Creating certainty where there is none: Artificial Intelligence as political concept

Cristóbal Garibay-Petersen, London School of Economics and Political Science

Marta Lorimer, Cardiff University (Corresponding author)

Bayar Menzat, Teschnische Universität Wien

Accepted Manuscript. The version of record will appear in Big Data & Society

Abstract: Recent developments in research on Artificial Intelligence (AI) have prompted a growing politicisation of AI. In this paper, we critically analyse how AI is being construed in public discourse and with what political implications. Approaching AI as a mobilising political concept, and focusing on the public pronouncements made by influential tech commentators, we identify and subject to technical and critical scrutiny four key themes in contemporary discourses on AI. First, we show how AI discourses endorse anthropological commitments that create false equivalences between human and artificial intelligence, and suggest that all are equally affected by AI. Second, we demonstrate that AI discourses unproblematically indulge in agential constructs, which ascribe agency to AI while obfuscating the role of humans in its development. Third, we explain how the economic assumptions made by these discourses support specific political interests. Finally, we show that discourses on AI endorse a set of temporal assumptions that reduce the space for democratic intervention. We conclude that AI is becoming more than what its 'technical'

specifications would warrant; however, this is happening in a way that limits the space for democratic engagement with, and control of, the technology itself.

Keywords: artificial intelligence; assumptions; political concepts; democracy; critical discourse; politicisation.

i. Introduction

Technology is, and has always been, deeply political. Technologies embody political ideas and have political implications (Williams, 1971; Winner, 1980). They are also, frequently, the object of political contestation (Ulnicane and Erkkilä, 2023; van Lente et al., 2013). Artificial intelligence (AI) is no different. It reflects certain political ideas (Crawford, 2021), and has increasingly been brought into the realm of public debate.

The growing politicisation of AI has been most noticeable since the release of ChatGPT in November 2022. If politicisation requires salience, expansion, and polarisation (de Wilde et al., 2016), AI appears to be a thoroughly politicised issue. AI has been the object of a 'hype cycle' (Bender, 2023) of growing expectations – most of them duly reported on in the generalist press and beyond. Discussions previously confined to the expert realm have seeped into the public domain. 'Enthusiasts' have messianically defended the virtues of AI, while 'catastrophists' have alerted to the dangers of developing uncontrollable AI systems. Sitting somewhere in between, critical takes oppose the positions of both catastrophists and enthusiasts, shifting attention to the current effects of existing AI systems. While some of these actors may hold more in common than this simplistic categorisation reflects (Gebru and Torres, 2024), their opposing positions testify to an increasingly polarised debate on AI.

The politicisation of AI can be a positive development, bringing AI into the realm of democratic contestation. However, this will only hold if it is being done in a way that facilitates democratic engagement. In this paper, we critically analyse these processes of politicisation by

looking at how AI is being construed, and to what political effects. While 'AI' encompasses a variety of approaches, given their current public prominence, we focus on foundation-model-based systems, especially large language models (LLMs). Starting with Koselleck's understanding of mobilising political concepts, we draw from conceptual history's theoretical resources to study 'AI' as a contested concept that one cannot simply divorce from its future-inflected use. Doing so, we contribute to existing research in critical AI studies (Bareis and Katzenbach, 2022; Coeckelbergh, 2022; Heffernan, 2019; Lindgren, 2023) by bringing attention to the mobilising nature and democratic implications of the claims that are made about AI. We consider that adopting such a stance is essential to avoid falling prey to hyped-up narratives. Narratives, whether hyped up or not, have a performative capacity (Bareis and Katzenbach, 2022; Kim, 2023; Richter et al., 2023; van Lente et al., 2013). One should make sure that the activities aimed at determining the future path of AI are well-directed.

Approaching AI as a political concept, and focusing on the public pronouncements made by influential tech commentators, we identify four components of key contemporary discourses on AI and subject them to critical and technical scrutiny. First, we look at how AI discourses often endorse anthropological commitments that create false equivalences between human and artificial intelligence and suggest that all are equally affected by AI. Second, we show how AI discourses also ascribe agency to machines while concealing the role of humans in the development and management of AI. Third, we discuss how the economic assumptions made by these discourses support specific political interests. Fourth, we demonstrate how discourses on AI endorse temporal assumptions that reduce the space for democratic intervention. Taken together, these discourses indicate that AI is becoming more than what its 'technical' specifications would warrant; however, this is happening in a way that limits the space for democratic engagement with, and control of, the technology itself.

The paper proceeds as follows. We present a brief history of AI, then outline our approach to AI as a political concept and explain our analytical strategy. We then discuss the themes and assumptions we identify. The conclusion summarises our findings.

ii. From technical to political concept: a brief history of AI

The story of AI can be told as a technical history of the milestones that made AI what it is today. The emergence of AI is linked to the birth of computation and the convergence, between the 1940s and 1960s, of mathematics, psychology, engineering, and the emerging field of computer science. Two primary methodologies surfaced in the realm of AI research in those years: symbolic reasoning and connectionism. The symbolic approach focused on rule-based systems, logical inference, and high-level representations of knowledge (Smolensky, 1987). Connectionism, inspired by the structure and function of the biological brain, gave rise to artificial neural networks (ANN), the technology behind most contemporary AI systems.

ANN was initially only one branch of AI, and for decades it was not clear that it would be the most promising structure for advanced systems. The rise of the personal computer and growing computational power sparked a renewed interest in AI, and in the 1990s, the field transitioned from theoretical research to practical applications. Practices of performance benchmarking made it possible to establish that neural networks initially, and subsequently transformers, a neural network architecture introduced in 2017 that uses attention mechanisms to capture long-range dependencies, outperformed other architectures. These technologies became dominant and nowadays AI models are widespread in everyday applications.

There is, however, a parallel political history (or histories – Ali et al., 2023) of AI in which this technological development became the subject of political debate. Early pronouncements on AI that thought of it as either a moral or political force were embedded within the context

of cybernetics (Wiener, 1948, see also 1989). These attempts at thinking of AI were laden with the technology's organising and predicting capacities. In industrial and military contexts, in the 1950s and 1960s, these may have been framed in terms of developing hitherto only dreamt of power to optimise the performance of a system, attain procedural efficiency, or process ever greater amounts of data. In the West, the political framing was very much attached to gaining and maintaining a competitive advantage over rival powers. It was this instrumental, utilitarian, use of technology that gave rise to concerns linked to the loss of compassion and wisdom in decision-making (e.g., Weizenbaum, 1985), but these early considerations remained confined to a specialist field.

Technical developments in the 1980s led to thinking of the social and political implications of intelligent machines at a larger scale. Like other technological innovations before it, AI became gradually dislodged from the logic of a (exclusively) technical field and began increasingly being framed within the logic of a political field. From authors expressing concerns over policy impact in the field of machine learning and the role of expertise in the sociology of science (Courtial and Law, 1989; Shannon, 1948; Woolgar, 1985), through the 1990s and 2000s AI gained increasing attention by political agents.

iii. Approaching AI as a political concept

The existence of parallel histories of AI suggests that AI can be studied as both a technical and political concept. Political concepts are 'ideas that inject order and meaning into observed or anticipated sets of political phenomena and hold together an assortment of related notions' (Freeden, 1998: 52). They may be used to describe an empirical reality or a normative end-state – although their correct use will usually be 'open to dispute', with disagreements reflecting 'divergent normative, theoretical and empirical assumptions' (Bellamy and Mason, 2003: 2; see also Gallie, 1955).

