Check for updates

ESSAY

REVISED Us and artificial intelligence: Questioning agency and the need for critical and humanist perspectives

[version 2; peer review: 3 approved]

Marco Guglielmo[1], Petra Ahrweiler[2], Oscar Barberà[1], Simon Dario Brockschmidt[2], Nathan Critch[3], Ana Luisa de Moraes Azenha[4], Alejandro Fernández Del Río[1], Crystal A. Ennis[5], Pedro García Guijarro[1], Eva Gomís Jaen[1], Darcy Luke[3], Blanca Luque Capellas[2], Michal Malý[6], Masoumeh Iran Mansouri[7], Andrea Medrado[8], Ben O'Loughlin[9], D'arcy Ritchie[10], Mar Sánchez Montell[1], Giulia Sandri[11], Carla Sentí Navarro[1], Saori Shibata[12], Henry Snowball[9], Andreu Teruel[1], Bradley Ward[13]

[1]Constitutional Law, Political and Administrative Sciences, Universitat de Valencia, Valencia, Valencian Community, 46010, Spain
[2]Technology & Innovation Sociology / Social Simulation Laboratory (TISSS Lab), Johannes Gutenberg University Mainz Institute of Sociology, Mainz, Rhineland-Palatinate, 55128, Germany
[3]Department of Politics, The University of Manchester, Manchester, England, M13 9PL, UK
[4]Portulans Institute, Portulans Institute, Washington D.C., District of Columbia, 20005, USA
[5]Institute for Aread Studies, Leiden University Institute for Area Studies, Leiden, South Holland, 2300RA, The Netherlands
[6]Institute of Political Science, Charles University, Prague, 15800, Czech Republic
[7]University of Birmingham School of Computer Science, Birmingham, England, B152TT, UK
[8]Department of Communications, Drama and Film, University of Exeter Department of Film Studies, Exeter, England, EX4 4QJ, UK
[9]Department of Politics, International Relations and Philosophy, Royal Holloway University of London Department of Politics and International Relations, Egham, England, TW20 0EX, UK
[10]Department of Poltiical Science and International Studies, University of Birmingham Department of Political Science and International Studies, Birmingham, England, B15 2TT, UK
[11]Centre for the Study of Political Life, Université libre de Bruxelles, Ixelles, 1050,Ixelles, Belgium
[12]School of Languages, Arts and Societies, The University of Sheffield Faculty of Arts and Humanities, Sheffield, England, S3 7RA, UK
[13]Department of Politics and International Relations, University of Southampton Politics and International Relations, Southampton, England, SO17 1BJ, UK

v2 **First published:** 16 Oct 2025, **5**:318
https://doi.org/10.12688/openreseurope.21477.1
Latest published: 21 Aug 2026, **5**:318
https://doi.org/10.12688/openreseurope.21477.2

Abstract

This essay explores how Artificial Intelligence (AI) is reshaping political and social relations, arguing for the need to reclaim AI as 'ours' and indicating how critical and humanist research can contribute to this process. While AI technologies are increasingly integrated into public infrastructures, their development is mainly driven by market logics that prioritise efficiency, prediction, and optimisation—frequently at the expense of democratic participation, epistemic plurality, and environmental sustainability. The essay results from a two-day

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 2 (revision) 21 Aug 2026	✓ view	✓ view	✓ view
version 1 16 Oct 2025	✗ view	✓ view	

international workshop held at the University of Valencia, which brought together interdisciplinary researchers to engage in experiential, reflexive, and collaborative theory-building. Using world cafés, scenario-building, and AI-assisted role-play, participants collectively identified five key challenges of AI-mediated social life: the ideological framing of efficiency, exclusions within participatory processes, the epistemological authority of AI systems, the erasure of historicity, and the instability of predictive infrastructures. In response, we propose a minimal framework to reorient AI research toward more inclusive and pluralistic agendas. These include reclaiming predictive power, critically assessing participation and exclusion, and addressing the environmental and geopolitical dimensions of AI. We also reflect on the meaning of research impact in AI scholarship, advocating for a broader conception that embraces critique, friction, and reflexivity. Overall, the essay argues that reclaiming agency in the age of AI requires not only better tools or regulations but a fundamental rethinking of the social relations, epistemologies, and political imaginaries that shape technological development.

Keywords
Artificial Intelligence, Critical Theory, Political Agency, Feminism, Decolonial Theory, Political Participation

1. **Rebecca Scheffauer**, German Institute for Economic Research Berlin (Ringgold ID: 28354), Berlin, Germany
Freie Universitat Berlin Institut fur Soziologie (Ringgold ID: 54201), Berlin, Germany
2. **David Bailey**, University of Birmingham, Birmingham, UK
3. **Rosalie Waelen**, University of Bonn, Bonn, Germany

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the Horizon Europe gateway.

This article is included in the Marie-Sklodowska-Curie Actions (MSCA) gateway.

Corresponding author: Marco Guglielmo (marco2.guglielmo@uv.es)

Author roles: Guglielmo M: Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation; **Ahrweiler P**: Conceptualization, Methodology, Writing – Original Draft Preparation; **Barberà O**: Conceptualization, Funding Acquisition, Methodology, Writing – Original Draft Preparation; **Brockschmidt SD**: Conceptualization, Methodology; **Critch N**: Conceptualization, Writing – Original Draft Preparation; **de Moraes Azenha AL**: Conceptualization, Writing – Original Draft Preparation; **Del Río AF**: Conceptualization, Methodology; **Ennis CA**: Conceptualization, Supervision, Writing – Original Draft Preparation; **García Guijarro P**: Conceptualization, Supervision; **Gomís Jaen E**: Conceptualization, Methodology; **Luke D**: Conceptualization, Supervision; **Luque Capellas B**: Conceptualization, Writing – Original Draft Preparation; **Malý M**: Conceptualization, Methodology; **Mansouri MI**: Conceptualization, Methodology; **Medrado A**: Conceptualization, Supervision, Writing – Original Draft Preparation; **O'Loughlin B**: Conceptualization, Formal Analysis, Supervision, Writing – Original Draft Preparation; **Ritchie D**: Conceptualization, Investigation, Methodology; **Sánchez Montell M**: Conceptualization, Methodology; **Sandri G**: Conceptualization, Supervision, Writing – Review & Editing; **Sentí Navarro C**: Conceptualization, Methodology; **Shibata S**: Conceptualization, Investigation; **Snowball H**: Conceptualization, Investigation, Methodology; **Teruel A**: Data Curation, Methodology; **Ward B**: Conceptualization, Formal Analysis

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101110457.The international workshop from which conceptualisations for this essay emerged has been funded by Generalitat Valenciana, 'subvenciones para la organización y difusión de congresos, jornadas y reuniones científicas, tecnológicas, humanísticas o artísticas de carácter internacional', Grant: CIAORG/2024/015.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Guglielmo M *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Guglielmo M, Ahrweiler P, Barberà O *et al.* **Us and artificial intelligence: Questioning agency and the need for critical and humanist perspectives [version 2; peer review: 3 approved]** Open Research Europe 2026, **5**:318 https://doi.org/10.12688/openreseurope.21477.2

First published: 16 Oct 2025, **5**:318 https://doi.org/10.12688/openreseurope.21477.1

REVISED **Amendments from Version 1**

We have undertaken a systematic and substantive revision aimed at improving the manuscript's internal consistency, conceptual precision, and readability. In the 'Manuscript file', we highlighted the main areas of amendment.

In summary:

1. We have reorganised the structure of the paper, introducing a consistent numbering of sections and subsections and aligning the analytical development across challenges, impact, and research agendas.

2. We have strengthened the conceptual integration of the core argument, especially the links between efficiency, participation, knowledge production, historicity, and prediction.

3. We have clarified the positioning of the essay in relation to existing work on governance, regulation, design, and accountability, while making more explicit the relational focus of our contribution.

4. We have reworked sections that previously risked reading as lists, improving their theoretical depth while preserving the empirical and illustrative material.

5. We have added a dedicated limitations subsection and strengthened the conclusion by clarifying the normative grounding of the critical and humanist perspective advanced in the essay.

Any further responses from the reviewers can be found at the end of the article

Introduction

Artificial Intelligence (AI) is increasingly embedded in the infrastructures that shape our everyday lives. As AI technologies become central to decision-making in healthcare, education, labour markets, and public administration, among other areas, they carry with them a set of powerful expectations: greater speed, better outcomes, and more reliable forecasts. But these expectations are not simply technical; they are shaped by a particular set of values and assumptions about how society should function. As the critical theorist of technology Evgeny Morozov (2024) has brilliantly noted, they reflect a broader logic of Panglossian neoliberalism, a view promoted by tech entrepreneurs and platform builders that suggests 'we already live in the best of all possible worlds' and that 'there is no alternative to the market-driven provision of our tech infrastructures'. This essay starts from the premise that the Panglossian promise conceals the assumptions underlying the implementation of AI in social and political relations, ignoring the ways in which AI may exacerbate social exclusion, hasten war and democratic decline, among other issues.

AI, broadly conceived as 'any machine or algorithm that is capable of observing its environment, learning, and based on the knowledge and experience gained, taking intelligent action or proposing decisions' (Duberry 2022: 5), is becoming part of citizens' social and political relations. As allegedly intelligent machines reshape how institutions operate, how knowledge is validated, and how people interact, they also reconfigure agency: who is able to act, on what, and with what consequences (List 2021). Academic attention and public debates on AI have been burgeoning in the last ten years (see: Supriyanto & Saputra 2022; Ulnicane & Erkkilä 2023; Zuiderwijk *et al.* 2021). However, research efforts often remain caught between technical and institutional governance frameworks, leaving less space to reflect on the broader consequences for democratic life, the future of work, social inclusion, or the collective imagination of AI-ifying social and political relations.

Indeed, existing scholarship on AI and politics has largely developed along two broad lines. A first strand focuses on institutional regulation and governance, asking how democratic systems can contain the risks of AI through legal, ethical, and policy frameworks (Floridi *et al.* 2018; Djeffal *et al.* 2022; Ulnicane & Erkkilä 2023). A second strand focuses on the design, operation, and accountability of AI systems themselves, particularly in relation to bias, opacity, and fairness in algorithmic decision-making (Carabantes 2020; Coeckelbergh 2025; Micheli *et al.* 2020). By examining this recent literature on addressing AI biases through institutional regulation, data governance and software design ethics (i.e. Coeckelbergh 2024; Djeffal *et al.* 2022), we can identify key blind spots and seek fresh perspectives. If we affirm that people should have the right to understand and shape the technologies that increasingly govern their lives, then we must ask: who is included in this 'we'? Whose experiences, values and forms of knowledge count? And who is most at risk of being made invisible, replaceable, or excluded altogether?

These questions raise timely and conceptual issues. The hype surrounding Artificial General Intelligence (AGI) has helped reinforce the idea that AI development is on a one-way track, driven by inevitable technical progress. We question this narrative of temporal inevitability and telos, opening paths to explore the relationship between technological innovation and adaptation and resistance. As Morozov (2024) also argues, the inevitability narrative can serve to distract

from more grounded political questions: 'If AGI isn't just around the corner, the debate about the future of this technology becomes more open-ended', potentially allowing us to question 'the implications of leaving its development to market forces'. The urgency, therefore, lies in reclaiming a space for critique, imagination, and the articulation of alternatives that challenge the narrative of inevitability (Crawford 2021; McQuillan 2022). What remains overlooked is a more thorough examination of the social and political relations that underlie the integration of AI systems, the tensions they carry, and the possibilities they foreclose (Lindgren, 2024). This essay aims to prompt reflections to offer just that. It brings together the insights, provocations and shared questions that emerged from a two-day international workshop held at the University of Valencia in September 2025. The workshop brought together researchers and practitioners from different disciplines and countries[1] to reflect on AI as a set of processes entangled with politics, knowledge, and the reproduction of unjust social relations. Through a series of collective sessions, including experiential methods such as world cafés, role-plays, and scenario-building exercises, we questioned how AI uses and narratives shape our own positions, assumptions, and research practices.

This essay develops a minimal framework for critical and humanist perspectives on AI, politics and society. This framework unfolds from identifying key challenges in researching AI in the social sciences and humanities, problematising how research can be impactful, and, on this basis, which principles should inform alternative agendas and perspectives that tackle social exclusion and the risk of agency displacement to machines. These are the core elements of the essay's structure. After presenting the methodology used during the workshop to discuss these topics in Section 1, we propose five core challenges in Section 2. These are: (1) the dominance of efficiency as a framing device for AI systems; (2) the gap between promises of participation and the persistence of exclusion; (3) the emergence of AI as an epistemological actor; (4) the erasure of historicity and narrative complexity; and (5) the tension between predictive expansion and infrastructural fragility. Each of these challenges is presented not as a siloed theme, but as a way of rethinking what it means to act, to research, and to imagine in AI-mediated societies.

