaped by a set of fictional expectations shared by key executives which predicted a world in which the spread of the Internet would force IT companies to abandon proprietary technologies and adapt to open standards. Influenced by this imagined future, the company sought to create a new business model that shifted revenue away from hardware and software sales and towards IT services in a move that prefigured cloud computing, and modern approaches to 'Platforms/Software as a Service.' Under this context, IBM came to see Linux as the ideal OS on which this new model of computing would be built. In the years since, Linux has become a key part of the technological infrastructure of digital capitalism and facilitated the rise of big tech. While scholars and activists have largely viewed the open source development process as being inherently democratic by virtue of its open governance structure, my research documents the ways in which actors and organizations can position themselves as obligatory points of passage in order to exert influence over key technical decisions.

References:

Beckert, Jens. 2016. Imagined Futures: Fictional Expectations and Capitalist Dynamics.
Jasanoff, Sheila, and San-Hyun Kim, eds. 2015. Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power.
MacKenzie, Donald. 2021. Trading at the Speed of Light.

12 JUNE 2025 14.00 - 17.00 SESSION 2

ID 517 - Can Machine Learning be democratised? Automated Machine Learning and the impossibility of a multiplicity of objectives

Laura Kunz, Universität Graz

Juliane Jarke, Universität Graz

Keywords: Keywords: AutoML, democratization of AI, multiobjective optimization

Automated Machine Learning (AutoML) has been praised as an advancement in Machine Learning (ML), foremost used as practice of prediction and support in decision making. It is implemented through so-called Low Code or No Code Development Platforms (LCDP or NCDP), providing a visual interface guiding the development process of ML-based systems, thus requiring minimal or no coding expertise. Therefore, AutoML is promoted as an ideal tool for anyone who owns data and/or wishes to process data. It has thus been praised as democratising access to Machine Learning (Alamin & Uddin, 2024; Scriven et al., 2024).

This paper questions the potential of AutoML to democratise Machine Learning by arguing that the logics inscribed into ML foreclose "democratic control of algorithm objectives" (Kasy 2024). In other words: even if more people gain access to ML through LCDP or NCDP, they have little control over what to use it for. It aligns with scholars in critical data studies who have argued that Machine Learning is inherently foreclosing a multiplicity of practices and relations in the different social domains it is introduced to. Hence rather than opening up possibilities and allowing democratic control about the objectives of its use, ML allows only for particular engagements with the world. These are based on logics of optimisation and efficiency (Birhane et al 2022) and normalises "fascist" ideologies of categorising and classifying people (McQuillan 2022).

The paper is based on a survey of 198 commercial AutoML applications and an analysis of the ways in which their application (use cases) is promoted and described by AutoML providers. We compare how AutoML providers construct "real world problems" and how they can be solved through AutoML. The analysis demonstrates how the epistemological concept inscribed into Machine Learning (i.e., logics of calculation and data analysis in ML) only ever promotes certain solutions, i.e. approaches of optimization. These are, in the case of AutoML, in particular a) optimizing speed leading to cost optimization and b) informing decision-making to support cost optimization. The paper, hence demonstrates the tension between current "algorithm objectives" as afforded by AutoML and the claim for democratization.

Going beyond this point we outline how the epistemology inscribed into Machine Learning, thus AutoML, only allows for a single objective approach, single understanding of what can be optimised, contradicting a notion of democratization as allowing for a multiplicity of objectives in what to use AutoML for. Therefore, we argue that we need to investigate new ways of approaching (Auto)ML in order to allow for participation beyond giving access to executing the same single objective of cost reduction and profit maximization - allowing participatory usage with a wider set of objectives in usage.

12 JUNE 2025 14.00 - 17.00 SESSION 2

ID 563 - Ask Me Anything! How ChatGPT Got Hyped Into Being.