Political concepts are also mobilising concepts (‘*Bewegungsbegriffe*’) seeking to temporalise history or conceptualise historical movement (Koselleck, 2004: 251). These concepts emerged with Modernity and the concomitant shift to conceiving of history in transitional terms, i.e., as a passage from a known past (experience) to an unknown future (expectation) (Koselleck, 2004: 241). This shift generated a dynamic where the newly conceived indeterminateness of the future became open for contestation. Time, and with it, the future, began appearing as domains of contestation (Koselleck, 2004: 248). In this context, mobilising concepts emerged as future-oriented ideas which made it possible to occupy that futural ‘space’ and mobilise towards it. Conceived as mobilising concepts, then, political concepts do not only seek to describe a certain future - they also attempt to bring it about.

Approaching AI as a political concept seems both appropriate and useful. Appropriate because AI conforms to most of the expectations we have for political concepts. The term ‘Artificial Intelligence’ holds together several associated ideas, within contested boundaries.[1] The notion of AI has been flexible (Mager and Katzenbach, 2021), with its meaning ranging from narrow generative algorithms (Weizenbaum, 1985) to ‘Artificial General Intelligence’ (AGI), a hypothetical AI system capable of performing any intellectual task that a human can, exhibiting general reasoning and adaptability across various domains (Bostrom, 2016). Additionally, AI increasingly appears to have a mobilising aspect to it – as the vivid debates between enthusiasts, catastrophists, and critical observers attest to. These discourses are not just relating a current situation, they are also creating future projections, sometimes loosely attached, others heavily detached, from the space of experience. AI, in this sense, is not just a describing concept (*Begriff*), but also an anticipation (*Vorgriff*) promising more than experience warrants (Koselleck, 2004: 252).

[1] In this paper we do not take a stance concerning the most appropriate definition of AI. Instead, we pick up on its malleability as permitting its appropriation by discourses of various stripes.

It is useful – desirable, even – because approaching AI as a political concept helps bring to the fore the political implications of the claims that are made about it. Acknowledging that AI is political in the sense of being contested and mobilising means understanding that any definition of it has implications for how politics is conducted. In the political space (understood here as the realm of collective decision making), how issues, facts, and options are presented and discussed shapes how decisions are made by precluding some future courses of action and facilitating others. The question is what kinds of politics are being implied and mobilised towards. Studying AI discourses through a lens that links them to the broader horizon of a temporalised environment allows us to think of 'AI' not just as a descriptor, but also as an instrument for steering the direction of historico-political development.

In sum, when we think of AI as a political concept, we mean it is political in three related ways: it is political because it is contested, because it is mobilising, and because it has implications for how we do politics. Current processes of politicisation of AI reflect this political nature: different actors are seeking to define its proper meaning and mobilise towards the future they desire. Bringing attention to both the contested and mobilising aspects of current discourses on AI makes it possible to consider what kind of future politics they imply. Although the implications of these discourses may manifest in different areas of politics, in the remainder of the paper, we focus on the realm of democratic decision-making.

iv. Themes and assumptions in contemporary discourses on AI.

To study AI as a political concept, and identify the ideas it holds together and the politics it preconises, we analyse a selection of public pronouncements on AI, identify their recurring themes, and the assumptions that underlie them. We then subject these claims to technical and critical scrutiny to identify their political implications.

To perform this analysis, we rely on the close reading of a purposive sample of texts on AI targeted primarily (but not exclusively) to generalist audiences and published by prominent tech commentators in the year following ChatGPT's release. We select this as a starting point since ChatGPT's release prompted a renewed focus on AI. We also refer to some texts published before this period whose reflections have been directly cited in current discussions.

We focus on tech commentators (including computer scientists and developers, but also VC funders, economists, and journalists) because their pronouncements are underlain by a level of authority which provides them with greater weight in public conversation. Although their authors would be unlikely to consider these texts 'political', we approach them as mobilising texts reflecting a certain worldview. Our sample consists of 31 English-language texts in total. These include blog posts, interviews, manifestoes, articles, and papers (full list in appendix). The selected texts were chosen to reflect different positions in debates on AI. Some are by 'enthusiasts' supportive of AI (n=13), others by 'catastrophists' who express deeply negative views on AI's future (n=10), and others present multiple views (n=3). We also include texts by 'critical voices' problematising the positions developed by enthusiasts and catastrophists (n=5). We acknowledge that such a stylised division creates artificial boundaries between positions that may be more nuanced or similar in outlook than the terminology reflects. However, we are interested in how, collectively, these actors are employing 'AI' as a political concept, often by responding to each other from different perspectives. These texts were also widely reported on or referenced by those taking part in public debates on AI, meaning they were most likely to be influential in shaping public understandings of AI.

While we consider this sample to offer a good representation of key debates on AI, it still has limitations. It covers only a selection of debates happening within a highly visible, but limited, field. Given many of the texts we analyse are short and aimed at generalist audiences, they also do not necessarily convey some of the more nuanced takes on AI. These limitations

notwithstanding, we consider our analysis to provide meaningful insights into the question of how AI is being construed. Given the dominance of Silicon Valley companies in AI development, seeing how AI is being portrayed by those who are part of that scene, or respond to its developments, shows how some very influential actors discuss AI. The fact that these debates present simplified information does not prevent them from shaping how AI is thought of outside a specialist realm, and particularly by the general publics and policymakers that would, ideally, be involved in shaping the future of AI.

When reading these texts, we first identified key themes through a close reading of the texts themselves and then subjected them to technical and critical scrutiny. For the technical scrutiny part, we relied on findings from AI research. While, given the interdisciplinary nature of this paper, our discussion cannot do full justice to the debates in a rapidly evolving field, it still enables us to identify areas where there is a discrepancy between what is being claimed about AI in our texts, and existing scientific knowledge. This enables us to pinpoint claims that would seem to fall into the realm of 'expectation' rather than 'experience'. We then study the same claims through critical discourse analysis. Critical discourse analysis studies how discourses, intended as 'relatively stable uses of language serving the organization and structuring of social life' (Wodak and Meyer, 2016: 6) shape the interpretation of the world around us, and to what effects (Power et al., 2019). Our primary focus is on the political implications of how AI is being construed. We complement this analysis of primary sources with insights from secondary literature in critical AI studies.

We present the findings from our analysis below, addressing anthropological, agential, economic, and temporal themes. Although this selection is not exhaustive, we consider these themes require the closest critical scrutiny because of their implications for democratic deliberation. The themes are also not discrete: while analytically separate, there are clear connections between them and areas where they mix or overlap.

a) Anthropological themes

A first set of themes and assumptions we identify pertains to AI's relation to human intelligence and the nature of humanity. Because they reflect beliefs concerning what 'people' are, we dub these 'anthropological' themes.

The nature and comparability of human and artificial intelligence is a first theme we identify. It is built on the understanding that AI systems are getting closer to matching human capabilities in certain areas (and at the extreme in *all* areas), and even surpassing them in other areas. For example, in a 2023 contribution, leading computer scientists discuss how human and 'narrow' artificial intelligence scale up, highlighting how

> AI has already surpassed human abilities in narrow domains […]. Compared to humans, AI systems can act faster, absorb more knowledge, and communicate at a far higher bandwidth. Additionally, they can be scaled to use immense computational resources and can be replicated by the millions (Bengio et al., 2023).