In Section 3, we propose concepts to pluralise and politicise research impact, what it means, who defines it, and how it should be enacted. Drawing from the workshop's interactive format, including the dialogue with AI-generated political personas, we explore the tensions between stakeholder-driven understandings of impact and critical, pluralist approaches that emphasise reflexivity, resistance, and ethical discomfort. In doing so, we propose a broader interpretation of *impactful* research, one that includes challenging dominant views and narratives.

In Section 4, we sketch three priorities for informing open-ended agendas on AI: reclaiming predictive power, rethinking participation and exclusion, and addressing the environmental and geopolitical asymmetries behind AI infrastructures. These are not definitive roadmaps, but minimal conditions for research committed to alternative presents and futures. Altogether, the essay develops the argument that critical and humanist perspectives on AI are not a luxury or an add-on, but a necessary ground for the same possibility of redirecting technological development toward the common good.

Section 5 presents the limitations of our discussion and this essay and offers concluding reflections on reclaiming human agency through critical and humanist perspectives, such as feminist and decolonial epistemologies, to challenge the a-democratic control of AI by techno-oligarchs. It emphasises the necessity of collective choice and pluralism in research and development to ensure technology serves the common good and protects human creativity from being stifled by rigid power structures.

1. Experiential theory development: Analytical positioning and methodological approach

The workshop sought to build on the existing approaches to AI outlined above and to recast the key analytical problem to be addressed: namely, how AI becomes embedded within evolving relations between research, politics, and society, and how these relations reshape agency, interpretation, and contestation in practice (Halsema 2003). To address the problem, we conceived an academic workshop as an immersive methodology in its own right. The aim was to foster epistemic reflexivity, enabling researchers to explore how our own positions, assumptions, and methods are entangled with the very phenomena we study (Tarr *et al.* 2018). Through a series of plenaries and lab sessions, participants collectively interrogated the role of AI in shaping social relations, challenging research practices, and enabling or constraining political agency (Cornish *et al.* 2023).

[1]Disciplines included: politics, political economy, constitutional law, gender studies, media and arts, international relations, sociology. Participants reflected gender balance and various cultural backgrounds. Their academic affiliations are in: Germany, Belgium, Spain, United Kingdom, the Netherlands, Czech Republic.

1.a From concept to praxis: Designing for generativity

The event was designed with an open-ended, iterative structure, inviting participants to actively shape the agenda. The initial concept note outlined three perspectives as entry points to examine how agency circulates between humans and machines:

1. the challenges we see as inherent in the proliferation of AI,
2. how we as researchers can productively impact the development of AI, and
3. the need to identify the most pressing agendas for future research.

This design provided enough scaffolding to orient discussion, while simultaneously leaving space for unanticipated concerns, ideas, and modes of engagement (Ball 2012). Importantly, the workshop was oriented toward experiential collaboration. The 25 participants from five countries and spanning disciplines such as political science, sociology, law, international relations, arts, gender, media studies and computer science, were invited to co-produce situated knowledge.

The organisational logic reflected an ethos of horizontal learning (Freire 1994). Early-career researchers were not only welcomed but also structurally supported, ensuring that power asymmetries were not reproduced within the workshop space: all those with more senior or higher-paid positions agreed not to receive participation awards, redirecting resources to support early-career and precarious researchers. While the participants reflected disciplinary, gender and geographical diversity, a limitation of the workshop is that all academics are based in institutions in the Global North. This limitation was discussed during the workshop, which formed the basis for designing further activities. On day one, participants entered discussions starting from their own experiences and aims and practised collaborative methods for theory development. On day two, we drew on these discussions to devise principles for fresh research agendas on AI, politics and society.

1.b Method as encounter: World-café and role-play

The opening plenary, 'Researching GenAI: Challenges', began with short interventions from participants, followed by a clustering of shared dilemmas. These ranged from concerns about algorithmic epistemologies and exclusionary infrastructures to questions of resistance (Miragoli 2025), regulation, and the political imaginaries that underpin AI adoption. Importantly, the plenary was designed to surface tensions and provide conceptual material for the collaborative lab to follow.

In the first laboratory session, participants engaged in a World-Café format structured around the challenges identified earlier (Löhr *et al.* 2020). The participants were randomly assigned to three tables with prompts derived from the plenary session ('efficiency, or what?', 'who can (not) participate?', and 'how do we know what AI knows?'), plus a fourth table with no predefined themes. Moving between discussion tables every twenty minutes, each anchored by a host who retained and synthesised earlier conversations, participants cycled through multiple perspectives. This approach added layers of complexity to the initial prompts with each round, enabling us to identify areas of disagreement and open debate.

The emphasis on experiential methods was further expanded in the second lab-based session, which followed a collective discussion on the theme 'Whose Impact and What For?'. The discussion addressed whether and how researchers can create a positive impact when interrogating AI, society, and politics. Here, we adopted a role-play method, situating ourselves as humans interacting with quasi-agentic AI chatbots (Bandi *et al.* 2025) that embodied distinct political actors. These included a fictionalised Prime Minister, a conservative editor-in-chief, a community organiser from the Amazon region combating deforestation, and a high-ranking EU bureaucrat. Also in this session, there was a fifth table where participants could design their own agent. Participants engaged in role-play, utilising the agent designed in ChatGPT 4-o to assume the role of researchers seeking to open dialogue with these figures and explore pathways for impactful research.

1.c Reflexivity as orientation

Throughout the workshop, reflexivity was a constitutive principle (Farrugia 2013): participants were not only asked to reflect on AI as a social and political phenomenon, but on their own positionalities, disciplinary assumptions, and methodological commitments. This reflexivity was at the core of the final laboratory session, which encouraged participants to devise shared agendas, thematic synergies, and overarching principles for the future of research on AI, politics and society. These principles informed the commitment to follow up with further networking activities, primarily focused on prompting humanist and critical perspectives on AI. This essay represents one such collective outcome.

Ultimately, the methodology adopted was a practice of epistemic care attuned to the risks of abstraction, hierarchy, and closure that often accompany AI discourse. The workshop sought to reclaim agency, not just analytically, but in how we organise our knowledge practices, build our research communities, and imagine our futures with (or without) AI. It is crucial not to replicate the same data-extractivist logics we critique in tech industries (Lehuedé 2024).

1.d Analytical positioning: A relational approach to AI, politics and society

In what follows, we articulate how these collective reflections about the challenges and impacts of research led to the development of key concepts for critical and theoretical approaches to AI, society, and politics. As stated above, this essay engages with a growing body of work on AI and politics that has primarily focused on questions of governance, regulation, and system design. These approaches tend to analyse AI as an object acting upon political and social systems. However, the perspective developed here departs from this by treating technological systems and social processes as analytically inseparable, examining instead how they are configured in practice across relations between research, politics, and society. In doing so, we seek to shift the analytical focus toward the configurations through which AI is enacted. It is from this perspective that the workshop explored in the previous subsections is approached, not as an isolated empirical setting, but as a site through which these dynamics become observable and open to conceptual development.

2. Us and AI: Challenges

Artificial Intelligence is often heralded as a technological breakthrough that promises to reshape society for the better, delivering faster services, more efficient governance and administration, and enhanced decision-making (Duberry 2022; Dunleavy & Margetts 2023). The promise of AI is not only that societal goals can be achieved faster and better: Techno-entrepreneurs and politicians emphasise that AI is also cost-effective, often obscuring how these alleged financial savings come with enormous costs for natural resources (Crawford 2021) and huge human exploitation (Muldoon *et al.* 2024). In this discourse, AI offers to improve innovation itself, since all possibilities can be generated and considered immediately, so 'we' can invent *more*, *better*. These framings are not out there by chance. The promise of the Panglossian world, always almost-there-to-come (see Introduction) is a purposeful ideological construction seeking to obscure deeper tensions around the role of AI in society (Gilbert & Williams 2022).

This section develops a set of challenges that emerged from a grounded, reflexive theory-development process conducted during the worksho. The starting point of this process was the articulation of problems by participants, understood as open diagnostic entry points that captured tensions, uncertainties, and contradictions in the current framing and mobilisation of AI in social and political contexts.

From these initial reflections, three core problematisations structured the first stage of discussion. First, participants highlighted the problem of conceptualising alternatives to AI as an efficiency-maker, particularly in light of increasing evidence of the limitations and contradictions of this promise (Bracci 2022; Henman 2020). Second, they identified a tension between the use of AI to enable or enhance participatory processes in contexts of democratic decline (Sieber *et al.* 2025) and the fundamentally undemocratic design features embedded in many AI systems (Chehoudi 2025). Third, participants pointed to an epistemological challenge, namely the framing of AI as the driver of a new data-driven paradigm of political and social relations (Coeckelbergh 2024), which sits uneasily alongside persistent patterns of exclusion and the marginalisation of diverse human perspectives (Noble 2018).

Through the iterative dynamics of rotation, discussion, and collective reflection, these problematisations were further specified and reorganised. In this process, two additional lines of tension became sufficiently distinct to be treated analytically: the politics of time and temporality in human–AI relations, including the tension between the transfer of predictive capacity to AI systems (Coeckelbergh 2021) and the reconfiguration of temporal control in interaction, and the structural fragility of AI infrastructures, including the possibility that their underlying premises and promises may generate forms of instability or collapse (Peterson 2025).

The outcome of this process is the identification of five analytically distinct but interrelated challenges. In this essay, challenges are understood as specific points of tension in human–AI relations that put pressure on how social and political processes are usually observed, conceptualised, and evaluated. Each challenge isolates a particular relational dynamic and shows how it reshapes interaction while exposing the limits of existing concepts and methods. Taken together, these challenges provide a structured way to analyse how AI reconfigures discourse, participation, temporality, knowledge production, and the conditions under which more just social relations can be pursued. The remainder of this section develops each of these challenges in turn.

2.a The dominance of efficiency as a framing device for AI systems

The first core challenge concerns the powerful imaginary of AI as a driver of efficiency, a trope that pervades public discourse, government narratives (Fourcade 2016), and corporate promises (Amoore *et al.* 2024; Hirsch-Kreinsen 2024). Whether in national economies, policymaking, service delivery, research practice, or even deliberative democratic processes, AI is increasingly framed as a solution to longstanding performance gaps (Zuiderwijk *et al.* 2021). However, as the workshop discussions made clear, the assumed link between AI and improved efficiency is highly contingent and ideologically charged.

This becomes particularly evident in the use of AI systems to produce 'efficiency' in work which would ordinarily rely on experience, nuance and expertise. For example, the use of AI-driven facial recognition systems in warfare and their integration into drone operations is presented as enabling precision and efficiency in identifying and targeting individuals as well as reducing collateral damage and operational costs (De Swarte *et al.* 2019; Sauer and Schörnig 2012). However, these systems do not simply assist in identifying targets; they structure and define who or what is a legitimate target, shape how the threat itself is conceptualised, embed particular assumptions about risk, behaviour, and identification, and measure success against criteria generated within the system itself. A similar configuration can be observed in the use of AI systems to manage citizens' access to public health services where AI systems determine which data to collect and which criteria define urgency or eligibility (Khosravi *et al.* 2024). Again, AI reshapes what is observable within the system, reconceptualises legitimate needs and reevaluates appropriate outcomes, evaluating the efficiency of this new system on its own metrics.

Across these domains, participants highlighted how the efficiency paradigm privileges what can be tracked, standardised, and scaled, while marginalising dimensions of social life that resist quantification. Creativity, diversity, and contextual judgement are often treated as inefficiencies due to their non-quantifiable character (Edwards 2001; Moore 2017). At the same time, the empirical evidence on productivity gains remains uneven across levels, from individual cognitive labour under pressure (Vercellone 2014), to organisational performance (Bankins *et al.* 2024), and state capacity (Neumann *et al.* 2024). This suggests that the promise of efficiency operates as a powerful organising narrative even where its outcomes are uncertain.

This dynamic is inseparable from the underlying data infrastructures (Srnicek 2026). The pursuit of efficiency in AI systems is structurally tied to an expansive logic of data extraction, an implicit imperative to collect and process ever-larger volumes of information to optimise outcomes. Yet this comes at a high and often hidden cost. At the individual level, it entails the erosion of control over personal data, raising concerns about privacy (Manheim and Kaplan 2019). At the collective level, it raises fundamental questions about the governance of data as a public or common resource, including who owns, accesses, and controls the infrastructures through which social life is rendered legible. For this reason, legislative interventions, while necessary, are not sufficient. Regulation can impose constraints on data use and algorithmic accountability (Smuha 2021), but it does not fundamentally alter the infrastructures and epistemic assumptions through which AI systems define what is observed, how it is conceptualised, and how it is evaluated (Amoore 2024). The challenge of efficiency therefore lies not only in assessing whether AI delivers better outcomes, but in interrogating the conditions under which efficiency itself is defined and enacted.

The deconstruction of efficiency was a ground for envisioning alternative imaginaries, rooted not in making everything faster or cheaper, but in transforming the goals of public and private systems. Creativity, lived experiences, ecological responsibility, and collective well-being, none of which lend themselves to quantification, are the marginalised principles that can form the base of alternative human-to-technology relations. Transformative visions of AI prioritise happiness, relationality, and planetary care over sheer optimisation. While it is not the primary function of the research community to set alternative agendas, researchers committed to alternative social relations can play a crucial humanist and critical role in unmasking the deleterious effects of quantified efficiency and providing ground for alternative uses of human-centred technologies.