Jascha Bareis, Institute for Technology Assessment and Systems Analysis

Keywords: ChatGPT, chatbot, hype, BigTech, sociology of expectations, Large language model, LLM, Xrisk, catastrophic risk, hallucination, spectacle, Silicon Valley, Longtermism

This paper reconstructs how chatbots based on Large language models (LLMs) like ChatGPT got hyped into being. It dissects the actors and dynamics that triggered, fuelled and disseminated the hype. Through the lens of hype studies the paper interrogates three empirical realms: 1. Company websites where the chatbots are presented, 2. Blog entries and newspaper interviews by prominent tech figures from the Silicon Valley, and 3. New York Times articles in the timespan between November 2022 and August 2024. The

paper shows how the chatbot hype is driven by a dynamic between privileged actors (hypers) and a media frenzy both influencing and being carried by society and politics alike. Different interdependent building blocks in the chatbot hype construction are identified: 1. Strategic ignorance: Depicting Large Language Model (LLM) chatbots as knowledge models. 2. The weird and eerie: Panicking about the uncanny side of chatbots. 3. The battle: Staging a spectacle of competition between tech giants, and 4. Crossing the line of the normal: Praising the dualism of an apocalypse or a tech-religious calling. The paper unravels the core circulated narrative that turns the hype into a powerful societal phenomenon.

12 JUNE 2025 14.00 - 17.00 **SESSION 2**

ID 646 - Institutional Agency and Collective Responsibility for AI Governance

Miguel Garcia, Università di Bologna

Keywords: Institutional agency, collective responsibility, AI governance, AI agencies, EU AI Act

The global deployment of AI systems presents serious ethical and legal challenges, especially in relation to the assignment of responsibility to the institutional agents (e.g., corporations, states, and international organisations) that employ these technologies when carrying out specific activities. In particular, the autonomous nature of AI, which makes it capable to make decisions without direct human oversight, complicates the way in which current ethical and legal frameworks conceptualise their ontology and evaluate their normative implications.

With most contemporary ethical and legal theories seeing AI not as an agent in its own right but rather as an extension of institutional intentions (shaped by the objectives, data, and algorithms programmed into it), a critical question arises: Should responsibility rest solely with the developer, deployer, or end-user of an AI system, particularly when outcomes stem from emergent, unpredictable behaviours? Considering the disparities in AI regulation, with some jurisdictions assigning full responsibility to institutions and others highlighting the limitations of anticipating AI behaviours, this issue becomes an important obstacle in providing a general account of AI governance.

To address these challenges, a robust and multi-layered framework for institutional responsibility in AI contexts is necessary. To help in achieving this goal, this paper builds on recent developments within both institutional agency and collective responsibility to conceptualise AI as a collaborative tool of institutional action and justify the distribution of responsibility (viz., blame and compensatory duties) among all the relevant stakeholders (e.g., developers, providers, deployers, end-users, and regulators) that collectively contribute to the lifecycle of an AI system (from design and development to implementation and results).

A key innovation of this framework lies in its use of scenario-based risk analysis to evaluate the interactions of stakeholders within specific AI applications. By integrating this method into a governance structure grounded in institutional agency and collective responsibility, the framework aims to supply institutional actors with the tools to effectively assess and manage risks while ensuring compliance with ethical and legal norms (e.g., codes of practices, international regulations, etc). Through a re-examination of the foundational concepts of agency and responsibility, this approach not only advances understanding of the interconnected roles of stakeholders but also provides a pathway for ensuring that AI systems are governed by context-sensitive norms designed to regulate the varying institutional environments that constitute a significant part of our social reality.

12 JUNE 2025 14.00 - 17.00 **SESSION 2**

ID 699 - Techno-Economic Futurity for AI-enhanced Democracy?