Moustafa Suleyman, the CEO of Microsoft AI, made a similar (but more sweeping) comparison when claiming that AI models 'clearly aren't biological in any traditional sense' but already behave like humans in some respects: 'they communicate in our languages. They see what we see. They consume unimaginably large amounts of information. They have memory. They have personality. They have creativity. They can even reason to some extent and formulate rudimentary plans. They can act autonomously if we allow them' (Suleyman, 2024). Although these actors are making different claims, both are effectively comparing human and artificial intelligence.

Those comparing human and artificial intelligence are making two anthropological assumptions concerning the nature of intelligence. The first is that intelligence is something

that can be measured and compared, a frequently problematised theme in critical AI studies (Ballatore and Natale, 2023). The second is that human intelligence can be replicated by adequately developed machines. As Bill Gates optimistically put it,

> Once developers can generalize a learning algorithm and run it at the speed of a computer […] we'll have an incredibly powerful AGI. It will be able to do everything that a human brain can, but without any practical limits on the size of its memory or the speed at which it operates (Gates, 2023).

While we do not endorse the idea that human and artificial intelligence are (or should be) compared, we do not reject it either. Even admitting such comparisons might be possible, however, there are still ways they do not seem to match up. Though experts acknowledge their differences (Bellec et al., 2020; Beniaguev et al., 2021; Lillicrap et al., 2020), parallels between AI and human cognition often overlook core differences in how each learns and applies information to perform tasks. Contrary to popular belief, LLMs like ChatGPT, do not 'learn' continuously. Rather, once they are trained, they operate with fixed weights, meaning they do not adapt or learn from interactions.[2] Even when it might look like a model is learning from past interactions (for example, in conversations with ChatGPT asking it to recall past information) it is processing each input within its pre-defined context window. This creates an illusion of learning, but the model's capabilities are bound by this context size or how many words can be added to the input. Once the conversation surpasses this limit, the model's limitations become apparent, as it cannot recall or build upon past interactions. This speaks to the broader issue of 'catastrophic forgetting' in ANNs, which impedes their ability to

[2] This applies to most popular LLMs, such as ChatGPT, which do not update their parameters during normal use. Although some specialised AI systems can continue learning (e.g., through online or incremental training), they are not typically deployed in large-scale consumer applications.

incorporate new information without losing what was previously learned. The static nature of AI models, in simple terms, means that every single interaction is forgotten as soon as the conversation ends – a fundamental difference when compared to humans' ability to continually acquire new information without forgetting everything else (Parisi et al., 2019).

Artificial and human intelligence also differ in their ability to act in dynamic, embodied settings. While an AI model may be well-suited to deal with certain tasks, its ability to tackle problems is linked to the availability and quality of relevant data. AI models currently rely on digitised forms of data such as audio, text, and visual inputs. However, there is a pronounced gap in AI's cross-modal capabilities, especially in complex areas like sensory modalities (olfaction, taste) and many other data-poor domains. In the realm of robotics, this challenge also involves the control problem. Operating in the real world forces robots to meet hard real-time control requirements, process high-bandwidth multimodal sensor streams on-board, and remain robust in dynamic, non-stationary environments—capacities current AI and deep learning methods still struggle to achieve end-to-end (Zuffer et al., 2025), whereas for humans this kind of multi-sensory control is routine—like the simple act of moving a chair. Even seemingly straightforward tasks such as self-localization illustrate the difficulty: robots remain prone to drift, typically require carefully calibrated multi-sensor setups, and rely on brittle sensor-fusion pipelines. Present-day AI methods, including emerging visual–language models, do not yet solve this issue end-to-end (Yarovoi and Cho, 2024; Yu et al., 2025). These technical limitations of AI suggest that, for now at least, machines are far from being able to replicate human intelligence. AI might be able to replicate some aspects of human intelligence, but claims beyond that are a leap of faith.

Resisting the temptation to oversell AI's 'intelligence' and its 'human-like' qualities is important from a political standpoint. Not only does comparing human and machine intelligence enable dubious claims that a 'superhuman intelligence' can exist based merely on

benchmark performance rather than a comprehensive understanding of intelligence; it also confounds what humans and computers can do (Collins, 2018: 93), creating a false equivalence between human decision-making and computer decision-making. This false equivalence is likely to be facilitated by claims that oversell AI's capacity, and can become the basis for dubious practices, such as the replacement of (fallible) human judgement with (allegedly superior) machine judgement. As existing works on algorithmic bias have shown, substituting human with machine judgement can have nefarious consequences, but because decisions have been taken by machines, there is little accountability (Eubanks, 2019; Noble, 2018; O'Neil, 2017). This problem would not necessarily be solved by more accurate AI decision-making, as it remains unclear who should be held responsible when tasks delegated to AI go wrong. Therefore, while both human and artificial intelligence may be imperfect, only the former comes with ways of ensuring accountability, a point we will return to in the next section.

The second anthropological theme to emerge from our analysis is the idea, widespread amongst 'enthusiasts' and 'catastrophists', that AI is an issue that affects and offers opportunities to humanity *as a whole* – be it through claims that 'Humanity can enjoy a flourishing future with AI' (Future of Life Institute, 2023), that AI will 'save the world' (Andreessen, 2023b) or that 'Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war' (Centre for AI safety, 2023). Underlying these themes is the assumption that AI is not an issue that benefits or harms a select few. Rather, it conjures an abstract 'humanity' that is in this together and in equal measure.

Apart from the obvious anthropocentrism of such a view, conceiving of AI in these terms is politically problematic because it conceals how different groups are affected by AI, a point that is noted more consistently by critical voices in our texts (e.g., Distributed AI Research Institute, 2023). From a politico-economic standpoint, some jobs are more at risk than others

(Cazzaniga et al., 2024), and AI also creates new categories of precarious and marginalised workers, including the 'human coders' whose work guarantees its functioning (Tubaro et al., 2020). Certain social groups are more at risk of being discriminated against by AI because it is usually marginalised communities that are on the receiving end of its negative implications (Gebru and Torres, 2024). Finally, AI developments also entrench global inequalities because the menial labour required to power AI models is performed in poorer countries, but the economic benefits are reaped primarily in rich countries (Muldoon and Wu, 2023; Tacheva and Ramasubramanian, 2023).

Speaking of a 'common humanity' conceals these differences, with stark political implications. The category of 'humanity' creates a veneer of inclusiveness while reproducing hierarchies of belonging where those at the top of the hierarchy have more rights and entitlements than those at the bottom of (or excluded from) it (Braidotti, 2020). Speaking of a common humanity does little, therefore, to centre the concerns of the communities most negatively affected by AI. It also depoliticises the problem of who is most affected by AI, thereby hampering efforts at political organisation. Identifying a constituency is essential for the articulation of political problems and solutions, as well as for political mobilisation (Bartolini and Mair, 1990). Suggesting that 'humanity' is in this together minimises political conflict, implying that all agree on the diagnosis and (potentially) on the solution. As a result, it makes it harder for specific (marginalised) groups to mobilise because it negates the very existence of the kind of divisions one might mobilise around.

b) Agential themes

A second, related theme we find is that of agency. Agential themes cover the role that people play(ed) in the past, present, and future of AI and on the role that AI had, has, and will have in relation to people, specifically when looking at decision-making.