2.b The gap between promises of participation and the persistence of exclusion

While the first challenge has to do with the concepts and modes of operation of AI, and led to a discussion on the implications of campaigning for AI as an efficiency tool, the second challenge arises from the widespread assumption that AI can enhance participation, be it in democratic governance, public service co-design, or civic engagement (König & Wenzelburger 2020). However, as the workshop discussions made clear, this promise remains conceptually under-specified. Participation, in this context, cannot be reduced to mere inclusion, interaction, or data contribution. Rather, it concerns the capacity of individuals and groups to shape how processes are constituted, including what is made visible, how issues are defined, and how outcomes are assessed: that is, agency (Berenskoetter 2020).

From this perspective, the starting point is to ask: participation in what, and under which conditions? As data-hungry systems, AI requires all citizens to participate in feeding the machine (Muldoon *et al.* 2024), be it by conceding personal information or by contributing to processes such as image and language recognition. In the white West, this is framed as an equal exchange for receiving 'free' digital services and spaces of socialisation (Jordan 2015). Yet this apparent reciprocity obscures the fact that participation is already structured at the level of observation. AI systems determine which forms of input become legible as data, thereby shaping whose contributions are recognised and whose remain invisible.

At the same time, the infrastructures that sustain AI rely on forms of labour that are unevenly distributed and often hidden. The construction of databases and the operation of digital platforms are both displaced onto and sourced from the exploited labour of women, migrants, and digital workers in the 'Global South' (Fuchs 2018; Altenried 2020, 2022). Participation, in this sense, is embedded in asymmetrical relations of extraction. This premise is critically damaging to the very possibility of alternative AI systems. Forced data input is antithetical to AI co-creation, as it reduces participation. The extractive logic of AI results in the accumulation of data that is monetised by capitalists, extracted and utilised in less-than-transparent ways, with weak protections for data and privacy for the sources of this extracted knowledge (see Jacobsen 2025; Schaupp 2022; van Doorn & Badger 2020). The issue of data provenance—whose information powers these models and under what conditions—is one crucial omission in conversations about AI decision-making. The majority of large language models, such as those from OpenAI or Meta, are trained on enormous datasets that are scraped from the public web. One such dataset is Common Crawl, which contains billions of pages in dozens of languages, not just English, but with a disproportionate bias toward it (Vargas-Parada, 2025). Personal writings, forum posts, books, and code from innumerable people and creators are frequently included in this data, which is collected without their express permission or payment, posing difficult questions of intellectual property, privacy, and cultural bias to a condition of system functioning rather than a space of agency.

Acknowledging these dynamics requires recognising that participation is structured not only at the level of data extraction but also at the level of conceptualisation. Participation is often framed as input into pre-existing systems, rather than as the capacity to shape the terms under which these systems operate. As such, AI systems risk encoding dominant interests and reproducing social inequalities through their datasets and infrastructures (Fuchs 2019; Guglielmo *et al.* 2026; Mühlhoff 2025). As one participant warned, citizens' agency in AI governance is frequently assumed rather than materially supported. Structural exclusion by class, gender, race, age, access to education and health, geography, and disability remains pervasive, shaping who can participate and how.

These issues become particularly visible when AI is deployed within existing participatory processes. Rather than resolving problems of exclusion, AI may shift the terms of engagement from human deliberation to machine filtering and optimisation. This redefines how participation is evaluated. If AI is used to 'streamline' citizen input or 'enhance' participatory budgeting, who decides which inputs matter? Typically, this falls to platform developers, government technicians, or contracted AI vendors whose algorithms, trained on biased datasets, embed predefined priorities such as keyword frequency or sentiment scores, often sidelining unconventional voices. Furthermore, who interprets dissent? The same opaque systems, mediated by human overseers with limited accountability, may dismiss contrarian views as noise rather than legitimate critique. Finally, who gets filtered out? Marginalised groups whose language, cultural references, or digital literacy fall outside the model's scope, such as migrants, indigenous communities, or rural participants, risk systematic exclusion. In this way, evaluation becomes internal to the system, reinforcing rather than challenging existing inequalities.

To address these questions, on the one hand, there are positions that envision experimentation in more humanistic AI models. For example, AI agents have been proposed as 'rights mediators' assisting citizens in navigating complex bureaucratic systems (Floridi 2020). On the other hand, more radical perspectives argue that participation must be understood in its most material and embodied sense, including forms of resistance to AI itself, particularly in contexts such as workplace governance where digitalisation intersects with gendered and racialised surveillance (Woodcock 2021).

It is within this tension that participatory AI must be situated. Participatory AI that adopts a critical and emancipatory stance is not entirely out of reach (Birhane *et al.* 2022), but it operates under significant constraints. In contexts where banning AI is not feasible, alternative approaches to inclusive technology co-design have been explored. The international research project AI FORA (Ahrweiler 2025), conducted by some of the participants was taken as an example of a case which examined whether participatory modelling in multi-stakeholder settings of the 'Quadruple Helix' (McAdam & Debackere 2018), including vulnerable groups, can support the context-sensitive design of 'Better AI' in welfare systems. However, such approaches must confront the limitations of existing infrastructures.

Current algorithmic tools and administrative data used in welfare assessments often fail to reflect cultural values and discourse, instead reproducing unintended political biases and reinforcing structural inequalities (Mitchell *et al.* 2021; Koebis *et al.* 2022; Pagano *et al.* 2023). Without careful co-design and continuous evaluation, AI risks institutionalising bias rather than correcting it. In AI FORA, a modelling environment was developed to capture the complexity of welfare systems, reveal hidden decision-making dynamics, and test alternative policies and AI rule sets (Ahrweiler *et al.* 2024). Stakeholders, including vulnerable groups, were able to experiment with new rules and explore their systemic consequences prior to implementation. This process allowed for the generation of alternative system configurations, including the production of new training data through simulation. Participatory approaches also included frontline workers and policymakers, enabling the translation of modelling results into policy and practice (Ahrweiler & Gilbert forthcoming). While these findings demonstrate the potential of participatory approaches to produce more legitimate and context-sensitive AI systems, they also underline that such outcomes depend on sustained efforts to reconfigure the underlying conditions of participation.

The critical question is therefore not whether AI can enable participation, but whose participation is enabled, under what terms, and at what cost. Democratising AI requires more than adding citizen inputs to pre-trained models. It requires rethinking the architectures through which participation is organised, including how data are collected, how contributions are interpreted, and how outcomes are evaluated. Crucially, the generative reflections from participants in the workshops became sites of mutual learning. Citizen-centred AI perspectives can provide examples of resistance through platforms as well as resistance-based approaches, while feminist perspectives in labour remain essential to avoid overlooking the areas of substantial exclusion that persist even within formally open participatory processes.

2.c The emergence of AI as an epistemological actor

A third challenge centres on AI as a producer of knowledge, and the implications this holds for epistemic authority (List 2021). In the workshop, participants probed the consequences of AI systems not only processing data but also generating representations of the world, which increasingly function as accepted truths in the forging of meaningful interpretations of events and in decision-making processes (Carabantes 2020). Understanding AI as epistemological machines that shape, if not chain, how we know and interpret phenomena is the necessary starting point for critical and humanist perspectives. The challenge, more precisely, lies in how AI systems intervene in the conditions under which knowledge is produced, validated, and circulated.

To begin with, the dominant notion of knowledge in AI-driven societies must be deconstructed. Knowledge is not a neutral accumulation of quantifiable data to be assembled in fact representations (D'Ignazio & Klein 2020). Rather, knowledge is itself a social process, embedded in power relations, partial perspectives, and normative assumptions. In contrast, AI-epistemological machines tend to reify data, presenting probabilistic outputs as objective insights (Micheli *et al.* 2020). This has direct implications for how social realities are observed, conceptualised, and evaluated: what is observable is constrained by available datasets, what is conceptualised is shaped by model architectures and correlations, and what is evaluated as valid knowledge is increasingly tied to statistical plausibility rather than social justification. The evidence of the perils of such assumptions is clear when AI-biased knowledge production is used in contexts such as criminal justice and welfare distribution, with dramatic effects on underrepresented and marginalised communities (Gillespie 2024; Phan and Wark 2021).

AI platforms, especially generative systems, can draft biographies, policy suggestions, and even legislative drafts, but are often devoid of accountability, reflexivity, and contextual understanding. A data-driven 'picture of reality' may reflect statistical correlations but lacks the experiential depth or political contestation through which democratic knowledge is constructed. These are processes of knowledge (de-)construction, exchange, and critique that cannot exclude human reflexivity and agency. In this sense, epistemic authority risks shifting from socially embedded actors to systems whose internal logics remain largely opaque.

Moreover, a bias-loop problem raises the issue of the potential enclosure of contestation. AI systems are trained on data which normalises existing forms of exclusion and domination and which do not contrast with platform companies' protocols (Danaher 2016; Barry and Stephenson 2025). Without processes of re-structuration, and the inclusion-by-design of marginalised voices and experiences, AI risks enclosing humans in a-spatial loops of knowledge reproduction. Further, as AI infrastructure is heavily North-West centric, the promise of data-driven politics for the common good is hampered by an inherent re-colonialising tendency (Hepp *et al.* 2022). As one participant put it, we risk entering a regime in which 'AI creates knowledge, AI judges knowledge, and AI reproduces knowledge', a closed epistemic loop with limited space for human dissent or pluralistic validation.

These scenarios raise critical questions: what counts as legitimate knowledge in AI societies? How do we retain the capacity to interrogate what AI tells us? And how do we resist the deskilling of human judgment, especially when productivity pressures incentivise deference to machine output?

The core concept driving critical perspectives is, therefore, the constant reclaiming of humanism, not as a nostalgic return to pre-digital modes of knowing, but as an insistence on situatedness, plurality, and narrative agency (bell hooks 2014). In this context, humanism is invoked as a political and epistemic commitment to recognising marginalised voices, lived experience, and the relational conditions through which knowledge is produced and validated. The challenge is not only to critique what AI knows, but to reassert how and why we know what AI knows.

Reflecting the challenge of participation, the nucleus of humanism shapes a debate with different positions that need to construct bridges of epistemic solidarity. On the one hand, there is the idea that alternative uses of existing AI systems can improve the process of knowledge production through constant and coordinated injection of situated, non-dominant data, discourse, and perspectives. On the other hand, there is the position that claims the right to disconnect from the machine (Kuntsman & Miyake 2022) and to reassert the primacy of humans in co-creating knowledge from and about social processes.

2.d The erasure of historicity and narrative complexity

In relation to the problem of knowledge, another challenge is identifying what is lost when AI becomes central to how we imagine social issues and solutions. While being heavily territorialised in the so-called Global North and imposing the fast pace of quantified productivity, AI systems are marked by an absence of historicity and an incapacity to grasp or represent temporal depth, political contestation, or narrative complexity (Coeckelbergh 2022). More precisely, the challenge lies in how AI systems reshape the conditions under which social realities are rendered intelligible, by privileging forms of knowledge that can be processed as decontextualised data while marginalising those that depend on historical situatedness and narrative construction.

Several examples illustrated this point. Generative AI systems can assist in drafting policy or proposing interventions, but they are poorly equipped to understand how problems evolve over time, how they are shaped by struggles, or how they reflect unresolved injustices. One participant captured this vividly: ‘GenAI is not a Deus ex machina, yet most people expect it to be’. This expectation not only misrepresents what AI can do; it also reshapes how problems are approached in practice. By encouraging shortcut solutions and privileging outputs over processes, AI systems contribute to flattening the cognitive labour of critical diagnosis (Berry 2014), displacing the need to reconstruct causal trajectories, historical conflicts, and institutional legacies.

This has direct implications for how social issues are observed, conceptualised, and evaluated. At the level of observation, AI systems prioritise present-oriented, high-volume data streams, often detached from the historical conditions that produced them (Lindgren 2024). At the level of conceptualisation, problems are reframed as patterns to be identified and optimised, rather than as historically embedded processes shaped by power and contestation. At the level of evaluation, success is measured by immediate outputs or predictive accuracy, rather than by longer-term trajectories, collective memory, identity or unresolved injustices and political struggle. In this sense, the absence of historicity is not a limitation external to AI systems, but a structural feature of how they organise knowledge.

The epistemic flattening produced by GenAI echoes broader cultural patterns of amnesia, acceleration, and depoliticisation (Burnham 2017). When AI becomes the primary source of explanation or interpretation, we risk displacing the slow, contested work of political narration, where people make sense of their lives, communities, and aspirations. This dynamic also interacts with other challenges identified in this section. The emphasis on efficiency privileges speed and optimisation over historical understanding, while the reconfiguration of participation limits whose narratives are included and how they are recognised within AI-mediated processes. In both cases, the erosion of historicity amplifies existing asymmetries, as certain experiences and trajectories are systematically excluded from collective sense-making.