Mark Coté, King's College London

Keywords: Data, AI, Democracy, Futurity

AI will only diminish democracy without a more critical understanding of and collective agency over the data that feeds its growth. This paper posits that a different form of datafication is a necessary socio-tech-

nical precondition for building an AI-enhanced democracy. It is the superabundance of data that makes possible DeepSeek and Chat GPT, advanced Machine Learning, and all the tech titans seated in a row on the dias inaugurating oligarchical power. This paper repositions data and AI in a new critical frame, triangulated on three concepts: Non-rivalry, Excludability, and Futurity. This foregrounds how data's expansive and generative qualities (see non-rivalry and futurity) are controlled, restricted and monetised by forces of excludability. This framework aims to bring a more finely-granulated and incisive understanding of data as the socio-technical object, which conditions the ways in which AI does and could impact democracy.

Non-rivalry and excludability are drawn from economic theory and capture inherent qualities or affordances of data. Non-rivalry denotes goods that can be simultaneously and repeatedly accessed or utilised by multiple individuals without reducing their value or availability. Data is inherently non-rivalrous, a quality noted by many but in cursory terms with little critical discussion. What has been missing is Excludability, that is, how data are subject to private market forces, including the ability to enforce restricted access. This artificial force, imposed on the inherent non-rivalry of data, supercharges capitalist value generation in part from the growing dependencies created by the infrastructure of datafication (Pybus et. al. 2025) play in determining how we are increasingly subject to the power of Data-AI oligarchies. Futurity provides an innovative critique of how excludable AI systems both depend on the data we collectively generate, as well as foreclosing more democratic futures.

I argue that futurity is a defining characteristic of data and of AI systems, alongside the economies they enable. I therefore outline AI's dual futurity – technical and economic. Technical futurity refers to the self-reinforcing life cycle of AI, where increased data availability leads to improved model training, enhancing predictive accuracy, and the possibility of repurposing into new domains. In turn this informs economic futurity, which serves as a fundamental metric as data's non-rivalrous nature ensures sustained and long-term value generation. Central to both are the functions that data play as a self-expanding asset within neural networks, continuously feeding back into models to refine and optimize performance. This techno-economic futurity therefore raises pressing concerns regarding ownership, access, and long-term rights. Current data-for-service models fail to account for ownership of future value, posing challenges for equitable data governance. If our collective datafication is to contribute to a more democratic AI future, mechanisms for ensuring fair ownership and benefit-sharing must become foundational preconditions.

I will conclude by bridging this theoretical framework with a practice, a HORIZON project that aims to develop a new paradigm for data sharing, a non-rivalrous and non-excludable end-to-end pipeline from user communities with agency over their data to advanced AI applications.

12 JUNE 2025 14.00 - 17.00 **SESSION 2**

ID 794 - Preparing future citizen for the post-truth world in the age of AI

Bianca Sofia Irene Fumagalli, Università degli Studi di Milano Statale
Emiliana Murgia, Università degli Studi di Genova
Elisabetta Nicchia, Università degli Studi di Genova
Davide Parmigiani, Università degli Studi di Genova
Andrea Garavaglia, Università degli Studi di Milano Statale

Keywords: Digital Citizenship Education, Ethical Awareness, Media Literacy, Information Literacy

We live in an online dimension, with no clear boundary between real and virtual, and where being conscious citizens means learning to navigate an increasingly complex digital reality, exploring its limits and potentials. In this context, learning to be a citizen is deeply connected to accessing online content mindfully. It also requires understanding the ethical implications of internet navigation. For this reason, within the educational landscape, it becomes crucial that students – tomorrow's citizens – learn to know how to take advantage of the opportunities while minimizing manipulation and other risks. In today's environment, where teachers seek training, it is necessary that educators first and students subsequently achieve this competences.

Information literacy emerges as a critical framework for empowering students' right to access, critically evaluate, and to use digital information, avoiding disinformation (Dmitrova, 2022). Possible training models can be grounded to facilitate teachers' professional development in digital citizenship education (Ranieri, 2021). Premising teacher training, a multifaceted approach can be conceived, encompassing diverse strategies beneficial for learners' development.