Agential themes present conflicting views of the role of humans in the development of AI. On the one hand, in a widespread (but not pervasive — e.g., Bender et al., 2021) narrative, AI is portrayed as spoiling humans of their agency. AI is frequently presented, most obviously in catastrophist narratives, as self-developing and at constant risk of escaping human control – mirroring Goethe's Sorcerer's Apprentice parable. Nick Bostrom presents a good example of this theme when claiming that

> While one might consider creating a physically confined genie […] it would be difficult to have much confidence in the security of any such physical containment method against a superintelligence equipped with versatile manipulators and construction materials (Bostrom, 2016: 143).

The 'statement on pausing giant AI experiments' makes a similar point when describing AI labs 'in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control' (Future of Life Institute, 2023). Underlying these themes is the assumption that AI can acquire agential properties akin to those of human beings.

At the same time, the development of AI also becomes an opportunity to reassert human agency and use that agency to direct AI's future. Mitigating risks while unlocking new benefits is the name of the game – a point made clearly by Sam Altman, co-founder and CEO of OpenAI, when in his declaration in front of Congress he claimed 'We take the risks of this technology very seriously […]. We believe that government and industry together can manage the risks so that we can all enjoy the tremendous potential' (Altman, 2023). Underlying these discourses is the idea that 'regulating', 'leveraging', or 'controlling' AI becomes a way to reinstate human agency, without, however, taking responsibility for creating potentially destructive AI in the first place.

From a technical perspective, ascribing 'agential' qualities to AI is problematic because it overstates its ability to generate truly novel content or generalise outside of training data. Establishing exactly how good AI is at solving problems is complicated, no less due to the difficulty in identifying appropriate benchmarks to measure it, but existing evidence suggests that AI models are still far from becoming human-like 'agents.' For instance, claims that GPT-4 passed the Turing Test, a test designed to assess whether machines can exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human, were found to be unfounded (Jones and Bergen, 2024). Similarly, while transformers excel at learning tasks similar to their pre-training data, their ability to generalise beyond that scope remains limited (Yadlowsky et al., 2023). For example, GPT-4's performance in solving competitive programming questions significantly drops when tested on problems posted after its training period (Roberts et al., 2023). This suggests that their perceived in-context learning abilities may be more a function of data coverage than inherent generalisability. LLMs also fail to demonstrate a robust understanding of learned associations, a phenomenon Berglund and his coauthors (2024) call the 'Reversal Curse.' If a model is trained on a statement such as 'A is B,' it will not automatically infer that 'B is A' in a new context. This failure highlights that LLMs learn superficial patterns rather than truly grasping the underlying relationship between concepts. Claims that AI is about to acquire human-like agency, by which we mean a capacity to act intentionally, would therefore seem to overstate AI's actual abilities.

This complex mix of machine agency and human reaction is politically problematic, because it construes agency without responsibility, and presents human agency and AI agency as antithetical. Regarding the first, the attribution of agency to AI places accountability out of reach, because it is unclear how far AI can be held responsible – or who should be held responsible when AI makes mistakes (Eubanks, 2019). This de-responsibilisation obfuscates the ways in which humans are deeply implicated in developing, training, and maintaining AI

systems up and running (Heffernan, 2019: 5; Jarrahi et al., 2022). As Lucia Rafanelli (2022: 4) powerfully put it, delegating tasks to AI does not remove power from human hands:

> Someone denied a job because a "sexist" AI system deems her resume inadequate is not subject "only" to the power of the AI. She is subject to the power of the people who wrote the AI system's code and the people who decided to use it to evaluate her resume.

Regarding the second problem, presenting human agency as purely reactive, and necessarily incompatible with that of AI, promotes a form of politics that leaves limited space for political choice. As Jascha Bareis and Christian Katzenbach note, attributing agency to a technology often reduces human agency to 'adaption, reaction, or mitigation' (2022: 867). Like other forms of reactive politics, it takes the political agenda as externally set and outside the realm of democratic deliberation: the technology is developing in a certain direction, and policymakers and citizens need to respond to it on its own terms. Although there might be some truth to the claim that the agenda is externally set, it should also be clear that it is not set by AI itself, but by those who develop these technologies. Simultaneously, the reduction of human agency to 'reaction' nullifies political choice because it removes the possibility of shaping developments (and subjecting those developments to political deliberation) rather than simply responding to them urgently (a point we will return to later).

c) Economic themes

A third set of themes concerns the economics of AI. Discourses on AI have tended to portray the development of the technology as essentially linked to a specific conception of what economic reality is and, perhaps more importantly, to a specific conception of what that reality ought to be. The economic themes, perhaps more so than the others, play out differently in the case of catastrophists and enthusiasts, and it would be a mistake to conflate these differentiable

elements. While the former have generally tended to favour forms of regulation and legislation amenable to their interests, the latter have tended to favour strong de-regulation and free-for-all competition. In what follows we hope to speak to both. However, we leave the technical limitations of AI to the side because the immediate concern here is the specific economic realities being pursued.

The desirability (indispensability) of market-driven competition for innovation is a prominent theme in contemporary discourses on AI, especially amongst enthusiasts. The thought here is that unrestrained, unregulated competition should mark the pace, a point well made by Mark Andreesen when he says that 'Big AI companies should be allowed to build AI as fast and aggressively as they can – but not allowed to achieve regulatory capture, not allowed to establish a government-protected cartel that is insulated from market competition due to incorrect claims of AI risk.' This will, he continues, 'maximize the technological and societal payoff from the amazing capabilities of these companies, which are jewels of modern capitalism' (Andreessen, 2023b).

This is tightly linked to claims on how increasingly efficient, perhaps infinitely so, the use of resources can become if in the hands of an artificially intelligent being. AI discourses often indulge in the further notion that quantifiable but limitless and exponential economic growth is a byproduct of the merits of the technology (and, while on the side of enthusiasts, that will be linked to the undesirability of regulation, on the side of catastrophists it will be linked to the desirability of a specific — benevolently perceived— form of regulation):

> None of those estimates should be taken to suggest that AI development will be anything less than hugely impressive over the next few decades. But as one set of constraints is relaxed — in this case access to intelligence — the remaining constraints will matter all the more. Regulatory delays will be more frustrating, for instance, as they

> will be holding back a greater amount of cognitive horsepower than in times past (Cowen, 2023b).

There is, lastly, a third theme pertaining to economic regulation. Taking inspiration from some form of knowledge-based economics (e.g., Hayek, 1969: 84), resistance toward planning and centralisation in enthusiasts' discourses takes the form of an antipathy toward centralised regulation that would preclude its 'natural' (i.e. technical) development. On the catastrophist side, however, it is not so much an antipathy toward centralised regulation as much as a shifting attention away from possible labour distribution problems associated to the technology (Cole, 2023) by overemphasising existential threats. Often, on both sides, the absence of clear regulation, what a Goldman Sachs report referred to as the 'inter-AI years', is construed as a space of opportunity (Cohen et al., 2023; see also Cowen, 2023a; Mazzucato et al., 2022; Taeihagh, 2021).

