AI & SOCIETY
https://doi.org/10.1007/s00146-025-02339-1

OPEN FORUM

Humanism strikes back? A posthumanist reckoning with 'self-development' and generative AI

Sam Cadman[1] · Claire Tanner[1] · Patrick Cheong-Iao Pang[2]

Received: 18 January 2025 / Accepted: 26 March 2025

Abstract
Since the release of OpenAI's ChatGPT in 2022, AI activity has reached a fever pitch. Calls for effective ethical responses to the pressurised AI environment have in turn abounded. Posthumanism, which seeks to build ethical futures by de-centring the 'human', is an obvious candidate to act as a lynchpin of theoretical intervention. In their responses, posthumanist scholars appear to have embraced AI's potential to destabilise Humanist philosophical ideas. We critically interrogate this initial enthusiasm. Conceptually distinguishing 'post-dualist self-development' (PDSD) from 'technical self-development' (TSD), we show how AI prompts an urgent need to advance posthumanist engagement with how technical development unsupervised by humans is ontologically discrete from other forms of material agency. We argue that specific engagement with TSD as distinct from PDSD is a key to avoid ignoring or underestimating Humanist and anthropocentric aspects of current AI innovation, and the influence of anthropomorphism. Without a theoretical reckoning with these tensions, posthumanism in the AI-era runs the risk of potentially promoting technologies that reinvigorate Humanist and anthropocentric expansion. To conclude, we show how a posthumanist ethics of generative AI that pays requisite attention to both TSD and PDSD may enable more anticipatory and nuanced assessments of the risks and benefits of discrete AI technologies to inform public discourse, appropriate social, institutional, policy and governance responses, and direct AI research and development priorities.

Keywords AI · Posthumanism · Humanism · Self-development · Large language models

1 Introduction

Contemporary posthumanist theories of technology place critical reliance on non-human entities including machines actively participating in their own development rather than being passive sites for development by humans. Since Haraway famously referred to machines as 'self-developing' (Haraway 2006, p. 120), analogous terms have included 'material agency' (Barad 2007, Footnote 43), 'thing-power' (Bennett 2010, p. 5), 'technological intentionality' (Verbeek 2011, p. 16), 'self-organising materiality' (Braidotti 2013, p. 82), technology with 'autonomous features' (Dahlin 2024, p. 62) or 'machinic autopoiesis' (Braidotti 2013, p. 94; Guattari 1995). Broader theoretical designations include 'post-dualism' (Escobar 2018) or new/neo-materialism (Bennett 2010; Blok 2024). Often characterised as an extension of poststructuralism's decentring of the human subject (Cord 2022), notions of machines self-developing in this way have been influential across the humanities and social sciences (Blok 2024; Dahlin 2024). For posthumanists, recognising agentic or autonomous capabilities in non-humans including machines has become intrinsic to a critique of dichotomised distinctions (e.g., human/non-human, animate/inanimate) that underpin post-Enlightenment Humanism and anthropocentrism (Braidotti 2013). We call self-development in this sense post-dualist self-development (PDSD).

In computer science, and notably taken up in broader explanations across historical, regulatory and government discourses, artificial intelligence (AI) is also defined or described—from a highly specific technical standpoint—in terms that reflect a continuum of 'self-development' that is independent of human control. In the case of large

✉ Claire Tanner
claire.tanner@monash.edu

Sam Cadman
sam.cadman@monash.edu

Patrick Cheong-Iao Pang
mail@patrickpang.net

[1] Monash University, Melbourne, Australia

[2] Macao Polytechnic University, Macao, China

Published online: 17 April 2025

language models (LLMs) like that underlying today's generative AI (e.g., ChatGPT), computer scientists use the term 'emergence' to describe AI capabilities (e.g., producing or responding to language, computer code, audio, or images) that have already evolved in ways uncontrolled or unanticipated by humans (Bommasani et al. 2021). Future horizons in AI general intelligence (AGI), where AI systems are projected to potentially achieve 'super-intelligent' reasoning capacities, such as through a subjective reality that facilitates reflexivity and emotional intelligence, are framed more directly in terms of 'self-development' (Dubrovsky et al. 2022). Existing and foreshadowed emergent or self-developing aspects of AI are arguably not well understood in the public imaginary, nor have they been meaningfully engaged with in research focused on the affordances and ethical issues associated with AI to date. Yet they are arguably central to the urgency surrounding the need for appropriate policy and regulatory responses to AI, and to framing what responsible and ethical development of AI means in the current era of breakneck AI development and implementation (Trotta et al. 2023). Even given cycles of hype in technological development there are indications, and certainly much investment in the promise, that AI will continue its current trajectory of rapid increase in scale and complexity across multiple areas of human activity (Ooi et al. 2023), and that greater machinic 'self-development' could arise, by human design or as a side-effect of increased capability. We call this 'self-development' continuum, from today's emergent LLMs to possible future directions in machinic subjectivity or super-intelligence, technical self-development (TSD).

Below, we discuss how the advent of AI technologies characterised by TSD requires a reconfiguring of posthumanist conceptual approaches to PDSD to ensure the relevance of posthumanism to ethical AI advancement. Drawing on posthumanist approaches to AI to date, we identify a slippage or conflation of meaning between PDSD and TSD that necessarily compromises posthumanist critique of AI-related social and political developments. Historicising the development of posthumanist theory, we suggest that technical developments in AI prompt an urgent need to advance posthumanist grappling with ideas around 'self-development', agency and autonomy, to enable nuanced engagement with how TSD as related to AI is ontologically discrete. We discuss how these theoretical tensions appear to be manifesting in preliminary responses to AI that may be ignoring or underestimating regressive Humanist and anthropocentric aspects of current AI innovation, particularly with respect to the influence of anthropomorphism. Finally, we outline how a posthumanist ethics of generative AI that pays requisite attention to both TSD and PDSD may critically inform more accurate, nuanced and anticipatory assessments of the risks and benefits of discrete AI technologies, to inform public discourse, appropriate social, institutional, policy and governance responses, and direct AI research and development priorities.

1.1 Technical self-development (TSD) and AI

Karen Barad's theory of 'agential realism' (2007) has been influential in the evolution of posthumanism's materialist theories of matter as 'auto-poietic or self-organizing' (Braidotti 2013, p. 158; Köves et al. 2024), as discussed below. But Barad's intervention is significant methodologically as well as substantively. Deriving a posthuman theoretical analysis from physicist Neils Bohr's theories of quantum mechanics, Barad attends closely to measuring apparatuses such as Bohr's two-slit diffraction or interference experiment (2007, p. 82). Barad emphasises the need to 'remain rigorously attentive to important details of specialized arguments within a given field', to avoid 'coarsegrained portrayals' that make a caricature of other disciplines (2007, p. 93). Following this approach, we begin our posthumanist enquiry into AI by attending to how AI is conceptualised by computer scientists and policymakers, particularly how technical artefacts are seen as contributing to their own development in contradistinction to being directed or controlled by humans. To begin we look back.