In this sense, historicity and narrative literacy are not technical add-ons, but core conditions for democratic agency. The challenge is therefore not only to recognise what AI systems fail to capture, but to interrogate how their widespread adoption reshapes the epistemic foundations of social and political life.

2.e The tension between predictive expansion and infrastructural fragility

The fifth challenge interrogates a central assumption behind the current AI boom: that its infrastructures are robust, expanding, and inevitable, primarily as they take over predictive power (Mühlhoff 2025). More precisely, the challenge lies in the growing dependence on AI systems to anticipate and organise the future, while the material and economic

infrastructures that sustain this predictive capacity remain fragile, uneven, and potentially unsustainable. If humans delegate prediction to machines, they are not simply outsourcing a function; they are binding social and political processes to infrastructures that require continuous expansion in compute, data, and energy.

Participants urged a more critical stance, recognising that AI expansion is neither inevitable nor guaranteed. AI systems are often presented as stable and scalable, yet they are prone to crises, bottlenecks, and systemic vulnerabilities. In this sense, we may be conceding predictive authority to technological systems whose own conditions of existence are precarious. This tension becomes particularly visible when examining how AI reshapes the conditions under which social realities are observed, conceptualised, and evaluated. Prediction relies on continuous streams of data and computational capacity, privileging forms of observation that are real-time, high-volume, and forward-oriented (Alvarado 2023). At the level of conceptualisation, social processes are increasingly framed as patterns to be forecasted and optimised, rather than as open-ended and contestable trajectories. At the level of evaluation, success becomes tied to predictive accuracy and performance, reinforcing a narrow understanding of what counts as valid knowledge about the future.

This tension and paradox are well explained by the environmental costs of AI development. The extractive attribute of AI is well known: energy consumption, material extraction, and labour exploitation, often outsourced to the Global South, has been studied at large (boyd & Crawford 2012; Pirina 2022). However, this is often presented in political and public discourse as a negative externality rarely factored into policy discussions while being integral to the global political economy of AI (Dauvergne 2022). In this respect, the predictive paradigm reinforces the efficiency challenge by obscuring the material costs of optimisation and intersects with the participation challenge by externalising these costs onto populations that are structurally excluded from decision-making.

The call to expose what is unsustainable in current AI developments raises the provocative question: what if the business models sustaining AI development are unsustainable? What if the current trajectory mirrors the dot.com bubble, driven by hype, inflated valuations, and faith in endless expansion (Floridi 2024)? While these evolutions are overdetermined, model collapse cannot be excluded, especially in social research on AI. There is the possibility that AI systems will eventually consume their own outputs, degrading accuracy and eroding creativity. This 'cannibalisation' effect, where models recycle their own data until generative saturation, could result in a profound epistemic and economic breakdown. This dynamic can be read in light of broader critiques of capitalism as a 'cannibalising' system that consumes its own social and ecological conditions of possibility (Fraser 2023). While Fraser does not refer specifically to AI, this lens helps illuminate how AI infrastructures may reproduce similar extractive and self-undermining dynamics.

As with previous challenges, the generative discussion led participants to claim the need for community-based, localised alternatives to the dominant AI paradigm, technologies that serve real needs rather than speculative futures. However, the viability of such alternatives depends on confronting the structural dominance of predictive reason. Challenging AI is therefore not only a matter of improving models or mitigating externalities, but of reclaiming the possibility of rupture, that is, the capacity to imagine and enact futures that are not pre-defined by existing data, infrastructures, or optimisation logics.

3. AI, politics and society: Can we make an impact?

Building on the five challenges identified above, this section turns to the question of intervention through the lens of research impact. If AI systems increasingly mediate how knowledge is produced, circulated, and translated into socio-political processes, then the conditions under which research becomes 'impactful' are themselves transformed. What counts as relevant knowledge, whose perspectives are prioritised, and how problems are framed for decision-making are key processes of both embedding in and resisting unequal infrastructural and epistemic arrangements.

Rather than addressing these dynamics at a purely theoretical level, the workshop operationalised them through structured simulation with AI-generated models. Participants were placed in situations approximating real-world interactions in which political actors seek 'impactful' insights from researchers, allowing us to observe how generative AI intervenes in the production, mediation, and evaluation of politically relevant knowledge. The simulation exercise was relevant for reframing and politicising impact as a spectrum of possibilities, rather than a quantified measurement of research contribution on social and organisational indicators (Wigginton and Lafrance 2019).

We began with a collective reflection on the relational infrastructures that researchers are expected to use to generate impact. Institutions, ranging from public administrations to media regulators and civil society organisations, often articulate specific expectations about what research can or should deliver: evidence-based policy recommendations, accurate datasets to address societal challenges, or metrics to evaluate risks and opportunities of emerging technologies (Bergold & Thomas 2012). From this perspective, impact is defined mainly by the demands of stakeholders as interventions at specific points, for instance, the delivery or assessment of a policy cycle (Luchetta 2012). Within

AI-mediated contexts, such definitions risk reinforcing filtered participation, as systems privilege standardised inputs and institutional logics while marginalising alternative forms of knowledge and expression.

Building on this reflection, some participants proposed ways in which research on AI and the co-creation of alternative AI can enhance the role of academics while boosting liberal democratic participation. One example is that of citizen-centred AI tools that enhance rights navigation across EU platforms, to participatory consultations such as the Democracy Shield. However, such interventions often rely on data infrastructures that pre-structure whose voices are included, thereby reproducing asymmetries in participation under the guise of inclusivity (see Section 2.b). This, therefore, complicates the value of developing such tools to respond to real and pressing needs for inclusivity, explainability, and regulatory accountability (Novelli & Sandri 2025).

This tension led to a broader critical discussion on the need to re-centre humanism and pluralism within the very conception of impactful research. Instead of viewing research as a transactional exchange that delivers answers to predefined institutional questions, the session foregrounded a mode of inquiry in which scholars act as facilitators of reflexivity, contestation, and democratic imagination. This reflects broader epistemic struggles around defining legitimate knowledge as we see AI systems mediating the content and validation of politically relevant claims (see Section 2.c). Such a shift involves a spectrum of researchers' positionalities in search of areas for dialogue: on the one hand, researchers can contribute to more inclusive policy-making processes, for example, by co-designing participatory mechanisms that include the voices and experiences typically excluded from techno-legal deliberations. On the other hand, there is a need to create space for research that adopts an adversarial or critical posture towards institutions and organisations, questioning whether existing institutional logics are compatible with democratic goals in the age of AI (Mansouri & Bailey 2025).

Indeed, a recurrent theme in the conversation was that critique is not antithetical to impact but a necessary yet dismissed part of the relation between academia, organisations and society. While often tolerated as a niche or dissident perspective, critical research is rarely recognised as impactful. Yet, in fields like AI, where epistemic, normative and infrastructural transformations unfold rapidly and asymmetrically, critique is indispensable (Lindgren 2023). It makes visible the blind spots of dominant approaches, resists the flattening of alternatives, and insists on the ethical-political stakes of technological design and deployment. Without such friction, AI-mediated environments risk narrowing the horizon of possible futures through predictive framings that privilege continuity over rupture (see Section 2.e). The challenge, then, is not to choose between these two understandings of impact, improved accountability versus critique, but to articulate their tensions to improve their reciprocal spaces of possibilities in moving social relations towards the common good.

To explore these tensions the laboratory session invited participants to engage in a simulation exercise: a role-play with agentic AI chatbots, designed with GPT-4o, allegedly simulating different political and institutional actors. Participants were organised into small groups and invited to conduct a dialogue with these synthetic interlocutors as if they were real-life counterparts, prime ministers, editors-in-chief, EU cabinet chiefs, or community organisers (see above). The objective was not to assess the technical performance of the chatbots, but to use them as a medium for reflecting on what it means to do impactful research with, through, or against AI systems.

The simulations were conducted using large language model-based agents configured with distinct ideological orientations, rhetorical styles, and strategic priorities. Concretely, the simulation was built using the GPT-4o model as the underlying architecture. Distinct 'agents' were not separate trained systems, but instances of the same model differentiated through structured prompt design that specified roles (e.g., policymakers, activists), ideological orientations, and communicative styles. These prompts conditioned each agent's responses to inputs, allowing the same model to simulate contrasting positions within a controlled interaction environment. Through prompt design and iterative interaction, these agents generated policy arguments, responded to dissent, and proposed courses of action under conditions of time pressure and informational asymmetry. This enabled participants to engage with AI as an active interlocutor shaping the framing, sequencing, and perceived legitimacy of political discourse (Guglielmo *et al.* 2026).

The participation was mediated through the AI models' interpretative logic, which privileges certain discursive forms while marginalising others (see Section 2.b). Therefore, this format offered a unique vantage point to examine how GenAI systems perform political speech, reproduce normative assumptions, or generate affective responses. In practice, these effects emerge from the model's conditioning on large-scale training data combined with prompt instructions, which shape how outputs are generated, including the selection of arguments, framing of issues, and modulation of tone in response to user inputs.

In some instances, the exchanges produced moments of surprise: participants found themselves persuaded, provoked, or disoriented by the chatbot's replies. These effects point to the capacity of AI systems to structure affective and cognitive

responses, shaping not only what is said but how it is received and interpreted within interaction (Alvarado 2023). These moments offered critical insight into the simulation of political agency by GenAI models, how such systems emulate rhetorical strategies, project ideological stances, and manage dissent. At the same time, the exercise exposed the limits of such interactions especially where the GenAI models simply reproduced dominant epistemic patterns and reinforced existing knowledge hierarchies (see Section 2.c). For example, in cases where participants engaged the chatbot posing as a head of government, responses often remained superficial or evasive, failing to sustain substantive discussion on regulatory frameworks or democratic safeguards. This limitation reflects a broader tendency toward narrative flattening, where complex and historically embedded issues are reduced to generic or superficial outputs (see Section 2.d). Importantly, the activity did not aim to reach definitive conclusions on the feasibility of AI-simulated political engagement. Instead, it served as a laboratory for reflexive discomfort, a space where participants could experience the contradictions of AI research impact in real time.

In particular, the simulations reveal how AI systems operate as epistemological actors, shaping what counts as valid knowledge and how it circulates within political interactions. One particularly illustrative moment emerged when a group of participants, engaging with a chatbot representing a grassroots environmental activist, questioned the ecological footprint of GenAI itself. The agent expressed a form of systemic pessimism, while its very response was being generated through energy-intensive computation highlighting contradictions between the content of generated knowledge and the material conditions of its production (see Section 2.c). The irony was not lost on the participants, who chose to end the conversation early as an ethical gesture. Such moments underscore the importance of reflecting on the conditions of experimentation when designing interactive research settings integrating AI systems.

Another table, composed of feminist scholars, engaged the chatbot playing the role of an EU foreign policy official. Their questioning focused on the integration of gender perspectives into security strategies—a line of critique often marginalised in traditional policy impact frameworks. While the responses from the chatbot were measured and respectful, they nonetheless surfaced a latent tension: AI systems, even when adopting politically progressive scripts, can re-inscribe hierarchies of knowledge and authority. For instance, gendered critiques were acknowledged but reframed within pre-existing policy logics, limiting their transformative scope and re-positioning them as supplementary. In doing so, AI systems filter participation by privileging dominant policy discourses while rendering critical or situated perspectives less actionable within institutional frameworks (see Section 2.b). In this sense, the activity became a site for rehearsing dissent, where participants could test the elasticity of synthetic discourse and observe how critique is often absorbed into pre-structured conversational patterns that limit deeper contestation (see Section 2.d).

Taken together, these observations suggest that generative AI actively structures the conditions under which knowledge becomes actionable and impactful. The simulation-based approach makes visible how epistemic authority, participation, narrative construction, efficiency, and prediction are co-produced within AI-mediated interactions.

In sum, the session offered a multi-layered exploration of impact, demonstrating the need for a pluralistic understanding of ‘making a difference’, connecting the challenges identified above to AI-mediated interactions. The session created a space of dialectics and agonism between the more pragmatic and policy-driven demands of institutional actors who require academics to provide impactful research, and the epistemic and ethical responsibilities of critical scholars who navigate institutional constraints and agency to invent new concepts and metrics for what counts as impact. It used the affordances and constraints of GenAI not to propose ready-made solutions, but to stage questions, surface contradictions, and provoke reflection. In doing so, it affirmed that impact is not merely a matter of external validation or stakeholder alignment, but a process in need of critique and reflexivity.

4. Open-ended critical and humanist agendas on AI and social relations

The critical reflections on the challenges identified in Section 2, and their exploration through the simulation-based reflections on impact presented in Section 3, allow us to move from diagnosis to inference: we derive three priorities from the ways in which the nexus between AI, research, and socio-political processes becomes visible when examined through the tensions between challenges and impact. Each priority corresponds to a specific configuration of the challenges discussed above within the politics of research impact. The dominance of predictive logics (Section 2.e) and efficiency-driven reasoning (Section 2.a) shapes what kinds of knowledge are considered actionable and future-oriented, calling for a reclaiming of predictive power. The filtering of participation and reproduction of epistemic hierarchies (Sections 2.b and 2.c) determine whose knowledge is recognised and mobilised within impact-oriented processes, requiring a rethinking of inclusion beyond procedural approaches. Finally, the material and geopolitical conditions of AI infrastructures (Section 2.e, and implicitly 2.a) structure where and for whom impact is produced, foregrounding the need to address global asymmetries and environmental constraints.