These three economic themes identified above point toward a specific conception of what economic reality should be that is endorsed, sometimes explicitly, sometimes tacitly, by AI discourses, albeit differently for enthusiasts and catastrophists. For enthusiasts, that economic reality is one where free-market competition, 'effective' use of resources, and a decentralised form of decision-making are fetishised as inherently appealing to all. For catastrophists, that economic reality is one where regulation ought to be in place, but only to mitigate the existential threats, rather than the finer grained problems, posed by AI. But these themes, and especially the assumptions that undergird them concerning the desirability of a certain kind of economic order, present a series of problems. That competition-driven markets should act as the benchmark against which success/legality/rightfulness ought to be measured, should not only be fundamentally questionable; it should give us reason to pause at its implications. For while it is almost a truism that a free-for-all order incentivises inventiveness, it is also true that not all invention should be uncritically welcome (Stirling, 2017). In the specific context in

question, even if one were to recognise the important advancements made by international corporations, and further recognise the (now deeper, now shallower) transformative effects AI might have on different economic settings, some modicum of criticality would seem to suggest that legal and moral limits ought to be in place.

Likewise, that a very specific (neoliberal) conception of what constitutes 'good/responsible economic practice' is set in stone, and the laws of which one should abide by forever, is problematic to the extent that economic practice and theory, like everything else, are subject to change over time. While this is not the place to construct another critique of political economy, it is nevertheless worth remarking that the unquestioned pursuit of a hegemonic, neoliberal economic order might give rise to an increasing entrenchment of inequalities, as literature in the field has shown (Bender et al., 2021). Be it from the standpoint of what the scholarship has designated mainstream economics or heterodox economics (Lawson, 2005), no amount of modelling can do away with the dynamism that pertains to economic systems. Irrespective of whether through a price mechanism or some other way, one should be careful not to simply endorse trends predicated on (often mistakenly) perceived market tendencies or a supposed increasing efficiency in the use of resources. Being able to avoid drawing policy 'solutions' ensconced in what is currently perceived as desirable economic practice, especially if done so overly hastily, is a consequential capacity that should not be simply given up.

The economic assumptions, along with the problems they give rise to, are worth remarking not least because of the effects they have had in policy making (Ulnicane, 2022). While valuable attempts are being put forth at thinking through some of the economic implications of AI and their links to economic policy (for a recent example, see Acemoğlu, 2023), these attempts risk being sidelined, if not altogether obscured, by more extravagant economic claims. When 'the most destructive force in history' ensures that 'a time will come when there will be no jobs' (*Elon Musk & Rishi Sunak Interview*, 2023), one is tempted to ignore the more

meticulous work being carried out by economists in favour of unfounded large-scale predictions. It may well be worth pondering the extent to which the more extravagant claims, rather than merely thinking through the economic dimension of AI, naturalise a contingent conception of what economic reality ought to be.

d) Temporal themes

A final, but crucial, set of themes concerns the temporality of AI, pointing towards a sense of acceleration, urgency, and speed. Yoshua Bengio captured this sentiment when, cited in an interview, he stated 'I didn't pay too much attention to the longer-term danger. I thought it was too far off into the future. In the last few months, it's dawned on me and many others that things have accelerated unexpectedly' (Whitworth, 2023). Others have provided confident accelerated timelines for the development of AGI. Sam Altman predicted that it could be achieved in 4-5 years, while Dario Amodei, the CEO of Anthropic AI, claimed he expected a human level AGI in 2-3 years (Henshall, 2024).

In this ever-shortening timeline, AI is presented as a force in motion, which is already changing societies, and will do so even more radically in the future. For catastrophists, the future is approaching fast. For enthusiasts, it is approaching fast – but could be faster. For critical voices, a problem-ridden AI future is already here. Underlying this theme is an assumption about the nature of time as linear, ineluctable, and cumulative: developments are bringing us in a clear direction and doing so in a linear fashion. Further, while critical voices tend to see AI as the continuation of existing practices of discrimination and exploitation by other means (Bender et al., 2021), catastrophists and enthusiasts present these changes as somehow 'unprecedented', as if AI represented a clear and undisputed break in technological development that leads in a singular direction – enthusiastically, salvation; catastrophically, damnation.

The claims of sudden acceleration and unprecedentedness prepare the ground for demands to act urgently to speed AI development up or halt it in its tracks. For enthusiasts, it is a matter of unleashing the potential of AI. For catastrophists, the priority is to avoid catastrophe because 'We are not ready. We are not on track to be significantly readier in the foreseeable future. If we go ahead on this everyone will die' (Yudkowsky, 2023). Even the critical voices focusing on narrower time horizons argue that 'it is indeed time to act: but the focus of our concern should not be imaginary "powerful digital minds." Instead, we should focus on the very real and very present exploitative practices of the companies claiming to build them' (Gebru et al., 2023; see also Aradau and Bunz, 2022). Underlying this theme is the assumption that the correct response to fast-moving change is immediate action.

Many of these claims are not fully warranted from a technical standpoint because they overstate the speed at which AI is developing, and create narratives of certainty where scientific knowledge is uncertain. Current models still struggle with handling problems that significantly differ from what has already been seen, digitised and solved, and much of the anticipatory language that is being used to describe future AI conceals significant uncertainty about when and how these challenges will be addressed. Proponents of an accelerated development timeline like Altman or Suleyman often rely on confirmation bias, drawing from anecdotal evidence, rather than on sound research findings. For example, in March 2024 a demo appeared to showcase a new AI model called Devin that was declared to be a 'quantum leap advance' in solving software engineering problems (Vance, 2024). Only a month later, serious allegations appeared that the entire demo was faked (Wang, 2024).

One will occasionally hear the argument that some of the existing problems with AI models will be solved through 'scaling', that is, augmenting the size of models (e.g., de Freitas, 2022). Reality, however, presents a less optimistic picture. Large models such as GPT-4 may be better able to memorise information (Carlini et al., 2023), but still exhibit deficiencies in fundamental

tasks such as arithmetic operations and data organisation (Arkoudas, 2023). Furthermore, scaling exacerbates the strain on computational resources, necessitating either further optimisation or an expansion in the availability of computational power. These limitations raise questions concerning how easy it will be to develop the kind of AGI on the mind of catastrophists and enthusiasts alike.

Given existing technical limitations, the confident timelines of catastrophists and enthusiasts, along with their calls for sweeping urgent action, often lack a solid technical foundation. Rather, these claims operate on a political level. Exploiting the widening gap between what experience warrants and what expectations promise (Koselleck, 1989: 300), these discourses on AI inscribe themselves in a lineage of discourses that prescribe what the future ought to be and legitimise those who claim to know what it will be like (Jung, 2014; Winner, 1998).

Positing a linear, and nearly completely certain, temporal structure is equally problematic insofar as the end (in both senses of *telos* and *terminus*) is posited as a necessary outcome of temporal succession. The eschatological echoes of such discourses have been noted before by Susi Geiger (2020) as efforts at resolving ontological uncertainty through discursive inevitability. The seemingly opposite strategy of positing a future that is radically different from the present and the past confirms, rather than denies, the assumption of a singular timeline that seemingly 'breaks' altogether. Temporal discontinuities allow those that posit them to no longer have to legitimise their predictions on experience, and instead allow them to find legitimacy in mere imaginative expectations (Bareis and Katzenbach, 2022). Additionally, operating under the assumption of a common temporal horizon that appears as constantly accelerating is part of a construction of a univocal imminent future that flattens any nuance in the range of possible outcomes of social and political becoming.