Digitalization has transformed how we learn. It is fundamental that educators know how to train young people to become aware citizens who understand their rights, are actively engaged, and are capable of responsibly addressing societal challenges. Children and teens should be viewed not merely as users but as informants, co-designers, and evaluators (Landoni al., 2024). Integrating digital citizenship education (DCE) into school curricula teaches students their rights and responsibilities, promoting responsible technology use and strategies to tackle disinformation.

Strategies include teaching critical thinking, online safety, and ethical behaviour to empower youth and minimize manipulation risks (Malik, 2024). Another approach involves engaging students through Real-World Tasks to enhance digital literacy and citizenship (Jeanneau & Ollivier, 2023). Dialogue and co-creation also assume foundational importance: Educators can encourage collaboration through technology use to minimize manipulations and promote more engaged and informed citizenship (Cazacu et al., 2020).

From an ethical approach, addressing the ethical implications of technology use is fundamental to promoting more socially responsible behaviour (Brueckner et al., 2018). On the side of Media Literacy, helping students identify misinformation enables them to improve critical thinking and make more conscious online choices (Fakih, 2022).

While these strategies focus on improvement and learning, it is essential to recognize that manipulation can persist. Continuous assessment of digital practices and promoting ethical standards are the paths leading to cultivating responsible digital citizens who can confidently navigate and critically engage with information landscapes.

12 JUNE 2025 14.00 - 17.00 **SESSION 2**

ID 808 - AI as normative assemblage

Armen Khatchatourov, Université Gustave-Eiffel

Keywords: AI, normativity, bias, episteme, HCI

To great extent, the premises of assemblage thinking can be found in Foucault's work, where plays out the idea of multiple layers of apparatus of power which are never monolithic or reducible to one-way domination. There is instead an imbrication of different normative mechanisms, related to sometime concurrent forms of governmentality.

This contribution will apply this framework to recent socio-technical assemblages. The proliferation of AI-based systems has led to new ways in which the normativity is entangled within technical systems. On the one hand, social normativity is translated into dynamic AI systems rather than a slow-evolving set of social rules. Here, the question is what is considered as error or bias, and with respect to which referent – and how deviation and alignment are addressed.

On the other hand, social normativity itself – not only in its content but in the very way we relate to it – is affected by the opacity and adjustability of machine-learning models, which nevertheless produce new expectations, behaviours and a form of algorithmic governmentality.

Hence, we will present a heuristic approach to disentangle the many ways normative concepts are at play in AI technologies, according to the following lines of questioning.

In classical AI, we can refer to "normal" as based either (1) on statistical observation of a given population or (2) on a moral or political stance related to the latter.

Shifting from a static vision of a pre-given norm, the norm itself (which we think we sociologically observe

or which we politically wish to come to reality) evolve under the effect of technological change. The users may have different expectations (3) with regard to a Human-Computer or to a mediated social interaction, depending on a particular technical artefact. This obviously has been one of the main topics in user studies since decades and plays out forcefully with current generative AI, as expectations and behaviors evolve rapidly.

More broadly, (4) the normativity itself is inscribed in larger processes related to governmentality at hand: the norms – and our relation to, the way we understand, submit to or confront them – do evolve both in their form and in their content. From disciplinary societies to algorithmic governmentality, the change from obedient/resisting subject (norm as a finite bounded domain) to the modulated subject (norm as perpetual adjustment), this evolution is intertwined with the advent of AI-based systems.

To address the 4 levels mentioned above, first approach would examine different conceptions of normativity, error and deviation, and how we can think what the "error", "bias" and "truth" are in AI settings.

The second approach would theorize the way we cope with the error and its inscription in temporality and teleology. Starting with Hegelian stance on the error as the "necessary moment of the truth" and the unavoidable "après-coup" (nachträglich) which derives from it, throughout the structural failure of "différance" (Derrida) and "le défaut qu'il faut" (Stiegler), we will try to reexamine how this "necessity" could contribute to the understanding of machine learning.