4.a Reclaiming and challenging predictive power

To begin with, there is a need for agendas that reclaim and redirect predictive power (Nowotny 2021). This priority brings together the challenges related to efficiency-driven reasoning (Section 2.a) and predictive expansion (Section 2.e), as they shape what forms of knowledge are recognised as actionable within impact-oriented research. In AI-mediated environments, predictive logics structure the conditions under which certain futures become thinkable, fundable, and politically relevant. To delve deeper into this priority, we started by reflecting on how AI is already reshaping the knowledge infrastructure of academia itself, from how we collect data to how we teach, publish, and disseminate research (Grossmann *et al.* 2023). The question 'what will conducting research in 2035 look like?' served as an anchor for scenario-building exercises aimed at unpacking the institutional, epistemic and technological shifts currently underway. These shifts reflect the consolidation of predictive infrastructures within knowledge production, reinforcing a forward-looking logic aligned with optimisation and control (see Sections 2.a and 2.e). The purpose of such future work is not to offer deterministic forecasts but to clarify the conditions under which particular trajectories may become dominant, and to imagine interventions that might disrupt or redirect them (van Witteloostuijn *et al.* 2022).

Defining scenarios and the conditions for their unfolding is one way to advance critical reflection on alternatives and perspectives that remain underexplored. Drawing from participant interests, we took the question of 'what is AI doing to labour relations?'. We are becoming accustomed to the use of AI in recruitment, human resources management, and surveillance processes (Cant 2020). In this context, prediction becomes a mechanism of governance, shaping labour relations through surveillance, classification, and anticipatory control (see Section 2.e). This happens in the background of discourse and imaginaries, promising work to become more innovative and more productive (Al Naqbi *et al.* 2024). What are the possible scenarios unfolding from the dialectics between the promise of productivity and the practices of authoritarian surveillance? And is there any space for alternative uses of AI models in favour of labourers' creativity? These questions illuminate the potentiality of shifting the perspective of social inquiry. As proposed throughout this essay, to take feminist and decolonial epistemologies (Stingl 2015) seriously means shifting how we look at phenomena to shed light on what practices of resistance to and through algorithms (Bonini & Trere 2024) have been emerging and forging solidarity bonds that can inform the background against which implementations of AI are assessed (D'Ignazio & Klein 2020).

4.b Rolling back exclusion and re-politicising participation

The second priority for novel research agendas concerns the challenge of who participates with AI and to what end? Today's AI development is still expert-driven, "characterised currently as technically-focused, representationally imbalanced, and non-participatory" (Birhane *et al.* 2022: 1). Though Birhane *et al.* report recent participatory methods in the lifecycle of machine learning, including auditing, public consultation such as citizen juries, joint problem formulation, model cards and datasheets for information disclosures, and even artefact co-development, they voice a suspicion that these methods might merely serve to aid in the refinement of systems from the viewpoint of industry. This priority derives from the challenges of participation and epistemic authority (Sections 2.b and 2.c), as they intersect in defining whose knowledge is recognised, mobilised, and legitimised within AI-mediated processes of research impact. The question is not only who participates, but under what epistemic conditions participation becomes meaningful or is rendered symbolic.

The plea is for participation that ensures empowerment and equity for participants, with wider humanitarian or societal benefits. However, without addressing how AI systems structure visibility, voice, and interpretability, such approaches risk reproducing exclusion within ostensibly inclusive frameworks (see Section 2.b). Participatory AI approaches need to specify the reasons for more stakeholder involvement (Donia & Shaw 2021; Iason 2022) presenting frameworks, best practice recommendations and guidelines on how to do it (Lee *et al.* 2019; Mohamed *et al.* 2020). Birhane *et al.* ask for standards for participatory approaches in AI that include claims for reciprocity, reflexivity, and empowerment. As the example of the AI FORA project above shows, developing a critical and emancipatory Participatory AI approach is complex, research-intensive, and costly, and it needs to be highly diversified to support adaptive local solutions. Tailored methods for stakeholder involvement, anticipation, and foresight, such as participatory modelling and simulations, have to co-evolve. In the end, policy impact depends on more than technical insight: It requires institutional willingness, systemic reforms, and sustained engagement with the very communities whose lives are shaped by AI systems in use.

We risk treating 'participation' as a buzzword in both technological governance and democratic innovation, with its meaning emptied of critical content and used as a co-opting mechanism for a-democratic systems (Lamprianou 2013). In many AI-enabled participatory processes, algorithmic tools are used to filter, structure or even generate citizen input (Gerdes 2022). Yet this raises fundamental questions that critical and humanist agendas must include: who is designing these systems? Whose knowledge counts as 'relevant'? And who is rendered invisible through these seemingly inclusive mechanisms? These questions highlight that inclusion cannot be reduced to access but must engage with how knowledge is hierarchised and filtered in AI-mediated environments (see Section 2.c).

Our collective discussion approached these questions through a feminist and decolonial lens, emphasising how AI systems, if left unexamined, may entrench the very exclusions they claim to solve (Adams 2021). This entrenchment occurs through the reproduction of asymmetries embedded in training data and task design, whereby marginalised groups are under-represented, misrepresented, or systematically associated with risk, deviance, or low-value roles. Empirically, this becomes visible when AI systems generate outputs that disproportionately assign harmful, unsafe, or subordinate actions to specific social groups, or when they fail to recognise their perspectives as relevant inputs. Recent work shows, for instance, how LLM-driven robotic systems can reproduce discrimination and unsafe behaviours when interacting with socially coded categories, reflecting patterns present in training data (Hundt *et al.* 2025).

A particular concern is with how individuals who already face marginalisation due to their gender, class, ethnicity, geography, or bodily conditions are further excluded or distorted by AI-mediated systems. If these systems are used to inform policy, allocate resources, or shape public discourse, then the stakes of participation are existential. As with the other priorities, we identified these questions and principles through sharing ideas on several areas of inquiry. For instance, claims for participation are to be critiqued in the absence of key interventions on the reproduction of sexist socio-technical relations. Think of how AI-enabled harms (such as deepfakes) are politically classified: while non-sexual deepfakes of male politicians are treated as national security threats, gendered or sexualised deepfakes are often trivialised (Gosse and Burkell 2020). This asymmetry demonstrates how AI systems amplify existing social hierarchies, reinforcing differential visibility and legitimacy across groups.

4.c Re-embedding AI in global infrastructures of power

The third priority concerns the environmental and geopolitical implications of AI, particularly the tensions between development models, resource consumption, and global justice (McQuillan 2022). This priority emerges from recognising that AI systems are embedded in material, environmental, and geopolitical infrastructures that shape where and for whom research impact is produced (Srnicek 2026). It extends the challenge of predictive expansion (Section 2.e) by situating it within global asymmetries of resource extraction, labour, and technological control. While the ecological costs of AI, such as water usage, energy consumption, and e-waste, are increasingly acknowledged, they remain largely disconnected from mainstream research and policy agendas. This disconnection reflects a tendency to treat AI as immaterial, obscuring the extractive infrastructures underpinning its development (see Section 2.e). The workshop placed this paradox at the centre of the principles of critical humanist agendas: how can researchers contribute to preventing a future where the Global North continues to maintain its consumption habits under the guise of 'green AI' (Verdecchia *et al.* 2023) while the Global South is simultaneously blamed for its developmental aspirations and exploited for its resources (Muldoon & Wu 2023)?

This priority demands both conceptual and empirical innovation. On one hand, it calls for re-theorising AI as part of a broader political economy of planetary computation. On the other hand, it demands grounded case studies on the socio-environmental consequences of AI infrastructure, from mining and supply chains to cloud data centres and satellite constellations. What emerges from these three priorities is not a static research plan but a structured response to the ways AI reshapes the relationship among knowledge, power, and impact across multiple domains.

5. Concluding discussion: Reclaiming our AI through critical and humanist perspectives

This essay proposes the foundations of a minimal critical framework for examining how AI reshapes the relationship among knowledge production, social relations, and political agency. By identifying key challenges (Section 2), examining how these emerge through reflexive and simulation-based exercises (Section 3), and deriving three priority areas for research agendas (Section 4), we have advanced a structured account of how AI mediates social and political relations and the spectrum of possibilities for critical researchers to intervene in these relations.

5.a What this essay aims to achieve

Thinking about challenges, impacts, and agendas requires questioning agency, the autonomy and capacity to make a difference (Elder-Vass 2010). Do we, as researchers, have agency in examining how AI taps into and reshapes social relations? And at the points where agency is 'chained' by structured power to unequal relations, which are the traction points to unchain our liberatory potentialities (Guglielmo *et al.* in press)? And if we consider society at large, where are the lived experiences that may work as prompters of liberation?

Three main contributions emerge from this analysis. First, the essay provides an integrated framework that connects challenges, empirical observations, and research agendas, addressing a gap in the existing literature, which often treats these dimensions separately. Second, it advances a relational understanding of AI, showing how efficiency, prediction, participation, and infrastructure operate not as isolated dynamics but as interdependent mechanisms shaping socio-technical relations. Third, it repositions research impact as a site of political struggle, where the recognition, circulation, and application of knowledge are structured by AI-mediated processes.

5.b Limitations

The scope of this essay is necessarily selective, given the breadth and complexity of the issues at stake. The analytical focus has been on the nexus among knowledge, social relations, and political agency, examined through the lenses of challenges, impacts, and research agendas. This choice entails the relative marginalisation of other crucial domains, particularly those related to governance frameworks, regulatory design, and technical architectures. While these areas are central to ongoing debates on AI, the present analysis deliberately foregrounds the epistemic and relational dimensions through which AI mediates what counts as knowledge and how it becomes impactful. Therefore, this limitation reflects a prioritisation: understanding these mediations is a necessary step to critically inform governance and design interventions.

A second limitation concerns the positionality of the perspectives mobilised in this essay. The reflections emerge from a collective process situated within Global North academic institutions, which inevitably shapes both the analytical framing and the range of experiences considered. Although feminist and decolonial epistemologies are taken seriously as critical lenses, their incorporation remains mediated by institutional and geographical constraints. This raises a fundamental tension: attempts to address global asymmetries in AI may themselves reproduce partial viewpoints if they are not grounded in more diverse epistemic and material contexts. Future research must therefore move beyond incorporating critical perspectives as frameworks and instead engage in forms of collaboration that redistribute epistemic authority across different regions and knowledge traditions. Research and collaboration should not be extractive and solely based on localised knowledge, instead it should engage scholars from the Global South to share expertise and to co-design and foster equitable research partnerships which address knowledge, debates and concerns about the global proliferation of AI.

A third limitation concerns the nature of the dialogic space produced by the workshop and the analysis developed in this essay. While the encounter enabled the articulation of a spectrum of positions ranging from anti-AI to reform-oriented approaches, it does not resolve the conditions under which meaningful dialogue between these positions can be sustained. The coexistence of fundamentally divergent normative orientations raises unresolved questions about how such perspectives can interact without being subsumed into dominant frameworks or reduced to strategic compromise. Addressing this limitation requires moving beyond episodic exchanges towards the design of sustained infrastructures of dialogue that accommodate disagreement while enabling collective inquiry.

These limitations do not weaken the essay's contribution; rather, they delineate three interrelated directions for future research. First, they call for a closer integration between epistemic critique and the study of governance, regulation, and system design, examining how the dynamics identified here are concretely embedded in institutional and technical arrangements. Second, they highlight the need for genuinely transnational and plural research practices that are not merely inclusive but co-produced across different epistemic locations. Third, they point to the importance of developing methodological and institutional formats that enable structured engagement between divergent critical positions, particularly between anti-AI and reform-oriented perspectives.

5.c Our AI: Way forward to advance critical and humanist perspectives

Building on the analytical contributions and limitations outlined above, this essay advances two complementary perspectives to reorient research on AI, society, and politics: a critical perspective and a humanist perspective. These are not abstract normative positions but methodological orientations that respond directly to the challenges and priorities identified throughout the analysis.