Discourses on AI are obviously not unprecedented in co-opting the future to serve their own agendas. It should nonetheless be concerning to witness discourses wielding the combined power of ineluctability, disruption, and acceleration, all predicated on an assumed linear temporal structure, in contexts where the possibilities of democratic deliberation and intervention presuppose electability, continuity, and duration. It is in the scrutiny of the temporal themes that contemporary AI discourses show themselves most flagrantly at odds with many of our contemporary political paradigms. The reduction of the future to one where AI will take over –a deterministic view shared by catastrophists and enthusiasts alike (Benjamin, 2024), but not by critical voices– suggests that only one future is possible. From a democratic standpoint, a future that is one to the exclusion of any other will generate tension with what it means to choose for oneself. A *radically* different future, likewise, will sit ill at ease with traditional ways of understanding the role of political institutions, and so does the erosion of lasting or stable mechanisms or spaces for political deliberation.

The appeal to urgency exacerbates these tendencies by further reducing the space to imagine or bring about alternative futures, while limiting the role of coordinated democratic action in the present. Calls to urgent action reduce the space for 'slow' decision processes and for democratic choice itself because in an emergency situation, 'the unspoken presumption is that either one can think or one can act, and given that it is absolutely mandatory that an action be performed, thinking must fall away' (Scarry, 2012: 7). They also concentrate power into the hands of a group of people identified as best able to respond to the emergency (White, 2024: 162). AI discourses pointing to the need to urgently address the problems/opportunities that AI produces have a similar effect. They reduce the space to collectively define what the problem that needs addressing is, and how it should be addressed. The time of democracy and of politics is too slow for the kind of urgency required to forestall or bring the AI future about. Additionally, particularly when the calls come from industry, they automatically suggest a

series of actors best placed to respond to the emergency – the tech companies themselves, perhaps in concert with compliant policy-makers, but certainly not citizens. As Ruha Benjamin (2024) put it, 'if AI evangelists can convince us that AGI is possible, imminent, *and* dangerous, we might be compelled to entrust our fate to them'. Buying into the narrative that AI is an issue in need of urgent regulation (by experts) risks feeding into a process whereby a crucial development is taken out of the realm of the politically contestable.

Criticising does not mean that inaction is preferrable. Instead, it should be seen as a reminder that what action looks like, and what the future looks like, should be subject to collective determination. A world with AI is a possibility, but it is not inevitable (Crawford, 2021), and even defining what a future with AI looks like should not be left solely in the hands of actors with a vested interest in a future with AI (Tacheva and Ramasubramanian, 2023).

vi. Conclusion

In this paper, we set out to analyse the how AI is being construed in public debates, and to what political effects. We began by advancing the claim that AI can be approached as a political concept, operationalised by different discursive sets. Our analysis showed how AI discourses endorse a problematic anthropological picture, presuppose a questionable understanding of what agency is, perpetuate a view of capital accumulation based on an overly narrow view of market logics and legitimise a conception of time, development, and intervention that is fundamentally at odds with genuinely democratic decision-making processes. Table 1 sums up our findings.

Table 1 Summary of themes, assumptions and stress points

Topic	Themes	Assumptions	Stress points
Anthropology	• Comparability of human and artificial intelligence. • AI issue affects 'humanity.'	• Human intelligence is replicable. • A monolithic humanity exists and is equally affected by AI.	• Overstates similarities between human and artificial intelligence. • Facilitates false equivalences. • Depoliticises AI and hampers political organisation.
Agency	• 'Out of control' AI. • Reassertion of human agency.	• AI can have agency. • Humans can respond to AI developments but cannot be held responsible for them.	• Overstates AI's ability to generate novel content or generalise. • Creates agency without responsibility. • Limits space for political engagement by accepting externally-set agenda.
Economics	• Market-driven innovation. • Ever-increasing 'efficiency' in use of resources. • Decentralised planning.	• Desirability of competition-driven markets. • Reification of economic modelling.	• Eschews other criteria for ascertaining desirability of AI. • Predicated on questionable forms of accumulation. • Diverts policy attention to far-fetched economic scenarios.
Time	• Accelerating pace of AI development. • Urgent action required to respond to AI.	• Linear progress of time towards a defined but unprecedented end. • Unprecedented developments require urgent responses.	• Creates technological certainty where there is uncertainty. • Flattens the future. • Restricts democratic action and empowers tech-nocratic elites.

We are left with what at first glance might seem like an oxymoron: the discursive use of AI as a depoliticising political mechanism. Contemporary uses of 'artificial intelligence' simultaneously politicise the technology and arrest political engagement with it. Even when

explicitly claiming otherwise (Andreessen, 2023a), pronouncements on AI are political moves that interrupt political participation, sequestering decisions of profound political relevance, and confining those decisions to the hands of self-appointed adjudicators. Koselleck had noticed this feature of mobilising political concepts: they both make a political use of the word and depoliticise its meaning. We have shown that feature to be there in 'artificial intelligence'. Discourses on AI at the same time lay a claim toward responsibility for the advancement of the technology and relieve the claiming actors from any responsibility around the technological tool's employment. Indeed, '[a]ll [political] notions aim at an irreversible process that imposes responsibility on the actors while simultaneously relieving them of it because the self-generation of the promised future is included in the notion' (Koselleck, 1989: 302). Moreover, this self-generation legitimises the use of a political concept through the use itself, and haunts 'AI' as much as it does any other political concept. Much of the use of 'artificial intelligence' answers not so much to technical or historical development, i.e. experience, but much more to visions of what an anticipated future should look like, i.e. expectation. Because the concept 'artificial intelligence' is mostly used with a future-inflection, it analytically includes the future it claims to merely describe, and thereby legitimises itself, that is, prescribes, the very reality it denotes while neutralising contestation.

None of this is to say that AI will not pose very real political, economic, and social challenges going forwards. However, one should be mindful when considering the claims that are being made about it and who is making them. To quote Emily Bender, 'we are not saying that AI is hype. We are saying that your claims about AI are hype' (Mance, 2024). Maintaining a critical attitude is part of making sure that control over AI's future developments is not misplaced, hence why we consider this paper's contribution to be timely and relevant. In a letter to Carl Schmitt dated 5th of November 1954, precisely in the context of a discussion about technology's presumed autonomy (and how it should not obscure that power remains in

people's hands), Koselleck himself wondered 'who really rules?' (Koselleck and Schmitt, 2019: 67). The question in our context remains pertinent. Do discourses on artificial intelligence, especially those delivered by people in positions of power, promote informed dialogue and deliberation or do they constrain political participation and limit decision-making to a merely reactive endeavour? The argument developed here indicates the latter.

References

Acemoğlu D (2023) Harms of AI. In: Bullock JB, Chen Y-C, Himmelreich J, et al. (eds) *The Oxford Handbook of AI Governance*. Oxford University Press. Available at: https://doi.org/10.1093/oxfordhb/9780197579329.013.65 (accessed 15 May 2024).

Ali SM, Dick S, Dillon S, et al. (2023) Histories of artificial intelligence: a genealogy of power. *BJHS Themes* 8. 2023/12/22 edn. Cambridge University Press: 1–18.

Altman S (2023) *Written Testimony of Sam Altman Chief Executive Officer OpenAI Before the U.S. Senate Committee on the Judiciary Subcommittee on Privacy, Technology, & the Law*. 15 May. Available at: https://www.judiciary.senate.gov/imo/media/doc/2023-05-16%20-%20Bio%20&%20Testimony%20-%20Altman.pdf.

Andreessen M (2023a) The Techno-Optimist Manifesto. In: *A16Z*. Available at: https://a16z.com/the-techno-optimist-manifesto/.