The term 'artificial intelligence' was coined by the 'Dartmouth Summer Research Project' in 1956 (Elliott 2022; Anderson 2024). The notion of machines developing independently of humans was central to the novel concept of AI. The Dartmouth researchers challenged themselves to 'make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves' (McCarthy et al. 2006, p. 12, emphasis added). Celebrated computer scientist Alan Turing had previously emphasised non-human interior development as essential to 'intelligent' machines. In what became known as the 'Turing test', Turing (1950) proposed the influential 'imitation game' to determine if a machine 'thinks', based on whether it could impersonate a human's typewritten responses to questions. Turing hypothesised that a 'thinking' machine might be created through a process of development within the machine itself: 'Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's?'…Presumably the child-brain is something like a notebook as one buys it from the stationers. Rather little mechanism, and lots of blank sheets' (1950, p. 456). Turing distinguished a putative 'learning machine' of this kind, which would display 'intelligent behaviour' by developing independently of human programmers, from more predictable 'computation machines':

> An important feature of a learning machine is that its teacher will often be very largely ignorant of quite what is going on inside, although he may still be able

> to some extent to predict his pupil's behaviour…This is in clear contrast with normal procedure when using a machine to do computations: one's object is then to have a clear mental picture of the state of the machine at each moment in the computation…*Intelligent behaviour presumably consists in a departure from the completely disciplined behaviour involved in computation* (1950, pp. 458–459, emphasis added).

Since the 1950s, the term 'AI' has become increasingly unstable (e.g., Svensson 2023). It now routinely refers to machines reflecting vastly different kinds and degrees of 'intelligence', from 'smart' elevators to voice assistants like 'Siri' or 'Alexa' to video game opponents to self-driving cars to deep machine learning (Elliott 2022). In many contexts where 'AI' is used, little claim or explanation is provided about non-explicit learning or development capability resembling what the Dartmouth researchers or Turing originally had in mind. Part of this instability may be attributed to the difficulty in defining intelligence, human-like or otherwise (Elliott 2022; Kaplan and Haenlein 2019). Machines with today's power and functionality may appear 'intelligent', even if they will seem limited or sluggish in ten years' time. Another cause of imprecision may be the 'ignorance' rising in our age of advanced machines. 'Ignorance Studies' is now a distinct field of study (Gross and McGoey 2022). Philosopher Alfred Nordmann describes how 'researchers need to accommodate ignorance as part of their daily lives… The insides of their instruments are as opaque to most researchers as are the insides of computers to most of their users' (Maasen et al. 2020, p. 27). If modern life involves navigating multiple technologies without knowing how they work (Innerarity 2021), there may be an understandable tendency to treat 'AI' as a synonym for complex or 'advanced' machines. This uncertainty is arguably exacerbated by modern technological developments being implemented by stealth. Apple's voice assistant 'Siri', for example, was released in 2010 but continues to be upgraded online; the kind of 'AI' that 'Siri' and other voice assistants represents may be constantly changing. A further reason for indiscriminate references to 'AI' may be its growing allure in marketing. Recent bibliometric studies have identified exponential increases in 'AI'-related literature across marketing-related fields (Mariani et al. 2021), especially 'psychology and human–computer interactions', 'computer and information systems', and 'business research' (as well as 'marketing' per se). Notably, such analyses depend on keyword co-occurrence networks, so they accept references to AI-related terms at face value rather than critically evaluating what machine 'intelligence' refers to.

For computer scientists, however, capacity for 'self-improvement' or 'learning' rather than pure computation remains a determinative characteristic of properly so-called AI (Kaplan and Haenlein 2019). Computationally intense tasks like solving non-linear differential equations or abstruse mathematical conjectures do not require 'intelligence' if achieved by 'finite-differences number-crunching' or 'try-all-possibilities brute force' (Chen & Chen 2022, pp 39–40). By contrast, the AI of LLMs like that underlying ChatGPT consists in behaviour that is 'implicitly induced rather than explicitly constructed' (Bommasani et al. 2021, p. 3). An LLM is trained using raw text, broken into tokens and encoded as numerical data. Using a 'transformer' to determine the probability of words occurring together or referring to each other at a distance (Vaswani et al. 2017), the LLM translates the training data into a multi-layered neural network of nodes connected by edges modelled on the human nervous system (Wolfram 2023). After training, the resulting neural network is used to generate text from previously unseen prompts in a process called 'transfer learning' (Bommasani et al. 2021, p. 3).

This training process is substantially invisible to and unsupervised by human developers. The lack of supervision means LLMs are faithful to historical notions of AI as noted above and is (in part) constitutive of what we refer to here as TSD. Four further aspects of LLMs are especially relevant for our discussion.

First, LLMs do not always select the most probable word. Instead, there is a 'temperature' parameter that introduces 'randomness' into the model's choice (OpenAI 2024a; Wolfram 2023). OpenAI explain that 'Higher [temperature] values like 0.8 will make the output more random, while lower values like 0.2 will make it more focused and deterministic' (OpenAI 2024b). Randomness applied by a computer is always classified as 'pseudo-random', because computers are essentially deterministic and hence not capable of producing true randomness (Lau 2024a, b, p. 243). This contrasts with more truly random processes like radioactive decay or the quantum mechanics that underlie Barad's agential realism (2007). It is LLMs' pseudo-randomising temperature setting that means the same prompt generates different text each time (Wolfram 2023). Developers working with OpenAI's application programming interfaces (APIs) have concluded that the temperature setting cannot be zero, because it is the denominator in a fraction, so some randomisation is always applied, although smaller temperature settings (so-called 'greedy sampling') can make results more predictable (OpenAI 2024b). Pseudo-randomisation within defined parameters (e.g., pick a number between 10 and 20) is on its own a purely computational function.

Secondly, while the emergent potential of deep neural networks and self-supervised learning have been recognised for decades (Wang and Chen 2023), the dramatic impacts of generative AI services like ChatGPT since late 2022 are primarily due to increases in scale: 'Transfer learning is what makes [LLMs] possible, but scale is what makes

them powerful' (Bommasani et al. 2021, p. 4). Substantial increases in hardware throughput have enabled unprecedented quantities of text to be analysed in training. The model that underpinned ChatGPT when it was released was called GPT-3.5. Its predecessor, GPT-2, had 1.5 million parameters in 2019 (Bommasani et al. 2021, p. 5). By 2020, GPT-3 had 175 billion parameters, resulting from expansion in the size of the training data (Wolfram 2023). ChatGPT's creator OpenAI published that GPT-3 was trained using around 500 billion tokens. The more recent GPT-4 model has been estimated to have 1.8 trillion parameters, implying a correspondingly larger amount of training data. The primary functionality of ChatGPT, generating text in response to a prompt ('in-context learning'), was itself 'an emergent property that was neither specifically trained nor anticipated to arise' (Bommasani et al. 2021, p. 5). The same is true of LLMs' emergent ability to respond to human feedback, which has led to the current focus, especially in the education sector, on 'prompt engineering':

> One might have thought that to have the network behave as if it's 'learned something new' one would have to go in and run a training algorithm, adjusting weights, and so on…Instead, it seems to be sufficient to basically tell ChatGPT something one time—as part of the prompt you give—and then it can successfully make use of what you told it when it generates text (Wolfram 2023).