First, there is a need to be critical. This means that research is essential to uncover exploitation, critique structured power, and envision better forms of human organisation (Donoghue 2018). This understanding aligns with Andrew Sayer's conception of critical social science as an evaluative and normative practice, concerned not only with explanation but with assessing social arrangements in terms of human well-being, flourishing, and the reduction of unnecessary suffering (Sayer 2009). Especially in the field of evolving human/techno-apparatus relations, being critical requires changing the perspective from which we view things. While extant research on AI, society and politics has opened relevant routes of inquiry, too often the perspective is on two 'objects', institutions and technical systems, and on how better regulations (from the top down) and increased digital and AI literacy (from the bottom up) can reduce bias and destructive behaviours in AI usages (see, for instance: Cihon *et al.* 2020; Mazzucato *et al.* 2022). However, these approaches leave the core challenges we have identified throughout unchallenged. If we recognise that a strict definition of efficiency is problematic because it does not allow for creativity and generativity, what is the alternative? Who would consider research that is dedicated to being anti-efficiency impactful? At the same time, when we think about participation, there is a tendency to incorporate AI into participatory processes that do not address the key problem of how to facilitate, enable and favour the inclusion of those who are structurally excluded. While developing critical perspectives on AI, politics, and society is an open-ended and contested logic, there is fertile ground for informing new routes of research. The constant reflection on

the assumptions underlying how we conduct research, the search for alternatives, the constraints of different spaces of social possibility, and the shifts in points of observation are key principles for looking differently at the phenomena imbricated in the socio-technical relations we have presented throughout.

The claim for critical perspectives requires taking feminist and decolonial epistemologies seriously. They are not just critical contributions that the mainstream North-centric research permits, tolerates or co-opts, but are incredibly powerful heuristics for reclaiming human agency in the social relations mediated by technological systems. The situatedness of who uses AI, what for, and within which institutions cannot be an appendix of research. It must become the core of a new humanist and critical perspective on these relations. This means that there is no one-size-fits-all 'good AI', as technological systems and those who use them are always situated in specific social and political regimes. Each regime comes with specific asymmetric power relations and intersections of exploitation that resist one 'good AI' for all, everywhere, every time. As our collective reflection has proposed, there is an emerging spectrum of critical positions in scholarship on society on AI that ranges, from being anti-AI, that is to say, the idea that there are whole new technologies that humans should have agency to resist and reject, to be better-AI oriented, that is the case when researchers engaging with stakeholders are normatively committed to bias reduction and the design of participatory processes to overcome digital exclusion. So far, these two sides of the spectrum do not have sufficient space for dialogue. This, instead, is essential. And taking humanist and critical perspectives is key to empowering both, without dismissing the dialectics and tensions inherent in genuinely pluralist research practices.

The essay has proposed critical and humanist perspectives to repurpose and refocus their ways of looking at social relations mediated by AI.

With this essay, we propose three core principles to this reclaiming process:

1. At present, techno-oligarchs control the infrastructure and narrative related to AI development to an extent that is inherently a-democratic (at best). As such, the current path of AI development must not be normalised, but counteracted.

2. If AI may serve the common good, 'us' as humans must have the most extensive range of options and choices on the direction this may take. There must always be an off button and the choice of which systems are used for which purposes must be a collective common good.

3. Pluralism is not a decorative element of a-democratic regimes: if human creativity is put at risk by epistemological loops in AI circuits of knowledge production, enacting pluralism of experiences and perspectives must be at the core of every research process with a critical lens.

The reflections in this essay also serve as a call to further advance the practice of reflection. There may be some over-hype around the real impacts of AI. Yet, there is at least a constitutive element in how discourse on AI is permeating deeply into our lives. As such, this provides an important opportunity to reflect on what it means to be researchers, and what it means to be critical in the 21st century.

Ethics and consent

Ethical approval and consent were not required, as the essay stems from the authors' own conceptualisations.

Data availability

Data from the workshop, including pre-event anonymised survey responses, and notes from Lab Sessions one and two, as mentioned in the essay, are available on Zenodo Link: https://doi.org/10.5281/zenodo.17272872 (Guglielmo, 2025), with license Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

References

Adams R: **Can artificial intelligence be decolonized?** *Interdisci. Sci. Rev.* 2021; **46**(1–2): 176–197.
Publisher Full Text

Ahrweiler P, Gilbert N: *Participatory Modelling and Simulation to Improve AI-based Public Social Services–Scientific Advice for Policy and Practice.* Springer; forthcoming.

Ahrweiler P, editor: **Participatory Artificial Intelligence in public social Services.** *From Bias to Fairness in assessing Beneficiaries.* Springer; 2025.
Publisher Full Text

Ahrweiler P, Gilbert N, Juranyi Z, *et al.*: **Using ABM and Serious Games to create 'Better AI'** *2024 Annual Modeling and Simulation Conference (ANNSIM)*. 2024; **2024**. 1–16.
Publisher Full Text

Al Naqbi H, Bahroun Z, Ahmed V: **Enhancing Work Productivity through Generative Artificial Intelligence: A Comprehensive Literature Review.** *Sustainability.* 2024; **16**(3): 1166.
Publisher Full Text

Altenried M: *The Digital Factory: The Human Labor of Automation.* Chicago: University of Chicago Press; 2022.
Publisher Full Text

Altenried M: **The platform as factory: Crowdwork and the hidden labour behind artificial intelligence.** *Ca. Class.* 2020; **44**(2): 145–158.
Publisher Full Text

Alvarado R: **AI as an Epistemic Technology.** *Sci. Eng. Ethics.* 2023; **29**(5): 32.
PubMed Abstract | **Publisher Full Text**

Amoore L, Campolo A, Jacobsen B, *et al.*: **A world model: On the political logics of generative AI.** *Polit. Geogr.* 2024; **113**: 103134.
Publisher Full Text

Ball AF: **To Know Is Not Enough: Knowledge, Power, and the Zone of Generativity.** *Educ. Res.* 2012; **41**(8): 283–293.
Publisher Full Text

Bandi A, Kongari B, Naguru R, *et al.*: **The Rise of Agentic AI: A Review of Definitions, Frameworks, Architectures, Applications, Evaluation Metrics, and Challenges**', *Future Internet.* 2025; **17**(9): 404.
Reference Source

Bankins S, Ocampo AC, Marrone M, *et al.*: **A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice.** *J. Organ. Behav.* 2024; **45**(2): 159–182.
Publisher Full Text

Barry I, Stephenson E: **The Gendered, Epistemic Injustices of Generative AI.** *Aust. Fem. Stud.* 2025; **40**(123): 1–21.
Publisher Full Text

bell hooks: *Talking Back: Thinking Feminist, Thinking Black.* New York: Routledge; 2014.
Publisher Full Text

Bergold J, Thomas S: **Participatory research methods: A methodological approach in motion.** *Histor. Soc. Res./ Historische Sozialforschung.* 2012: 191–222.

Berenskoetter F: **Anxiety, time, and agency.** *Int Theory.* 2020; **12**(2): 273–290.
Publisher Full Text

Berry DM: *Critical Theory and the Digital.* 1st ed. New York: Bloomsbury Academic; 2014.

Birhane A, Isaac W, Prabhakaran V, *et al.*: **'Power to the people? Opportunities and challenges for participatory AI.** In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization' (EAAMO'22).* 2022; pp. 1–8.
Publisher Full Text

Bonini T, Trere E: *Algorithms of Resistance: The Everyday Fight against Platform Power.* Cambridge: MIT Press; 2024.
Publisher Full Text

Boyd D, Crawford K: **Critical Questions for Big Data.** *Inf. Commun. Soc.* 2012; **15**(5): 662–679.
Publisher Full Text

Bracci E: **The loopholes of algorithmic public services: an "intelligent" accountability research agenda.** *Account. Audit. Account. J.* 2022; **36**(2): 739–763.
Publisher Full Text

Burnham P: **'Neo-liberalism, crisis and the contradictions of depoliticisation',** ***Partecipazione e conflitto.*** *Open J Sociopolit. Stud.* 2017; **10**(2): 357–380.

Cant C: *Riding for Deliveroo: Resistance in the New Economy.* Cambridge, UK: Polity; 2020.

Carabantes M: **Black-box artificial intelligence: an epistemological and critical analysis.** *AI Soc.* 2020; **35**(2): 309–317.
Publisher Full Text

Chehoudi R: **Artificial intelligence and democracy: pathway to progress or decline?** *J. Inform. Tech. Polit.* 2025: 1–16.

Cihon P, Maas MM, Kemp L: **Fragmentation and the Future: Investigating Architectures for International AI Governance.** *Glob. Pol.* 2020; **11**(5): 545–556.
Publisher Full Text

Coeckelbergh M: **Time machines: Artificial intelligence, process, and narrative.** *Philos Technol.* 2021; **34**(4): 1623–1638.
Publisher Full Text

Coeckelbergh M: *Digital technologies, temporality, and the politics of co-existence.* Cham: Springer; 2022.
Publisher Full Text

Coeckelbergh M: **Artificial intelligence, the common good, and the democratic deficit in AI governance.** *AI Ethics.* 2024; **5**: 1491–1497.
Publisher Full Text

Coeckelbergh M: **Three challenges for a global AI ethics: towards a more relational normative vision.** *AI Ethics.* 2025; **5**(5): 5527–5533.
Publisher Full Text

Cornish F, Breton N, Moreno-Tabarez U, *et al.*: **Participatory action research.** *Nat. Rev. Methods Primers.* 2023; **3**(1): 34.
Publisher Full Text

Crawford K: *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* New Haven: Yale University Press; 2021.
Publisher Full Text

Danaher J: **The Threat of Algocracy: Reality, Resistance and Accommodation.** *Philos. Technol.* 2016; **29**(3): 245–268.
Publisher Full Text

D'Ignazio C, Klein LF: *Data Feminism.* Cambridge: MIT Press; 2020.
Publisher Full Text

Dauvergne P: **Is artificial intelligence greening global supply chains? Exposing the political economy of environmental costs.** *Rev. Int. Polit. Econ.* 2022; **29**(3): 696–718.
Publisher Full Text

De Swarte T, Boufous O, Escalle P: **Artificial intelligence, ethics and human values: the cases of military drones and companion robots.** *Artif. Life Robot.* 2019; **24**(3): 291–296.
Publisher Full Text

Djeffal C, Siewert MB, Wurster S: **Role of the state and responsibility in governing artificial intelligence: a comparative analysis of AI strategies.** *J. Eur. Publ. Policy.* 2022; **29**(11): 1799–1821.
Publisher Full Text

Donia J, Shaw JA: **Co-design and ethical artificial intelligence for health: An agenda for critical research and practice.** *Big Data Soc.* 2021; **8**(2)
Publisher Full Text

Donoghue M: **Beyond hegemony: Elaborating on the use of Gramscian concepts in Critical Discourse Analysis for Political Studies.** *Polit Stud.* 2018; **66**(2): 392–408.
Publisher Full Text

Duberry J: *Artificial Intelligence and Democracy: Risks and Promises of AI-Mediated Citizen–Government Relations.* London: Edward Elgar Publishing; 2022.

Dunleavy P, Margetts H: **Data science, artificial intelligence and the third wave of digital era governance.** *Public Policy Admin.* 2023: 09520767231198737.

Edwards SM: **The Technology Paradox: Efficiency Versus Creativity.** *Creat. Res. J.* 2001; **13**(2): 221–228.
Publisher Full Text

Emery-Xu N, Jordan R, Trager R: **International governance of advancing artificial intelligence.** *AI Soc.* 2025; **40**(4): 3019–3044.
Publisher Full Text

Elder-Vass D: *The Causal Power of Social Structures: Emergence, Structure and Agency.* Cambridge: Cambridge University Press; 2010.
Publisher Full Text

Farrugia D: **The reflexive subject: Towards a theory of reflexivity as practical intelligibility.** *Curr. Sociol.* 2013; **61**(3): 283–300.
Publisher Full Text

Floridi L, Cowls J, Beltrametti M, *et al.*: **AI4People-An Ethical Framework for a Good AI Society: Opportunities Risks, Principles, and Recommendations.** *Minds Mach.* 2018; **28**: 689–707.
Publisher Full Text

Floridi L: **The Fight for Digital Sovereignty: What It Is, and Why It Matters, Especially for the EU.** *Philos. Technol.* 2020; **33**(3): 369–378.
Publisher Full Text

Floridi L: **Why the AI hype is another tech bubble.** *Philos. Technol.* 2024; **37**(4): 128.
Publisher Full Text

Fourcade: **Ordinalization: Lewis A. Coser Memorial Award for Theoretical Agenda Setting 2014.** *Sociol. Theory.* 2016; **34**(3): 175–195.
Publisher Full Text

Fraser N: *Cannibal Capitalism: How our System is Devouring Democracy, Care, and the Planet and What We Can Do About It.* London: Verso Books; 2023.

Freire P: *Pedagogy of Hope: Reliving Pedagogy of the Oppressed.*Translated by Barr RR. New York: Continuum; 1994.

Fuchs C: **Capitalism, patriarchy, slavery, and racism in the age of digital capitalism and digital labour.** *Crit. Sociol.* 2018; **44**(4–5): 677–702.
Publisher Full Text

Fuchs C: **Henri Lefebvre's theory of the production of space and the critical theory of communication.** *Commun. Theory* 2019; **29**(2): 129–150.
Publisher Full Text

Gerdes A: **A participatory data-centric approach to AI Ethics by Design.** *Appl. Artif. Intell.* 2022; **36**(1): 2009222.
Publisher Full Text

Gilbert J, Williams A: *Hegemony Now: How Big Tech and Wall Street Won the World (And How We Win it Back).* London: Verso Books; 2022.