Andreessen M (2023b) Why AI Will Save the World. In: *A16Z*. Available at: https://a16z.com/ai-will-save-the-world/.

Aradau C and Bunz M (2022) Dismantling the apparatus of domination? Left critiques of AI. *Radical Philosophy* 2(12): 10–18.

Arkoudas K (2023) GPT-4 Can't Reason. *arXiv [cs.CL]*. Epub ahead of print 2023.

Ballatore A and Natale S (2023) Technological failures, controversies and the myth of AI. In: *Handbook of Critical Studies of Artificial Intelligence*. Cheltenham: Edward Elgar, pp. 237–244.

Bareis J and Katzenbach C (2022) Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics. *Science, Technology, & Human Values* 47(5). SAGE Publications Inc: 855–881.

Bartolini S and Mair P (1990) *Identity, Competition and Electoral Availability : The Stabilisation of European Electorates 1885-1985*. Cambridge: Cambridge : Cambridge University Press.

Bellamy R and Mason A (2003) *Political Concepts*. Manchester: Manchester University Press.

Bellec G, Scherr F, Subramoney A, et al. (2020) A solution to the learning dilemma for recurrent networks of spiking neurons. *Nature Communications* 11(1): 1–15.

Bender EM (2023) Policy makers: Please don't fall for the distractions of #AIhype. In: *Medium*. Available at: https://medium.com/@emilymenonbender/policy-makers-please-dont-fall-for-the-distractions-of-aihype-e03fa80ddbf1.

Bender EM, Gebru T, McMillan-Major A, et al. (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. In: *FAccT '21*, New York, NY, USA, 2021, pp. 610–623. Association for Computing Machinery. Available at: https://doi.org/10.1145/3442188.3445922.

Bengio Y, Hinton G, Yao A, et al. (2023) Managing AI Risks in an Era of Rapid Progress. *arXiv [cs.CY]*. Epub ahead of print 2023.

Beniaguev D, Segev I and London M (2021) Single Cortical Neurons as Deep Artificial Neural Networks. *Neuron* 109(17): 2727–2739.

Benjamin R (2024) The New Artificial Intelligentsia. *Los Angeles Review of Books*, 18 October. Available at: https://lareviewofbooks.org/article/the-new-artificial-intelligentsia/.

Berglund L, Tong M, Kaufmann M, et al. (2024) The Reversal Curse: LLMs trained on 'A is B' fail to learn 'B is A'. *arXiv [cs.CL]*. Epub ahead of print 2024.

Bostrom N (2016) *Superintelligence : Paths, Dangers, Strategies*. New York: Oxford University Press.

Braidotti R (2020) 'We' May Be in This Together, but We Are Not All Human and We Are Not One and the Same. *Ecocene: Cappadocia Journal of Environmental Humanities* 1(1): 26–31.

Carlini N, Ippolito D, Jagielski M, et al. (2023) Quantifying Memorization Across Neural Language Models. *arXiv [cs.LG]*. Epub ahead of print 2023.

Cazzaniga M, Jaumotte F, Li, Longji, et al. (2024) Gen-AI: Artificial Intelligence and the Future of Work. *Staff discussion notes*. Epub ahead of print 14 January 2024.

Centre for AI safety (2023) Statement on AI Risk. Available at: https://www.safe.ai/work/statement-on-ai-risk.

Coeckelbergh M (2022) *The Political Philosophy of AI: An Introduction*. Cambridge: Polity.

Cohen J, Lee G, Greenbaum L, et al. (2023) *The generative world order: AI, geopolitics, and power*. 14 December. Goldman Sachs. Available at: https://www.goldmansachs.com/intelligence/pages/the-generative-world-order-ai-geopolitics-and-power.html (accessed 15 May 2024).

Cole M (2023) (Infra)structural Discontinuity: Capital, Labour, and Technological Change. *Antipode* 55(2): 348–372.

Collins HM (Harry M) 1943- (2018) *Artifictional Intelligence: Against Humanity's Surrender to Computers*. Cambridge: Polity Press.

Courtial J-P and Law J (1989) A Co-Word Study of Artificial Intelligence. *Social Studies of Science* 19(2). SAGE Publications Ltd: 301–311.

Cowen T (2023a) New Laws to Regulate AI Would be Premature. *Bloomberg*, 30 October. Available at: https://news.bloomberglaw.com/artificial-intelligence/new-laws-to-regulate-ai-would-be-premature-tyler-cowen.

Cowen T (2023b) The economic impact of AI. In: *Marginal revolution*. Available at: https://marginalrevolution.com/marginalrevolution/2023/08/the-economic-impact-of-ai.html.

Crawford K (2021) *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. 1st edn New Haven: Yale University Press.

de Freitas N (2022) Tweet. Available at: https://twitter.com/NandoDF/status/1525397036325019649 (accessed 15 May 2024).

de Wilde P, Leupold A and Schmidtke H (2016) Introduction: the differentiated politicisation of European governance. *West European politics* 39(1). Routledge: 3–22.

Distributed AI Research Institute (2023) Statement from the listed authors of Stochastic Parrots on the "AI pause" letter. Available at: https://www.dair-institute.org/blog/letter-statement-March2023/.

Elon Musk & Rishi Sunak Interview (2023). Available at: https://www.youtube.com/watch?v=AjdVlmBjRCA.

Eubanks V (2019) *Automating Inequality*. New York: Picador.

Freeden M (1998) *Ideologies and Political Theory: A Conceptual Approach*. Oxford: Oxford University Press.

Future of Life Institute (2023) Pause Giant AI Experiments: An Open Letter. Available at: https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Gallie WB (1955) Essentially Contested Concepts. *Proceedings of the Aristotelian Society* 56: 167–198.

Gates B (2023) The Age of AI has begun. In: *Gates Notes*. Available at: https://www.gatesnotes.com/The-Age-of-AI-Has-Begun.

Gebru T and Torres EP (2024) The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence. *First Monday* 29(4).

Gebru T, Bender EM, McMillan-Major A, et al. (2023) Statement from the listed authors of Stochastic Parrots on the "AI pause" letter. In: *DAIR*. Available at: https://www.dair-institute.org/blog/letter-statement-March2023/.

Geiger S (2020) Silicon Valley, disruption, and the end of uncertainty. *Journal of Cultural Economy* 13(2). Routledge: 169–184.

Hayek FA von (Friedrich A (1969) The use of knowledge in society. In: *Individualism and Economic Order*. Chicago: University of Chicago Press.

Heffernan T (2019) *Cyborg Futures: Cross-Disciplinary Perspectives on Artificial Intelligence and Robotics*. 1st ed. 2019. Basingstoke: Palgrave MacMillan.

Henshall W (2024) When Might AI Outsmart Us? It Depends Who You Ask. *Time*, 19 January. Available at: https://time.com/6556168/when-ai-outsmart-humans/.

Jarrahi MH, Lutz C and Newlands G (2022) Artificial intelligence, human intelligence and hybrid intelligence based on mutual augmentation. *Big Data & Society* 9(2). SAGE Publications Ltd.

Jones CR and Bergen BK (2024) Does GPT-4 pass the Turing test? *arXiv [cs.AI]*. Epub ahead of print 2024.

Jung T (2014) The Politics of Time: Zeitgeist in Early Nineteenth-Century Political Discourse. *Contributions to the History of Concepts* 9(1): 24–49.