It is these emergent affordances resulting from scale that have rapidly transformed expectations in the AI space. OpenAI reported in March 2024 that 'on a simulated bar exam, GPT-4 achieves a score that falls in the top 10% of test takers' (OpenAI et al. 2024, p. 1). The recently released Deepseek model has been credited with 'spontaneous emergence of complex, self-reflective behaviors' (Zhang 2025). The ready availability of distinctive human-like text that is responsive to prompts and feedback has already dramatically impacted sectors like education (Russell Group 2023) and has been projected to cause substantial job displacement (e.g., United States 2023).

Thirdly, LLMs' emergent capacities are inseparable from discourses surrounding AI risks. A United States Executive Order dated 30 October 2023 (now revoked), which described AI as 'hold[ing] extraordinary potential for both promise and peril', tellingly referred to 'AI's opacity and complexity' and instructed US government agencies to conduct 'red-teaming' tests, defined as 'adversarial methods to identify flaws and vulnerabilities, such as harmful or discriminatory outputs from an AI system, unforeseen or undesirable system behaviors, limitations, or potential risks associated with the misuse of the system' (United States 2023). LLMs were thus depicted as difficult if not impossible to know or predict, with their unsupervised training and hidden neural network functionality translating into whack-a-mole-type risk assessment at the highest levels of government. The European Union Regulation on AI, a nuanced and comprehensive regulatory response to the post-ChatGPT environment, places critical emphasis on the emergent nature of LLMs (Wörsdörfer 2024). For example, Article 51 includes a definition of a 'general-purpose AI model with systemic risk' as a model that has 'high impact capabilities' (Regulation 2024, p. 83). Such high impact capabilities are determined by taking into account matters including the 'number of parameters', the 'quality or size of the data set, for example, measured through tokens', and 'the benchmarks and evaluations of capabilities of the model, including considering the number of tasks without additional training, adaptability to learn new, distinct tasks, its level of autonomy and scalability, the tools it has access to' (p. 144). The EU's regulatory position thus proceeds on the basis that risk of harm rises commensurately with the degree of LLMs' emergence. By referring to scalability and access to varied tools, the Regulation also draws attention to a factor computer scientists have termed 'homogenisation'. LLMs' dependence on scale requiring massive computational power has led to unprecedented centralisation (Bommasani et al. 2021). A small number of LLMs dominate natural language processing in text form and are increasingly transferred to other modalities such as images, speech, tabular data, protein sequences, organic molecules, and reinforcement learning (Bommasani et al. 2021). This process is accelerating quickly as LLMs are made available to third party developers via application programming interfaces (APIs) (Togelius 2024). This means that identical underlying risks are shared across diverse arenas: 'Since emergence generates substantial uncertainty over the capabilities and flaws of [LLMs], aggressive homogenization through these models is risky business. Derisking is the central challenge in the further development of [LLMs] from an ethical and AI safety perspective' (Bommasani et al. 2021, p. 6).

Finally, for computer scientists, the emergent AI represented by LLMs is as noteworthy for its weakness and limitations as for its abilities (Kaplan and Haenlein 2019; McLean et al. 2021). It has been argued that because current LLMs 'cannot think, understand, explain, and pose problems', everything they do 'is more like automation than creativity…mainly due to the acceleration of computer power and the use of heuristics implemented on digital devices' (Dubrovsky et al. 2022, p. 5), and that 'a fundamental limitation of current LLMs [is that] they lack the capability to generate original outputs, [being] constrained by their training data' (Thórisson et al. 2024, p. 3). Even AI developers like ChatGPT's OpenAI conceive of today's LLMs as merely a preliminary step on the way to 'artificial general intelligence' (AGI) that is 'generally smarter than humans' and will 'elevate humanity by increasing

abundance, turbocharging the global economy, and aiding in the discovery of new scientific knowledge that changes the limits of possibility' (OpenAI 2023). Speculations about superintelligent AGI frequently allude to existential risks to humankind (Mitchell 2024), including by AI developers themselves: 'the risks could be extraordinary. A misaligned superintelligent AGI could cause grievous harm to the world' (OpenAI 2023). Nick Bostrom famously conjectured that an AGI programmed to maximise paperclip production could theoretically convert the Earth and much of the observable universe into paperclips (Bostrom 2013). Both negative and positive predictions about AGI, however, tend to focus on patterns of 'self-development' that exceed or are different in kind from the emergence of LLMs to date. A putative beyond-human intelligence would arguably have to be capable of developing itself to reach that point, which could require or be facilitated by a 'subjective reality and a self-developing poly-subject (reflexive-active) environment' (Dubrovsky et al. 2022, p. 5). It has been argued that existing predictions around AGI are predicated on a logically impossible 'computational dualism' between software and hardware (Thórisson et al. 2024, p. 22), and that intelligence may be 'essentially embodied…rather than putting a fully formed brain into a machine, a machine becomes "brainy" by interacting and learning from its world' (Jecker 2024). Ideas around machinic subjectivity and embodied self-development thus harken back to the 'child-machine' hypothesized by Turing. Forms of subjectivity or machine consciousness are speculated both to represent greater 'intelligence' and to help prevent doomsday scenarios, since subjectivity may enable a machine to depart from its terminal objectives (e.g., producing paperclips) (Thórisson et al. 2024, p. 180). Today, from a technical perspective, the concept of machines developing independently of humans begins to look increasingly like a continuum, with emergent LLMs as a weak entry point and potential advancement consisting in progressive degrees of greater 'self-development'. For now, as contested and uncertain as AGI may be in terms of meaning and coherence, it seems indisputable that ethical treatments of LLMs are urgently required today and potentially critical for the future. Given its commitment to building ethical futures focused on human-techno relationalities that de-centre the 'human', posthumanist theory appears to be a critical place to begin such ethical reckonings (Nath & Manna 2023).

1.2 Posthumanism, and AI as a 'posthumanist' apparatus

While posthumanism as a philosophical framework is multidisciplinary and heterogeneous, two constitutive projects are clear (Braidotti 2019; Ferrando 2019). First is a post-Humanist critique of the subject. A successor to the anti-Humanism expounded by Foucault (1966/2005) and Lévi-Strauss (1966) in the 1960s, this post-Humanist limb rejects 'the exclusionary nature of the specific vision of the human subject situated at the centre of the Western social, symbolic and discursive order' (Braidotti 2013). Instead, it foregrounds a non-unitary image of the human as 'materially embedded and embodied, differential, affective and relational' (Braidotti 2019). This aspect of posthumanism seeks to embrace and promote just outcomes for those excluded or Othered by Eurocentric Western Humanism (e.g., women, LGBTQ + people, the colonised, Indigenous peoples, and other excluded peoples of non-European ancestry) (Braidotti 2013). The second defining project of posthumanism—sometimes seen as distinct (Braidotti 2013, 2019) and elsewhere as implicit in the first (Mauthner 2019)—is a post-anthropocentric critique challenging humans' supposed hierarchical supremacy over non-human subjects including the environment and other non-humans (Ferrando 2019; Herbrechter 2022). Anthropocentrism has been identified as a fundamental cause of the Anthropocene's threats to ecology and sustainability, because 'the centrality of the human implies a sense of separation and individuation of the human from the rest of beings' which leads to abuse on non-humans and geological effects like climate change (Ferrando 2019).