Gillespie T: **Generative AI and the politics of visibility.** *Big Data Soc.* 2024; **11**(2): 20539517241252131.
Publisher Full Text

Gosse C, Burkell J: **Politics and porn: how news media characterizes problems presented by deepfakes.** *Crit. Stud. Media Commun.* 2020; **37**(5):

497–511.
Publisher Full Text

Grossmann I, Feinberg M, Parker DC, *et al.*: **AI and the transformation of social science research.** *Science.* 2023; **380**(6650): 1108–1109.
Publisher Full Text

Guglielmo M: **'Data repository of the international workshop: political agency between human and Artificial Intelligence', Valencia, Spain 10 and 11 September 2025 (1.0).** *Zenodo.* 2025.
Publisher Full Text

Guglielmo M, O'Loughlin B, Snowball H: **What Is at Stake? The Co-Encoding/Decoding Model of Human–Generative Artificial Intelligence Communication.** *Int. J. Commun.* 2026; **20**: 669–692.
Publisher Full Text

Guglielmo M, Heinrichs PS, O'Loughlin B: *Navigating Platform Power: Agency and Resistance in Digital Spaces.* Manchester: Manchester University Press; in press.

Halsema IV: **Feminist Methodology and Gender Planning Tools: Divergences and Meeting Points.** *Gend. Technol. Dev.* 2003; **7**(1): 75–89.
Publisher Full Text

Henman P: **Improving public services using artificial intelligence: possibilities, pitfalls, governance.** *Asia Pacific J. Public. Admin.* 2020; **42**(4): 209–221.
Publisher Full Text

Hepp A, Jarke J, Kramp L: *New Perspectives in Critical Data Studies The Ambivalences of Data Power.* Cham: Springer; 2022.
Publisher Full Text

Hirsch-Kreinsen H: **Artificial intelligence: a "promising technology".** *AI Soc.* 2024; **39**(4): 1641–1652.
Publisher Full Text

Hundt A, Azeem R, Mansouri M, *et al.*: **LLM-driven robots risk enacting discrimination, violence, and unlawful actions.** *Int. J. Soc. Robot.* 2025; **17**(11): 2663–2711.
Publisher Full Text

Iason G: **Toward a Theory of Justice for Artificial Intelligence.** *Daedalus.* 2022; **151**: 218–231.
Publisher Full Text

Jacobsen BN: **Review article: Data, algorithms and the asymmetric politics of extraction.** *Sociol. Rev.* 2025; **73**(6): 1349–1367.
Publisher Full Text

Jordan T: *Information politics: liberation and exploitation in the digital society.* London: Pluto Press; 2015.
Publisher Full Text

Khosravi M, Zare Z, Mojtabaeian SM, *et al.*: **Artificial intelligence and decision-making in healthcare: a thematic analysis of a systematic review of reviews.** *Health Serv. Res. Manag. Epidemiol.* 2024; **11**: 23333928241234863.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Koebis N, Starke C, Rahwan I: **The promise and perils of using artificial intelligence to fight corruption.** *Nature Mach. Intell.* 2022; **4**: 418–424.
Publisher Full Text

König PD, Wenzelburger G: **Opportunity for renewal or disruptive force? How artificial intelligence alters democratic politics.** *Gov. Inf. Q.* 2020; **37**(3): 101489.
Publisher Full Text

Kuntsman A, Miyake E: *Paradoxes of Digital Disengagement: In Search of the Opt-Out Button.* London: University of Westminster Press; 2022.

Lamprianou I: **Contemporary Political Participation Research: A Critical Assessment.** In Demetriou KN, editor. *Democracy in Transition: Political Participation in the European Union.* Springer Berlin Heidelberg; 2013; pp. 21–42.

Lee MK, Kahng A, Kim J, *et al.*: **WeBuildAI: Participatory framework for algorithmic governance.** *Proc. ACM Human-Comput. Interact.* 2019; **3**: 1–35.
Publisher Full Text

Lehuedé S: **The Double Helix of Data Extraction: Radicalising Reflexivity in Critical Data Studies.** *Technol. Regul.* 2024; **2024**: 84–92.
Publisher Full Text

Lindgren S: *Handbook of Critical Studies of Artificial Intelligence.* London: Edward Elgar Publishing; 2023.
Publisher Full Text

Lindgren S: *Critical Theory of AI.* Cambridge: Polity Press; 2024.

List C: **Group Agency and Artificial Intelligence.** *Philos. Technol.* 2021; **34**(4): 1213–1242.
Publisher Full Text

Löhr K, Weinhardt M, Sieber S: **The "World Café" as a participatory method for collecting qualitative data.** *Int. J. Qual. Methods.* 2020; **19**: 1609406920916976.
Publisher Full Text

Luchetta G: **Impact Assessment and the Policy Cycle in the EU.** *Eur. J. Risk Regul.* 2012; **3**(4): 561–575.
Publisher Full Text

Manheim K, Kaplan L: **Artificial Intelligence: Risks to Privacy and Democracy.** *Yale J. Law Technol.* 2019; **21**: 106–188.

Mansouri MI, Bailey DJ: **How to be 'anti-AI' in the 21st century: overcoming the inevitability narrative.** *Global Polit. Econ.* 2025: 1–10.

Mazzucato M, Schaake M, Krier S, *et al.*: *Governing artificial intelligence in the public interest.* Institute for Innovation and Public Purpose; 2022.
Reference Source

McAdam M, Debackere K: **Beyond 'triple helix' toward 'quadruple helix' models in regional innovation systems: Implications for theory and practice.** *R & D Manag.* 2018; **48**: 3–6.
Publisher Full Text

McQuillan D: *Resisting AI: An Anti-fascist Approach to Artificial Intelligence.* Bristol: Bristol University Press; 2022.
Publisher Full Text

Micheli M, Ponti M, Craglia M, *et al.*: **Emerging models of data governance in the age of datafication.** *Big Data Soc.* 2020; **7**(2): 1–15.
Publisher Full Text

Miragoli M: **Conformism, Ignorance & amp; Injustice: AI as a Tool of Epistemic Oppression.** *Episteme.* 2025; **22**(2): 522–540.
Publisher Full Text

Mitchell S, Potash E, Barocs S, *et al.*: **Algorithmic fairness: Choices, assumptions and definitions.** *Ann. Rev. Stat. Its Appl.* 2021; **8**: 141–163.
Publisher Full Text

Mohamed S, Png MT, Isaac W: **Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence.** *Philos. Technol.* 2020; **33**: 659–684.
Publisher Full Text

Moore P: *The Quantified Self in Precarity: Work, Technology and What Counts.* London: Routledge; 2017.
Publisher Full Text

Morozov E: *Can AI Break Out of Panglossian Neoliberalism? What Big Tech has done to our institutional and infrastructural imagination.* Boston; 2024.

Mühlhoff R: *The Ethics of AI. Power, Critique, Responsibility.* Bristol: Bristol University Press; 2025.
Publisher Full Text

Muldoon J, Graham M, Cant C: *Feeding the Machine: The Hidden Human Labour Powering AI.* Edinburgh: Canongate Books; 2024.

Muldoon J, Wu BA: **Artificial Intelligence in the Colonial Matrix of Power.** *Philos. Technol.* 2023; **36**(4): 80.
Publisher Full Text

Neumann O, Guirguis K, Steiner R: **Exploring artificial intelligence adoption in public organizations: a comparative case study.** *Public Manag. Rev.* 2024; **26**(1): 114–141.
Publisher Full Text

Noble SU: *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York: NYU; 2018.
Publisher Full Text

Novelli C, Sandri G: **Digital Democracy.** In Floridi L, Taddeo M, editors. *A Companion to Digital Ethics.* Vol. 20. Oxford: John Wiley & Sons; 2025; pp. 241–256.
Publisher Full Text

Nowotny H: *In AI we trust: Power, illusion and control of predictive algorithms.* Hoboken: John Wiley & Sons; 2021.

Pagano T, Loureiro R, Lisboa F, *et al.*: **Bias and unfairness in machine learning models: A systematic review on datasets, tools, fairness metrics, and identification and mitigation methods.** *Big Data Cogn. Comput.* 2023; **7**(1): 15.
Publisher Full Text

Park YJ, Jones-Jang SM: **Surveillance, security, and AI as technological acceptance.** *AI Soc.* 2023; **38**(6): 2667–2678.
Publisher Full Text

Peterson AJ: **AI and the problem of knowledge collapse.** *AI Soc.* 2025; **40**(5): 3249–3269.
Publisher Full Text

Phan T, Wark S: **Racial formations as data formations.** *Big Data Soc.* 2021; **8**(2).
Publisher Full Text

Pirina G: **I costi umani e ambientali del digitale.** *Il Mulino, Rivista trimestrale di cultura e di politica.* 2022; **3**: 57–64.

Sayer A: **Who's Afraid of Critical Social Science?** *Curr. Sociol.* 2009; **57**(6): 767–786.
Publisher Full Text

Sauer F, Schörnig N: **Killer drones: The 'silver bullet' of democratic warfare?** *Secur. Dialogue.* 2012; **43**(4): 363–380.
Publisher Full Text

Schaupp S: **Cybernetic proletarianization: Spirals of devaluation and conflict in digitalized production.** *Ca. Class* 2022; **46**(1): 11–31.
Reference Source | **Publisher Full Text**

Sieber R, Brandusescu A, Sangiambut S, *et al.*: **What is civic participation in artificial intelligence?** *Environ. Plan B Urban Anal. City Sci.* 2025; **52**(6):

1388–1406.
Publisher Full Text

Smuha NA: **From a 'race to AI'to a 'race to AI regulation': regulatory competition for artificial intelligence.** *Law Innov. Technol.* 2021; **13**(1): 57–84.
Publisher Full Text

Srnicek N: *Silicon Empires: The Fight for the Future of AI.* Cambridge: Polity Press; 2026.

Stingl AI: *The Digital Coloniality of Power: Epistemic Disobedience in the Social Sciences and the Legitimacy of the Digital Age.* London: Lexington Books; 2015.
Publisher Full Text

Supriyanto EE, Saputra J: **Big Data and Artificial Intelligence in Policy Making: A Mini-Review Approach.** *Int. J. Adv. Soc. Sci. Human.* 2022; **1**(2): 58–65.
Publisher Full Text

Taeihagh A: **Governance of artificial intelligence.** *Polic. Soc.* 2021; **40**(2): 137–157.
Publisher Full Text

Tarr J, Gonzalez-Polledo E, Cornish F: **On liveness: using arts workshops as a research method.** *Qual. Res.* 2018; **18**(1): 36–52.
Publisher Full Text

Ulnicane I, Erkkilä T: **Politics and policy of Artificial Intelligence.** *Rev. Policy Res.* 2023; **40**(5): 612–625.

van Doorn N, Badger A: **Platform Capitalism's Hidden Abode: Producing Data Assets in the Gig Economy.** *Antipode.* 2020; **52**(5): 1475–1495.
Publisher Full Text

van Witteloostuijn A, Vanderstraeten J, Slabbinck H, *et al.*:
From explanation of the past to prediction of the future: A comparative and predictive research design in the Social Sciences. *Social Sci. Human Open.* 2022; **6**(1): 100269.
Publisher Full Text

Vargas-Parada L: **Large language models are biased—local initiatives are fighting for change.** *Nature.* 2025.
PubMed Abstract | **Publisher Full Text**

Vercellone C: **From the mass-worker to cognitive labour: Historical and theoretical considerations.** In van der Linden M, Roth KH, editors. *Beyond Marx.* Leiden: Brill; 2014; pp. 417–443.
Publisher Full Text

Verdecchia R, Sallou J, Cruz L: **A systematic review of Green AI.** *WIREs Data Min. Knowl. Disc.* 2023; **13**(4): e1507.

Wigginton B, Lafrance MN: **Learning critical feminist research: A brief introduction to feminist epistemologies and methodologies.** *Fem. Psychol.* 2019: 0959353519866058.
Publisher Full Text

Woodcock J: *The Fight Against Platform Capitalism: An Inquiry into the Global Struggles of the Gig Economy.* London: University of Westminster Press; 2021.
Publisher Full Text

Zuiderwijk A, Chen Y-C, Salem F: **Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda.** *Gov. Inf. Q.* 2021; **38**(3): 101577.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 2

Reviewer Report 26 August 2026

https://doi.org/10.21956/openreseurope.26289.r79564

© **2026 Scheffauer R.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Rebecca Scheffauer**
[1] German Institute for Economic Research Berlin (Ringgold ID: 28354), Berlin, Berlin, Germany
[2] Freie Universitat Berlin Institut fur Soziologie (Ringgold ID: 54201), Berlin, Berlin, Germany

I appreciate the work the authors have invested in improving this essay.
The structure is now clearer and the narrativ much easier to follow.
I consider all my previously raised points of criticism as taken care of!