Kim J (2023) Traveling AI-essentialism and national AI strategies: A comparison between South Korea and France. *Review of Policy Research* 40(5): 705–728.

Koselleck R (1989) Time and Revolutionary Language. In: *The Public Realm*. Albany: SUNY Press.

Koselleck R (2004) *Futures Past: On the Semantics of Historical Time*. New ed. New York, Chichester: Columbia University Press.

Koselleck R and Schmitt C (2019) *Der Briefwechsel* (ed. JE Dunkhase). Berlin: Suhrkamp.

Lawson T (2005) The (Confused) State of Equilibrium Analysis in Modern Economics: An Explanation. *Journal of Post Keynesian Economics* 27(3). Taylor & Francis, Ltd.: 423–444.

Lillicrap TP, Santoro A, Marris L, et al. (2020) Backpropagation and the brain. *Nature Reviews Neuroscience* 21(6): 335–346.

Lindgren S (2023) *Handbook of Critical Studies of Artificial Intelligence*. Cheltenham: Edward Elgar Publishing.

Mager A and Katzenbach C (2021) Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified. *New Media & Society* 23(2): 223–236.

Mance H (2024) AI keeps going wrong. What if it can't be fixed? *Financial Times*, 6 April. Available at: https://www.ft.com/content/648228e7-11eb-4e1a-b0d5-e65a638e6135 (accessed 16 May 2024).

Mazzucato M, Schaake M, Krier S, et al. (2022) *Governing artificial intelligence in the public interest*. UCL Institute for Innovation and Public Purpose, Working Paper Series. Available at: https://www.ucl.ac.uk/bartlett/public-purpose/wp2022-12. (accessed 6 January 2025).

Muldoon J and Wu BA (2023) Artificial Intelligence in the Colonial Matrix of Power. *Philosophy & Technology* 36(4): 80.

Noble SU (2018) *Algorithms of Oppression How Search Engines Reinforce Racism*. New York: New York University Press.

O'Neil C (2017) *Weapons of Math Destruction : How Big Data Increases Inequality and Threatens Democracy*. London: Penguin Books.

Parisi GI, Kemker R, Part JL, et al. (2019) Continual lifelong learning with neural networks: A review. *Neural Networks* 113: 54–71.

Power K, Ali T and Lebdušková E (2019) *Discourse Analysis and Austerity : Critical Studies from Economics and Linguistics*. Abingdon: Routledge.

Rafanelli LM (2022) Justice, injustice, and artificial intelligence: Lessons from political theory and philosophy. *Big Data & Society* 9(1): 20539517221080676.

Richter V, Katzenbach C and Schäter MS (2023) Imaginaries of artificial intelligence. In: *Handbook of Critical Studies of Artificial Intelligence*. Cheltenham: Edward Elgar Publishing, pp. 209–223.

Roberts M, Thakur H, Herlihy C, et al. (2023) Data Contamination Through the Lens of Time. *arXiv [cs.CL]*. Epub ahead of print 2023.

Scarry E (2012) *Thinking in an Emergency*. New York ; London: W.W. Norton.

Shannon CE (1948) A Mathematical Theory of Communication. *Bell System Technical Journal* 27(3): 379–423.

Smolensky P (1987) Connectionist, symbolic, and the brain. *AI Review* 1.

Stirling A (2017) Precaution in the Governance of Technology. In: *The Oxford Handbook of Law, Regulation, and Technology*. Oxford: Oxford University Press, pp. 645–669.

Suleyman M (2024) What is an AI anyway? Available at: https://www.ted.com/talks/mustafa_suleyman_what_is_an_ai_anyway/transcript.

Tacheva J and Ramasubramanian S (2023) AI Empire: Unraveling the interlocking systems of oppression in generative AI's global order. *Big Data & Society* 10(2): 20539517231219241.

Taeihagh A (2021) Governance of artificial intelligence. *Policy and Society* 40(2): 137–157.

Tubaro P, Casilli AA and Coville M (2020) The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society* 7(1): 2053951720919776.

Ulnicane I (2022) Emerging technology for economic competitiveness or societal challenges? Framing purpose in Artificial Intelligence policy. *Global Public Policy and Governance* 2(3): 326–345.

Ulnicane I and Erkkilä T (2023) Politics and policy of Artificial Intelligence. *The Review of policy research* 40(5). Knoxville: Policy Studies Organization: 612–625.

van Lente H, Spitters C and Peine A (2013) Comparing technological hype cycles: Towards a theory. *Technological forecasting & social change* 80(8): 1615–1628.

Vance A (2024) Gold-Medalist Coders Build an AI That Can Do Their Job for Them. *Bloomberg*, 12 March. Available at: https://news.bloomberglaw.com/artificial-intelligence/gold-medalist-coders-build-an-ai-that-can-do-their-job-for-them (accessed 15 May 2024).

Wang B (2024) Devin AI Failure on One of Its Upwork Projects. Available at: https://www.nextbigfuture.com/2024/04/devin-ai-failure-on-one-of-its-upwork-projects.html (accessed 15 May 2024).

Weizenbaum Joseph (1985) *Computer Power and Human Reason : From Judgment to Calculation*. Harmondsworth: Pelican.

White J (2024) *In the Long Run : The Future as a Political Idea*. London: Profile Books.

Whitworth D (2023) Yoshua Bengio on how AI could cost us democracy. *The Times*, 2 June. Available at: https://www.thetimes.co.uk/article/yoshua-bengio-ai-safety-artificial-intelligence-x9mknfnr5.

Wiener N (1948) *Cybernetics; or, Control and Communication in the Animal and the Machine.* New York: J. Wiley. Available at: http://hdl.handle.net/2027/wu.89074767054.

Wiener N (1989) *The Human Use of Human Beings: Cybernetics and Society*. London: Free Association.

Williams R (1971) *Politics and Technology*. Studies in comparative politics. London: Macmillan.

Winner L (1980) Do Artifacts Have Politics? *Daedalus (Cambridge, Mass.)* 109(1). Boston: American Academy of Arts and Sciences: 121–136.

Winner L (1998) Prophets of Inevitability. *MIT Technology Review* 101(62).

Wodak R and Meyer M (2016) *Methods of Critical Discourse Studies.* Los Angeles: SAGE.

Woolgar S (1985) Why not a Sociology of Machines? The Case of Sociology and Artificial Intelligence. *Sociology* 19(4): 557–572.

Yadlowsky S, Doshi L and Tripuraneni N (2023) Pretraining Data Mixtures Enable Narrow Model Selection Capabilities in Transformer Models. *arXiv [cs.LG]*. Epub ahead of print 2023.

Yarovoi A and Cho YK (2024) Review of simultaneous localization and mapping (SLAM) for construction robotics applications. *Automation in Construction* 162: 105344.

Yu S, Chen Y, Ju H, et al. (2025) How Far are VLMs from Visual Spatial Intelligence? A Benchmark-Driven Perspective. *arXiv [cs.AI]*. Epub ahead of print 2025.

Yudkowsky E (2023) Pausing AI Developments Isn't Enough. We Need to Shut it All Down. *Time*, 29 March. Available at: https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/ (accessed 15 May 2024).

Zuffer A, Burke M and Harandi M (2025) Advancements and Challenges in Continual Reinforcement Learning: A Comprehensive Review. *arXiv [cs.LG]*. Epub ahead of print 2025.