Posthumanism thus provides a theoretical framework for exploring how a dominant Humanist philosophy predicated on a specific vision of the human subject (e.g., as white, male, heterosexual, able-bodied) shapes practical outcomes that bend towards that contingent ideal, and how a dominant anthropocentric philosophy shapes practical outcomes that bend towards human interests. The posthumanist perspective we adopt involves focusing attention on how contingent Humanist or anthropocentric values are disseminated and reinforced, with a view to increasing the visibility and intelligibility of underlying philosophical ideas that influence trajectories of human-techno relations and activity in the Anthropocene—namely, for our purposes here, in relation to AI.

In advancing the constitutive twin critiques of Humanism and anthropocentrism, posthumanist theorists usually cast technological development and entities in a central role. The idea of post-dualist self-development (PDSD) outlined at the start of our discussion, where matter is recognised as active or agentic, draws on the work of diverse theorists (e.g., Spinoza, Latour, Haraway, Deleuze and Guattari, Bennett, and Barad) to reject binaries at the core of the Humanist and anthropocentric worldview (e.g., human/non-human, animate/inanimate, natural/artificial) (Braidotti 2013; Ferrando 2019). Braidotti, for example, invokes Deleuze and Guattari's notion of 'assemblage', as a 'social or collective machine…that determines what is a technical element at a given moment, what is its usage, extension, comprehension, etc.' (Deleuze 1987, p. 398), to explain:

> Contemporary machines are no metaphors, but they are engines or devices that both capture and process forces and energies, facilitating interrelations, multiple connections and assemblages…The merger of the human with the technological results in a new transversal compound, a new kind of eco-sophical unity, not unlike the symbiotic relationship between the animal and its planetary habitat (2013, p. 92).

Following this project in posthumanist scholarship, recent posthumanist enquiries have fastened onto the seeming agentic capabilities of generative AI as evidencing an intrinsically posthuman character, yet commonly without attending to how generative AI operates at a technical level. For example, a discussion focused on AI creativity designates AI as 'explicitly posthuman', on the basis that it is 'by design interactive and, under most current iterations, based on a capacity to learn from its environment' (Kalpokiene and Kalpokas 2023, pp. 2–3). The authors do not attend to how AI functions beyond alluding to a recent 'flourishing of what is often called computer-generated art—essentially, artificial creativity' (Kalpokiene and Kalpokas 2023). In turn, they argue that AI originality and creativity ought to be protected by copyright law[1]: 'the refusal to acknowledge AI creativity is a clear attempt to hold on to the futile idea of human exceptionality…an equitable solution leading towards creativity shared by humans and AI should be sought as an outcome more compatible with posthumanist ethic' (p. 6). Another ambitious posthumanist discussion regarding literacy and AI invokes theorists including Latour, Barad, and Deleuze and Guattari to foreground the significance of AI's autonomy:

> Thinking with these theories allows us to ask new, critical posthumanist literacy questions about ourselves and our interconnections with AI: how does thinking of ourselves as tangled up in a dynamic network of human–data–AI processes change how we think of ourselves (our identities) and our capacity to make change (our agency)? Most importantly, who or what has the power to enact these boundaries? What these ontologies do for us is move away from the idea of AI as merely an inert tool for humans to use; AI also uses and moves us, relying on the data we generate to operate and shaping—often invisibly—our (literacy) activities (Burriss and Leander 2024, pp. 563–564).

AI is thus distinguished from other apparatuses as a specifically posthumanist site of human/machine interaction, based on AI's active and invisible role in co-shaping activities. There is little reference to the computational processes underlying AI; instead, agency is attributed to the invisibility of its current processes: 'AI is complicated for humans to visualize and includes far more disparate components and processes than we can see with our eyes at any given moment' (Burriss and Leander 2024, p. 563). Another study entitled 'Posthuman Creativity: unveiling Cyborg Subjectivity Through ChatGPT', which explores online interactions with ChatGPT, describes ChatGPT as 'the potential manifestation of a posthuman subject', owing in part to the fact that its 'responses are not predetermined, but rather emerge from an assemblage, the complex interaction between its network of nodes and the input it receives, transforming the body beyond its existing categorization' (Yan 2024, pp. 6–8). Posthuman subjectivity is thus depicted as fundamentally about beyond-human control, with ChatGPT held up as an exemplar. Again, there is no detailed investigation of what the 'complex interaction' that produces text involves. Indeed, the ambitious contention that 'If AI is generating "human-like" texts, it blurs the line between what is considered "human" and what is machine' is supported by a quotation provided by ChatGPT itself ('When given an initial text as prompt, [ChatGPT] can produce detailed human-like textual outputs') (Yan 2024, p. 2).

A perhaps more measured enquiry into 'transhuman language ontologies' (Demuro and Gurney 2024), again based on interactions with ChatGPT, pays more careful attention to technical aspects of LLMs,[2] while also adopting an overtly posthumanist perspective ('"we", as humans, are not the only entities capable of practising and generating language(s)/languaging') (Demuro and Gurney 2024, p. 3) to emphasise the agentic significance of AI, specifically in relation to language:

> ChatGPT is a nonhuman other with capacity for affect–that is, it has the capacity to affect and be affected…It is responsive, and it can be responded to…
> As transhuman language ontologies enter into the languaging encounter, they are performed, and become 'real'. This presents a need to develop a framework to engage with and understand *transhuman language ontologies*; this involves reconsidering our conceptual toolkit for recognising and investigating instances of what constitutes language or languaging in the world (Demuro and Gurney 2024, pp. 5–9, emphasis in original).

[1] For a discussion of the legal and moral arguments with respect to the use of copyrighted material by AI image generating software see Shoemaker (2024).

[2] E.g., 'ChatGPT…treats 'natural language' as sequential data, and produces language through inductive reasoning, a process of generating linguistic expressions or constructions based on patterns and regularities observed in previously encountered data' (Demuro and Gurney 2024, p. 9).

Here, the responsiveness of ChatGPT results in affective capacities analogous to human affect, and an ostensibly posthumanist framework emphasises the game-changing impact of ChatGPT in revealing the need for a revised 'conceptual toolkit' to enquire into the meaning of language itself.