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I have published several works in the field of algorithms and AI touching upon different topics from ethics to applications.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 26 August 2026

https://doi.org/10.21956/openreseurope.26289.r79656

© **2026 Waelen R.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Rosalie Waelen**
University of Bonn, Bonn, Germany

This is a well-written and timely piece. I think the critical and humanist approach that is advocated for resonates with a lot of recent literature on AI ethics (broadly understood). I can think of even more sources the authors could cite, but they already do a great job at embedding it in on-going

debates. I have one comment: The notion of 'plurality' seems important, but not really defined in the paper. I think a working definition should be given. A possible source to draw from here would be Hannah Arendt's writing on plurality in the political sphere in her book 'The Human Condition'. Otherwise, I have no major comments or concerns, so my recommendation is to accept this manuscript (with minor revisions) for indexing.

Is the topic of the essay discussed accurately in the context of the current literature?
Yes

Is the work clearly and cogently presented?
Yes

Is the argument persuasive and supported by appropriate evidence?
Yes

Does the essay contribute to the cultural, historical, social understanding of the field?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Philosophy of technology, applied ethics, social critique of AI

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 22 August 2026

https://doi.org/10.21956/openreseurope.26289.r79565

© 2026 Bailey D. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

David Bailey
University of Birmingham, Birmingham, England, UK

Thanks to the authors for responding to my review so thoroughly. I have no further comments to make and am happy to now approve the article for indexing.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: political economy, social movements, forms of resistance/dissent, specifically with a recent focus on the contestation of AI as a new capitalist technology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 22 November 2025

https://doi.org/10.21956/openreseurope.23231.r62812

© **2025 Bailey D.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

David Bailey
University of Birmingham, Birmingham, England, UK

The article presents an excellent and innovative discussion of some of the key questions and challenges that researchers and practitioners face in considering how to study, understand and advocate for alternatives forms of development of, AI. It is based on a collaborative exercise stemming from a two-day workshop at the University of Valencia, in which participants were asked to engage in a series of highly innovative and extremely productive and experimental methods of collaboration, theory-building, reflection and debate, in order to bring to the surface a series of shared insights, concerns and suggestions.

The paper does a really excellent job of introducing the reader both to the methods employed in the workshop, and the debates and observations that arose as a result of the different collaborative and reflexive exercises. In doing so, it highlights five core challenges/questions that face the use of AI in terms of the possibility for it to enhance human agency in an emancipatory or socially-valuable way. These are: the question of efficiency (for whom, and in what form), the question of participation (or exclusion), the question of what forms of knowledge are created by AI, the problem of a lack of historicity, and the problems of prediction associated with a technological development which itself is not guaranteed to be sustainable in terms of our shared future. Alongside these challenges, the authors together focus on potential principles regarding how we might think, study and do AI differently, focusing especially on the need for a critical/humanist perspective, and ending with a call for AI research to foreground three core principles: to challenge the current a-democratic form of AI, to ensure that AI is human-led in its nature, and to centre pluralism.

The paper overall amounts to a major and highly significant contribution to our thinking and collective process of reflection by AI-critical scholars. It brings together an extremely impressive and highly relevant group of researchers from across a range of backgrounds. The workshop methods used and described are really outstanding in terms of methodological innovation and the findings that they have generated.

My only suggestions are therefore really around areas that the authors might consider in terms of further clarifications for the reader, none of which I would consider essential (but nevertheless some consideration or clarification might improve the paper if the authors are so inclined), as follows:

(1) page 7 - the authors mention the 'deleterious effects of quantified efficiency' - but perhaps

some elaboration on what these deleterious effects are would be helpful.

(2) also page 7 - the authors ask 'If AI is used to 'streamline' citizen input or 'enhance' participatory budgeting, who decides which inputs matter? Who interprets dissent?' - but this doesn't seem to be directly answered in the text. I would think that the answer is the particular LLM that is used to perform these tasks? In which case, don't we know the answer? - the LLM - and isn't the broader concern not that we don't know who will do these tasks, but rather that we know it is an LLM (and not deliberating humans) and that this is itself the problem?

(3) The paper makes a strong case for a critical perspective and humanism. But I think some more work could have been done to spell out what exactly that means. Page 8 mentions 'situatedness, plurality, and narrative agency' with a reference to bell hooks. But some further elaboration of why these principles in particular are chosen, and in what way hooks is responsible for developing these concepts, would have been helpful. Likewise page 12 (conclusion) has some further discussion of what it means to be critical and humanist - 'to uncover exploitation, critiquing structured power, and envisioning better forms of human organisation' - but again it's not entirely clear why this is considered to be critical/humanist. Some further elaboration of what exactly it means to be critical and humanist would therefore I think have been helpful. I find Andrew Sayer's article on this question to be quite helpful and perhaps worthwhile incorporating in some way, and which I think would also fit broadly speaking with the approach that is being advocated: Sayer, A. (2009). Who's Afraid of Critical Social Science? *Current Sociology*, *57*(6), 767-786. Related, I think there is a possibility that the article's call for a humanist approach might stand accused of essentialising what it means to be human/humanist, prompting again a need for a bit more of a detailed discussion of what it means to be 'human' and why we should privilege such a value (and indeed whether to do so is as unproblematic as the paper perhaps suggests). I think the above contribution from Sayer can also help on this question.

(4) page 9 - the authors discuss how AI as a system might have a cannibalising effect - with reference to Nancy Fraser's book using the same concept - but I'm not sure whether Fraser is directly referring to AI when she develops this theme (I think she's more generally talking about capitalism?) - perhaps some clarification on this point?

(5) page 10 - the authors describe what sounds like a really very interesting simulation exercise that used agentic AI chatbots 'designed with GPT-4o' - but perhaps some elaboration on how this was done would be helpful. In particular, I wasn't sure whether the chatbots *were* GPT-4o, or whether GPT-4o had designed them itself (vibe coding?). If the latter then this would also be quite interesting to the reader, I would think. Either way some elaboration on how these chatbot simulators were created would I think be of interest to the reader. This would also help later in the same section, where the authors say that these experiments 'offered critical insight into the simulation of political agency by GenAI models, how such systems are trained to emulate rhetorical strategies, project ideological stances, or manage dissent' - all of which sounds very interesting, but I think the reader would have benefited by having more information to clarify *in what way* does Gen AI do all of these things?

6) Also in the same section, the authors describe how feminist scholars engaged with the chatbot on questions of EU foreign policy, only to find that AI systems 'can re-inscribe hierarchies of knowledge and authority' - but more could have been said I think, again, to help the reader to understand *in what way* did AI do this re-inscribing?

7) page 11 - the discussion of scenarios that were mapped out during the workshop sounds very interesting - but the reader is not told exactly what these alternative scenarios looked like. Perhaps this could be added in?

8) The authors mention how a feminist and decolonial lens allows us to understand how AI systems might entrench the exclusions they claim to solve. But perhaps more detail on how this entrenching happens and how we know it's existence when it occurs might be of interest. Reference to recent research by one of the co-authors on this could perhaps be used to illustrate this?: Hundt, A., Azeem, R., Mansouri, M. *et al.* LLM-Driven Robots Risk Enacting Discrimination, Violence, and Unlawful Actions. *Int J of Soc Robotics* (2025).

9) the concluding section - the authors claim to identify 'two core perspectives' - with the first clearly labelled 'Firstly' - but I couldn't find a corresponding 'Secondly'.

10) also in the conclusion - the third principle - 'pluralism' - I wasn't entirely sure how or why pluralism is advocated here. Is a critical and humanist approach the same as a pluralist one? I'm not sure. I would think that some interpretations of what we mean by pluralism are relativist in terms of the values that are being advocated; whereas I would think that some interpretations of critical/humanist thinking is committed to some notion of moral truths (which perhaps is not relativist?). Is this an inconsistency?

Finally, some minor typo/grammatical issues that might be worth addressing:

i) page 5 - 'wealthier job positions' - should be 'higher paid' or something like that, rather than 'wealthier'?
ii) page 6 - a couple of instances of 'the AI' where I think it should be just 'AI'?
iii) top of page 7, second column - I'm not sure what 'striking dialectics' refers to - maybe change?
iv) the end of section "AI, politics and society" - the authors refer at the very end to 'a process in need of critique and reflexivity' - I think this isn't quite the same meaning as was developed in this section. I think (although I may have misunderstood) that a rephrasing along the following lines is closer to the argument that the authors developed: ''also requires critique and reflexivity as intrinsic elements of any research that claims to be impactful' (or something like that).

Is the topic of the essay discussed accurately in the context of the current literature?

Yes

Is the work clearly and cogently presented?

Yes

Is the argument persuasive and supported by appropriate evidence?

Yes

Does the essay contribute to the cultural, historical, social understanding of the field?

Yes

Competing Interests: Non-financial competing interests: I am a current collaborator and previous co-author with a number of the article co-authors. One of the article co-authors is my spouse.

Reviewer Expertise: political economy, social movements, forms of resistance/dissent, specifically with a recent focus on the contestation of AI as a new capitalist technology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 14 November 2025

https://doi.org/10.21956/openreseurope.23231.r62524

© 2025 Scheffauer R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rebecca Scheffauer
[1] German Institute for Economic Research Berlin (Ringgold ID: 28354), Berlin, Berlin, Germany
[2] Freie Universitat Berlin Institut fur Soziologie (Ringgold ID: 54201), Berlin, Berlin, Germany

I appreciate the opportunity to review this interesting essay and thank the authors for highlighting several extremely relevant aspects regarding AI. However, there are some issues I would like to address. It is somewhat difficult to criticize the composition of this piece and of what is mentioned/left out as it is a reflection of what was discussed during a conference. Obviously, what did not come up then is not part of the essay. Nevertheless, there are some omissions and other problems that need to be taken care of before this could be indexing.

My first point is probably the easiest to fix as it regards the structure of the essay. Oftentimes the authors highlight “5 key challenges”, or other short lists of points that are discussed. However, it is oftentimes difficult to find the listed ideas within the headings chosen in the rest of the essay. I do not know what the journal policy is in this regard, but numbering headings would be extremely beneficial here, as it would allow the authors to easily reference in brackets which section of the paper tackles a certain idea.

On page 4, the authors mention that researchers often do not have the space to reflect on the consequences of AI use. I find this statement problematic, as there are plenty of papers discussing problems and limitations of AI, and even papers that investigate direct applications usually at least address some of the consequences the authors are getting at in their discussion section when they consider the implications of their findings.

I find it interesting that the authors mention “political” questions of AI but leave out any discussion of things such as legislation regarding AI. This is a major factor in shaping what can or cannot be done with AI and how it will be done in the future. Thus, completely ignoring this area seems a glaring omission to me.

The last paragraph in the first column on page 5/first paragraph in the second column can be deleted. The limitations should be discussed in a limitations section of the paper (which needs to be added!); the other information is not entirely relevant. Similarly, the paragraph on page 6 starting with “ultimately, the methodology” does not contribute too much to the overall narrative and needs to be removed. There already is a lot of explaining going on concerning the format under which the ideas discussed in the essay were derived, which is – to a degree – necessary. But this section needs to be a bit shorter.

When the authors discuss efficiency of AI, a topic that is not properly addressed is “accurate vs. false” decisions reached by AI. Especially as the authors bring up the examples of AI in warfare and bureaucracy, this is highly relevant.

While the “Efficiency or what?” section covers very broad topics, the next section regarding participation seems oddly specific in contrast. There needs to be a better connection between the two issues. Moreover, while reading said section, I kept noting that the authors never address whose data it is that is reflected in AI decisions. Which data was used to train models? E.g., if data from the web was scraped, was this only English language data, etc.? This general idea very briefly comes up much later in the next section. However, as this is an extremely important topic here as well, this needs to be mentioned (and discussed in more detail) much earlier. In general, the section regarding participation lists many, many different ideas and topics, but fails to discuss many of them in depth. A mere list of potential issues is not particularly helpful, so I would encourage the authors to either streamline this section better by removing some aspects or by keeping all ideas but discussing them better.

On page 9, the authors mention that generative AI is not a Deus ex machina. What is not discussed properly, however, is that AI is never truly innovative, as all it does is regurgitate and recycle based on the data it was fed previously. Again, this is a vital point to the authors’ arguments and, thus, should be elaborated on.
While I found the part of the paper that details the interactions with different chatbot personalities somewhat interesting, it is very different from the rest of the paper. So different in fact, that it does not seem to fit into this essay at all. Either the authors should integrate this better in light of their previously mentioned ideas or remove this section entirely (which might be the better option, as it, quite frankly, feels like reading about a different project).

Finally – and this somewhat circles back to the authors never touching upon legislation – the essay never addresses or even mentions the topics of privacy and data protection, which play a giant role regarding all aspects of AI. I understand that it is difficult to add this to the essay if it simply did not come up during the conference/workshop setting, but the topic cannot be ignored.

Is the topic of the essay discussed accurately in the context of the current literature?
Partly

Is the work clearly and cogently presented?
Partly

Is the argument persuasive and supported by appropriate evidence?
Partly

Does the essay contribute to the cultural, historical, social understanding of the field?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I have published several works in the field of algorithms and AI touching upon different topics from ethics to applications.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.