The depiction of AI as especially or uniquely posthumanist or post-anthropocentric is perhaps most strongly apparent in discussions of AI-produced art. One study focused on AI creativity asked participants to express preferences about artistic creations based on misleading information about whether authors were human or AI (Millet et al. 2023). The study defined AI only in terms of creative outputs, which were said to '[extend] beyond emulating already existing artistic styles' and to include 'original artistic styles', 'original songs and music scores' which 'are often indistinguishable from human-made art' (Millet et al. 2023, p. 2). The authors incorporated anthropocentrism into the research design ('we expect people to respond to AI art by derogating its artistic value to defend their threatened anthropocentric worldview' (Millet et al. 2023, p. 2)) and interpreted their results as evidence of AI disrupting anthropocentric beliefs ('people inadvertently reveal their need to support anthropocentrism in the face of recent advances of AI that threaten the last fortress of human supremacy arguments, artistic creation' (p. 7). Another questionnaire-based study investigating personality traits and receptivity to AI-produced art showed participants 40 randomised artworks: 20 by humans and 20 by AI (Grassini and Koivisto 2024). The project design was once more framed using anthropocentrism ('people may tend to defend their anthropocentric beliefs by downgrading the artistic value of AI-generated art… AI art can pose a significant psychological challenge, impacting people's ontological security' (p. 2)). The conclusion that 'humans may devalue AI-generated artworks' was theorised as 'possibly due to a cognitive bias (as e.g., proposed by Anthropocentrism theory)' (p. 9). Again, the definition of AI was based on the ability to produce original paintings, songs and poetry that '[transcend] the mere imitation of existing styles' and are 'indistinguishable from human-made art' (pp. 1–2), without attention to how this is achieved.

These preliminary posthumanist responses to generative AI thus seem to be mapping a hierarchy of value-based judgements, derived substantially from PDSD, onto the apparatus of AI itself. At times, the logic of this burgeoning orthodoxy seems to be that AI's nature as an assemblage or entanglement of technology and social networks constitutes AI as an inherently posthumanist phenomenon, and that to express resistance to AI, for example by objecting in principle to positive appraisals of AI-created text or images, is to commit an anthropocentric—even speciesist—act.

1.3 AI and PDSD

In addition to highlighting a critical need for meaningful interdisciplinary work across social and computer sciences in approaching the study of AI developments, we suggest that early posthumanist responses to everyday generative AI, characterising AI as essentially 'posthumanist' owing to its supposedly self-developing or agentic character, also reveal a need to reconceptualise PDSD and its role in posthumanist philosophy.

A threshold conceptual problem, which has arguably always been implicit in theories surrounding PDSD but takes on critical significance in the wake of generative AI, is whether PDSD is conceived as a classification of type or degree. To further posthumanism's critiques of Humanism and anthropocentrism, a dualist Cartesian separation between humankind and machines is rejected in favour of a hybrid symmetry between humanity, machines and the natural environment (Blok 2024; Ferrando 2019). Machines in this sense appear as a collective rather than as individuals; Braidotti's reference to a 'merger of the human with the technological' above, for example, seems to refer to machines in general, rather than some and not others. Yet PDSD is often framed in terms that imply potentially character-changing differences in degree. Haraway famously described that 'Late twentieth-century machines have made thoroughly ambiguous the difference between natural and artificial, mind and body, self-developing and externally designed, and many other distinctions that used to apply to organisms and machines' (1985/2006, p. 120). In this theorisation, self-development has augmented over time: something about late-twentieth century machines made ambiguous a distinction, which was once not ambiguous or 'used to apply', between self-developing and externally designed machines.

On the surface, Haraway's approach seems different from other renditions of PDSD, such as Latour's actor-network theory. In a discussion substantially focused on the ontology of a hammer through history, Latour describes how 'all technologies incite around them that whirlwind of new worlds. Far from primarily fulfilling a purpose, they start by exploring heterogeneous universes that nothing, up to that point, could have foreseen and behind which trail new functions' (Latour 2002, p. 250, emphasis added). Yet, for Latour too, the sense of technical artefacts as active rather than passive is accompanied by a quality of ongoing expansion: the 'very complexity of the apparatuses…is due to the accumulation of folds and detours, layers and reversals, compilations and re-orderings' (p. 251). Such a complexity (i.e., that arises from accumulating features) presumably is not equivalent in degree from one apparatus to the next, or from one moment in time to another. Rather, in the case of any individual apparatus, its complexity (and the ontological

significance of its complex nature) must be tied to specific accumulations and their respective influences, determined by historical evolution. Braidotti draws on Guattari's 'machinic autopoiesis' to define seemingly all machines as essentially 'intelligent and generative', with 'their own temporality… virtuality and futurity' (2013, p. 94), yet also contends that some improved or enlarged technical capacities are more significant to PDSD, designating weaponry- and surveillance-related technologies as specifically 'post-anthropocentric' (i.e., compared to other machines that must not be post-anthropocentric), and arguing that 'As they become smarter and more widespread, autonomous machines are bound to make life-or-death decisions and thus assume agency' (i.e., compared to machines that must not have agency) (p. 44).

What emerges, then, is an apparent Orwellian fallacy: all technical devices are self-developing, but some are more self-developing than others. Living things like birds 'self-develop' from conception until death or decomposition, presumably the kind of parallel PDSD is seeking to establish with machines to disrupt Cartesian dualities between animate and inanimate matter, but it would be strange if an emu were considered more 'self-developing' than a wren because it is bigger or its species has been around longer, or if the converse were said about the wren because it can fly. Using PDSD as a basis for claiming that some technological artefacts (like AI) are especially 'posthumanist' seems to implicate this problematic framing from the outset. Reading a technological assemblage as a 'social or collective machine' is so wide in scope that comparative assessments of PDSD (and hence machines that are more or less 'posthumanist') become profoundly uncertain, if not impossible. Which apparatus is more self-developing through a PDSD lens: the one used by more cultures, or the one put to more uses? The one that took shape earlier in history, or the one more prominent today? The atom bomb or the printing press? The sword, the gun, or the pen? The burgeoning orthodoxy identified above, classifying AI as a particularly posthuman development because of its apparently agentic character, seems to rely fundamentally on this precarious foundation.

Further problems potentially flow from this initial premise. Bearing in mind David Hume's cautionary observation about slippage from 'is' to 'ought' propositions (Capaldi 1966), attributing agency or autonomy to all machines might have a primarily descriptive or illuminating purpose, facilitating acknowledgment of interconnectedness in ways that, as much as they may disrupt Humanist or anthropocentric moral hierarchies, can be framed as morally agnostic. A continuum that singles out AI as especially posthumanist, however, seems to imply a moral hierarchy (i.e., the more PDSD the better) with AI positioned near the apex as an intrinsic moral good. Also, classifying generative AI as a significant 'posthuman' phenomenon involves distracting attention away from wider social and historical factors that bear on its ontology and emphasising instead the supposed significance of narrow technical capabilities. This kind of thinking around machines, attributing ontological significance based primarily on intended functions rather than broader assemblages, may reflect precisely the Humanist values that PDSD was intended to reject.

1.4 AI, humanism and anthropocentrism

The practical limitations arising from conceptual tension between PDSD and TSD are not limited to attributing AI a privileged moral status as a 'posthumanist' technical development. There is also the difficulty that early posthumanist responses to generative AI, perhaps meeting the conceptual limitations of PDSD in a world where TSD is increasingly prevalent, may fail to acknowledge or mistakenly embrace regressive Humanist and anthropocentric influences disseminated under the aegis of AI.

The early posthumanist responses to AI reviewed above reveal that what 'AI' means is usually based on an ability to imitate human behaviours, such as to create human-like art or music, to perform human-like affect, and especially to produce human-like language or text. If generative AI is a convincing human imitator, it might satisfy the 'Turing test' for identifying a 'thinking' machine noted above. However, there seems to be a dissonance about posthumanism adopting without close interrogation the 'Turing test' or like attributions of ontological significance based on imitation of a/the 'human'. Posthumanism's project to de-centre a Eurocentric notion of the human as a foundational category of thought is usually associated with de-emphasising the supposed exceptionalism of abstractly conceived human capabilities. Does a human-imitating machine challenge human exceptionalism or anthropocentrism, or reinforce the idea that human capabilities are the ones that really matter? For posthumanists, Humanism in the Cartesian mould is critiqued as a hierarchy based on disembodied abstract values. The following explanation by Braidotti is instructive:

> [Humanism] spells out a systematized standard of recognizability–of Sameness–by which all others can be assessed, regulated and allotted to a designated social location…The human norm stands for normality, normalcy and normativity. It functions by transposing a specific mode of being human into a generalized standard, which acquires transcendent values as the human: from male to masculine and onto human as the universalized format of humanity. This standard is posited as categorically and qualitatively distinct from the sexualized, racialized, naturalized others and also in opposition to the technological artefact. The human

> is a historical construct that became a social convention about 'human nature' (2013, p. 26).

With this notion of Humanism as a standard of 'Sameness' in mind, posthumanism is well positioned to ask: *who or what is the human that AI imitates?* Generative AI arguably does not manifest many (perhaps any) of the qualities that posthumanists have admirably insisted on as intrinsic to reconceptualising posthuman subjectivity (Braidotti 2013, 2019). AI is not embodied or embrained, sexed, differential, affective, or relational, in any sense that is comparable to a human or other animal—unless, as noted above, PDSD recognises those qualities as common to all machines. Instead, by arranging alphanumeric characters, punctuation marks, image pixels or other data into the appearance of things that 'a human' might make, generative AI arguably derives ontological significance (and elicits affective responses from real humans) by representing a stylised form of disembodied abstract thought more analogous to the universalised abstraction of the human underlying Humanism than diverse subjectivities reflecting difference or Otherness (indeed, subjectivity of any kind).

It might be argued that in practice AI does not represent a singular version of a human; since outputs are shaped by prompts, responses can be elicited from diverse AI-generated voices. But Humanism has always been comfortable addressing or impersonating people of all kinds. Postcolonial theorist Homi K. Bhabha famously observed that 'from the high ideals of the colonial imagination to its low mimetic literary effects mimicry emerges as one of the most elusive and effective strategies of colonial power and knowledge' (Bhabha 2004, p. 122). While posthumanist and post-colonialist thinkers come into tension around how far the category of 'human' ought to be rejected ('If the human subaltern cannot speak in the postcolonial world, the possibility of them being heard in the posthuman world becomes a distant cry' (Islam 2016, p. 128)), the two movements share an essential methodological focus on the role of discourse in sustaining ideology (Deckha 2012). Braidotti has described, for example, how postcolonial and decolonial theories 'offer a painstaking critical analysis of the extent to which racial assumptions and white supremacy have shaped the philosophical discussions about the human, that Western philosophers have come to take for granted' (2019, p. 26). If posthumanism interrogates seriously whether generative AI might in fact imitate, or give voice to, a universalised and fundamentally Humanist conception of the human, a capability to mimic subaltern subjectivities may be powerful evidence for the view that it does, rather than that it does not.

Beyond these high-level questions about generative AI and the human, close attention to how weak or emergent TSD enables generative AI to function arguably reveals at a more granular level the potential for Humanist or anthropocentric influences to be reinforced rather than disrupted. A critical entry point is the essential role played by substantial training data in LLMs' functionality, and its prospective implications from the perspective of human language evolution.

The ongoing evolution of human language was described recently as among the 'most prominent research directions of the past few years' (Markov et al. 2023, p. 1). There is a growing consensus that human language is constantly emerging and evolving and may be strongly influenced by cultural factors (Pleyer and Hartmann 2024). For example, commentators have drawn attention to the Anthropocene and the digital age as profoundly influencing language evolution (Markov et al. 2023). A potentially critical development for the ethics of generative AI is the idea that the cultural evolution of language does not occur at the level of whole words but rather statistically recurring re-combinable parts, such as phonemes or morphemes, a process that reflects human learning practices beyond language:

> Both of these properties of language—having parts and having them distributed in a particular way—can emerge through cultural transmission when learners are simply reproducing wholes…structure in language may similarly arise via cultural evolution through a highly general process of iterated sequence learning, rather than learning biases that are specifically adapted to language (Arnon and Kirby 2024, p. 9).

This view that language structure evolves has an often assumed yet foundational significance for posthumanist philosophy. Critical approaches to discourse involve identifying hegemonic practices whereby ideologically freighted representations become so ingrained that they appear neutral or objective (Fairclough 2015; Fowler 2007). For posthumanism, Humanism and anthropocentrism constitute embedded discursive practices of this kind, whose contingent ideologies are disguised and perpetuated by dominant discourses. The idea of language evolution emerges as critical, in part because it directs attention to how dominant discourses become entrenched over time—even at the granular level of phonemes and morphemes—and in part because cultural language evolution offers a positive vision of a future where the impact of today's contingent ideologies, such as Humanism and anthropocentrism, can be ameliorated by the contributions of critical thinkers like posthumanists.

The publicly disclosed training data for the most prominent LLM, ChatGPT, includes complete scrapes of the internet, digital books, and Wikipedia. Discussions of LLMs to date have drawn attention to risks around racism, unjust outcomes or compounding existing prejudices (Bommasani et al. 2021). Extending these critical enquiries, a posthumanist enquiry into generative AI could well ask: *to what extent does the training data reflect the ideological discourses*

of Humanism and anthropocentrism? Posthumanism has arguably emerged as an urgent response to the pressures of advanced capitalism because of the ongoing prevalence of Humanist and anthropocentric ethical values in dominant cultural practices. It may be true that posthuman ways of thinking and being have expanded in recent years (Braidotti 2019), but it would be optimistic to claim from a quantitative perspective that Humanist and anthropocentric ideologies no longer occupy an ascendant role in the Western canon of literature and philosophy, in corporate publications, in media or social media publications, or most of the text on the internet. The datasets used to train ChatGPT are also not sampled in proportion to size but rather based on OpenAI engineers' assessment of 'quality'. It is not possible to be certain because ChatGPT is proprietary, but tendency to promote Humanism or anthropocentrism (or other critiqued ideologies like colonialism) is probably not regarded as compromising training quality. Indeed, the secrecy surrounding what counts as 'quality' data could be seen as amplifying rather than defusing the potential for ideological concerns (Fisher et al. 2024). As research has identified, biases and disparities in training data, such as data that is predominately English-language and user-generated, can amplify racism, misogyny, homophobia, ableism and ageism (Bender et al. 2021). Other consequences of language ideologies embedded in LLMs that have been identified include uneven language performance, text outputs that amplify dominant or corporate interests, privacy violations, copyright infringement, increased circulation of misinformation or 'fictionalized nonsense' (Lau 2024a, b). To the extent the corpus used to train LLMs does reflect problematic Humanist anthropocentric ideology, the neural network architecture underlying LLMs will often, save perhaps in domain-specific areas where posthumanist texts exert a sufficiently countervailing statistical influence, manifest those ideologies with mathematical precision. So generally, if you want to know what anthropocentrism thinks, ask ChatGPT.

The potential for generative AI to promote rather than undermine Humanist or anthropocentric ideologies, as potent as it likely was when ChatGPT was released on 30 November 2022, may also increase with time. Researchers have demonstrated that training LLMs using their own outputs can cause 'model collapse', a 'degenerative process whereby, over time, models forget the true underlying data distribution' (Shumailov et al. 2024, p. 755). Quite apart from the ongoing efficacy of generative AI, however, it seems inevitable that the contingent ideologies embedded in the pre-release corpus will be magnified as usage of generative AI proliferates. OpenAI have reported producing more than 100 billion generative AI words a day (Bhatia 2024). Much of this may feed back into the training data of renewed LLMs via the internet or otherwise, meaning generative AI functions as an echo chamber for the ideological content of its original corpus. But even if AI developers effectively quarantine the pre-release training data, that data could arguably have an outsized ideological influence compared to an alternative reality where generative AI never took hold. Studies in the cultural evolution of language noted above tell us that human language is constantly evolving, and that the current digital age is accelerating this process. In a non-generative AI reality, material published on the internet between, say, 2002 to 2022 would by 2032 have become substantially historical, still influential in key respects no doubt, but overtaken and largely forgotten in others. However, in our generative AI-dominated moment, it seems possible that by 2032 a much higher proportion of the 2002–2022 content will continue to loom large as precious, genuine human content, a persisting originator of hundreds of billions of words generated by AI every day. It may be that one of generative AI's most effective tricks is trading on futurism while semiotically binding us to a perpetual present.

Focusing on the role of generative AI's training data also throws into relief the role of the pseudo-randomising 'temperature' setting in LLMs making results appear 'original'. As one developer explained on an OpenAI forum: 'A higher temperature…results in more *diverse and creative output*, while a lower temperature…makes the output more *deterministic and focused*' (Open AI Forum 2023, emphasis added). As noted above, pseudo-randomisation is a purely computational process integrated into the functioning of LLMs each time a word/token is selected for final output; it can be turned down but never off. The emergent neural network is always crucially involved in the selection of each word/token, because it identifies candidate words based on the preceding words. However, by the time the LLM has produced a sentence or paragraph, the overall meaning or representation has been determined by the neural network *and* the pseudo-randomness algorithm working in tandem to slice and dice the original sources in the training data. A higher temperature setting creates the appearance of creativity at the individual word-level, by extending the range for each candidate word (e.g., pick a number from one to 20 instead of one to ten), and at the sentence/paragraph level, by extending the range of possible candidate words as the neural network is prompted by more randomised text.

The essential contribution of pseudo-randomness has potentially critical implications for a nuanced posthumanist critique of generative AI. The supposed 'creativity' and 'originality' that have prompted posthumanist scholars so far to perceive generative AI as an essentially 'posthumanist' apparatus may be substantially the result of pseudo-random number generation, a process that is arguably not 'intelligent', 'original', or 'creative' in any non-human sense. Kari Weil has argued in relation to non-human animals that 'the urge to anthropomorphize…risks becoming a form of

narcissistic projection that erases boundaries of difference' (Weil 2012, p. 19, emphasis added). Anthropomorphism has long been a cornerstone of advertising strategies in the computer industry. A recent marketing study of anthropomorphised, gendered voices like 'Siri' and 'Alexa', for example, found that 'Endowing the AI with anthropomorphism seems like a great opportunity to curb the resistance against adoption of AI by acting as a buffer against people's biases' (Uysal et al. 2023). It has long been recognised that 'Anthropomorphism, enabling computers with human-like characteristics and capabilities, has been a driving philosophy behind the advances of computer technology' (Gong 2008, pp. 1494–1495; Caporael 1986). A posthumanist critique could consider the possibility that modern generative AI, for substantial economic gain, employs pseudo-randomisation to facilitate forms of anthropomorphism that, in Weil's terms, achieve potent Humanist and anthropocentric impacts by erasing essential differences at multiple levels: for example, between generative AI and other machines, and between generative AI outputs and human communications that express authentically differential and embodied experiences. From a practical standpoint, addressing the extent to which human vulnerability to anthropomorphise is being mobilised in the development, promotion and uptake of different AI technologies is key to adequate assessments of risk, and the development of appropriate governance and policy responses.

As is so far typical of generative AI, a key concern may be that forms of anthropomorphic transference—to which developers, users, researchers or regulators might be equally susceptible—elevate the perceived significance (and hence social or economic value) of TSD-related interactions in contradistinction to other forms of connection. An early discussion of Deepseek commented that 'it might even reach human-level accuracy in reasoning tasks' because, through reinforcement learning, it 'learns like a baby' (Tahir 2025). Risks to users that are built into current AI systems such as in generating 'seemingly legitimate but actually fabricated or misleading content' (De Angelis et al. 2023, p. 4 also cited in Shin et al. 2024, p. 3) are arguably enabled and amplified by anthropomorphism. As one participant in a small exploratory qualitative study on AI use stated, 'the more human it sounds, the more I trust it' (Troshani et al. 2020, p. 486). Whilst this potential to manipulate and build trust based on anthropomorphic vulnerability may be seen as advantageous from a marketing perspective, other potentially negative social impacts cannot be underestimated, especially for vulnerable users, such as those seeking healthcare or health information. For example, one recent study found that although usage of an 'Intelligent Social Agent' by lonely and suicidal students raised 'unexplored risks that require comprehensive scrutiny', the 'combination of conversational ability, embodiment, and deep user engagement shows a pathway for generalist Intelligent Social Agents to aid students…scaffolding their stress and mental health and even countering suicidal ideation' (Maples et al. 2024, p. 5). This kind of early use and engagement with risk and benefit of anthropomorphic AI machines is concerning, not least because of the inherent unreliability of AI models, and the risk of health misinformation built into LLMs that potentially constitute an emerging public threat due to what some researchers have called an 'AI-driven infodemic' (De Angelis et al. 2023, p. 4; see also Shin et al. 2024, p. 3). As Shin et al. note, irrespective of how successful 'Intelligent Social Agents' may be in imitating a caring human, AI models do not care. Instead, they 'lack personal experiences, emotions, or consciousness' and operate 'solely on learned patterns without any emotional depth or personal insight' (2024, p. 3). In this respect, conceptual tools (based on a distinction between PDSD and TSD) that enable researchers (and all users, including policymakers, companies, governments and digital citizens), to reflexively consider their own anthropomorphic vulnerability in approaching AI tools, as well as to guide understanding and interpretation of user experience, are critical to informing appropriate risk assessments, and policy and regulatory responses. This is especially imperative in providing adequate safeguards and protections for users, such as when the success of AI models (that may be increasingly unsupervised and inherently unreliable) rely upon the anthropomorphic vulnerabilities of users who are at risk, physically, mentally and or emotionally, notably in health contexts.

Perhaps more fundamentally, posthumanism offers a useful framework for enquiring into how LLMs effect an apparently persuasive but potentially false ontological separation between 'generative AI outputs' and works by individual human authors that comprise the training data. At least two lawsuits so far have been brought by news media organisations alleging that ChatGPT breaches copyright law in connection with training its LLMs (Grynbaum and Mac 2023; Stevis-Gridneff 2024). Speaking about the most recent litigation in Canada, an academic commentator explained: 'While it seems obvious that OpenAI is infringing copyright, it is technically very difficult to prove' (Stevis-Gridneff 2024). In the action brought by The New York Times, the claim alleges that copyrighted articles were produced verbatim by ChatGPT (Susman Godfrey 2023). In response, OpenAI argued that the Times used 'deceptive prompts' and '[fed] the tool portions of the very articles the sought to elicit verbatim passages of…Normal people do not use OpenAI's products this way' (Latham and Watkins 2024, p. 11). Quite apart from the lawfulness or otherwise of LLMs, the possibility that generative AI functions as a mathematically sophisticated process by which human knowledge is fragmented and pseudo-randomly reconstituted as apparently machinic knowledge, perhaps without the need for

legal attribution to the original authors, has serious consequences from the perspective of Humanist and anthropocentric ideology.

One powerful argument for a posthuman ethics is the emphasis on subjectivity and respectful recognition of difference, in contradistinction to the systemic sameness promoted by Humanism: '"we humans"…are not one and the same…For the subject to be materially embedded means to take distance from abstract universalism. To be embodied and embrained entails decentring transcendental consciousness' (Braidotti 2019, p. 4). By harvesting and anonymising human knowledge *en masse*, generative AI arguably represents human knowledge in a form that reflects precisely this kind of abstract universalism or putative transcendental consciousness. Knowledge is not traced or apparently traceable to the people who produced it, but to 'the model' or 'the algorithm'. The sense of technical sleight of hand may be especially strong in cases of highly specific domain knowledge, where the pool of relevant source materials in the training data must be narrower, and the pseudo-randomising algorithm is more likely to function as something akin to a synonym generator to replicate the outputs of human intelligence without acknowledgment or payment. OpenAI's business model, being predicated on progress towards superhuman AGI, places great emphasis on LLMs' burgeoning 'reasoning' capabilities, which may be technically emergent in the sense described above but are still emerging out of vast quantities of recorded human reasoning.

Testing a critique along these lines seems to be essential in the areas where generative AI is attracting the most attention, such as education and assessment of risk. Educational institutions would or should presumably baulk at the possibility of plagiarism on a massive scale, whether technically legal or not, yet there is already strong evidence of a paradigm shift among learners away from searching for information towards using LLMs to generate information instead (Luo et al. 2024). The risk that generative AI might be used for antisocial purposes like creating bioweapons may be very real. However, from a philosophical perspective, there is a significant difference if the risk arises from human-authored instructions buried somewhere in a wholesale scrape of the internet, or from words put next to each other by chance owing to the interaction of statistical processing of human knowledge combined with pseudo-randomness, than if a learning machine identifies they are an efficient solution to climate change and sets about persuading high school students to develop them. Presently, AI developers may be using generative AI's apparently convincing but substantially pseudo-random anthropomorphism to instrumentalise risk as a strategy to raise capital, feeding into fatalistic narratives about rogue AI superintelligence (Anderson 2024) to distract attention away from how today's generative AI is in fact much more limited and derivative (though not necessarily less dangerous). Again, adopting a posthumanist lens, this development could be viewed as representing a concerning form of regressive anthropocentric Humanism at this time in advanced capitalism.

1.5 So what? Some practical implications of TSD and PDSD as a conceptual paradigm

So far, we have sought to challenge conceptualisations of generative AI as a singularly posthumanist phenomenon. We have identified a fallacious conflation of TSD and PDSD in posthumanist approaches to generative AI, leading to neglect of LLMs' potential to promote and reinforce (rather than challenge) Humanist anthropocentric ideology. To conclude, we outline how TSD and PDSD might be reframed as a sequential posthumanist critical paradigm to understand and address potential benefits and risks of different AI machines, and guide policy and governance responses and future AI research and development.

A speech by UK Prime Minister Keir Starmer about 'AI' on 13 January 2025 (Starmer 2025) provides an illustrative example of how the conflation of different forms of AI machines in public discourse is being used to (over)state their value to humanity. Starmer's speech began with a prison officer who had a stroke and was saved by doctors pinpointing a blood clot ('That's the power of AI in action'), shifted to how AI could help individuals ('If you're sitting around the kitchen table tonight…worried about opportunity at your children's school…AI can help teachers plan lessons…tailored to your children's specific needs'), and climaxed with a patriotic call to arms ('Britain is going to shape the future. We are going to make the breakthroughs… We are going to create the wealth…So mark my words—Britain will be one of the great AI superpowers'). Yet modern day subjects also have to navigate more wide-ranging and conflictive narratives about the risks and benefits of generative AI in public discourse: from the internet being overwhelmed by 'enshittifying' slop (Mahdawi 2025) to the Nobel prizes in chemistry and physics (Li and Gilbert 2024); from empowering users to 'change the world' (Rosenberg 2024) to children 'losing the ability to think critically' (Grose 2024); from economic boom to bubble (Rau 2025). To navigate the tendency to conflate all AI technologies in utopian visions of social progress, as well as more diverse, conflicting and unstable AI realities, and attend to how they are being constituted and their effects, we argue theorisation of generative AI must begin by analysing how TSD functions in specific domains based on specific training data.

Machine learning algorithms may share a common ancestry, but their architectures vary across domains (Kriegeskorte and Golan 2019) and their outputs axiomatically correlate with domain-specific training data. The AI model AlphaFold2, for example, whose developers won the chemistry

Nobel prize, predicts three-dimensional protein structures using a transformer-based neural network trained on 'all known amino acid sequences and experimentally determined protein structures' (Li and Gilbert 2024, p. 1). These training data are highly homogeneous, express no ideological content analogous to human-authored text, and directly record biological phenomena. By comparison, the human words, language and art used to train LLMs like ChatGPT are essentially heterogeneous, freighted with ideology, and only indirectly represent sensory and cultural lived human experiences phenomenologically inaccessible to neural networks. If nuanced attention to domain-specific TSD reveals, as we have sought to demonstrate, that AI text generation is limited for the time being to plagiaristic fragmentation and reassembly of human-created artefacts, it can also reveal how in arenas like protein modelling, chess, or the archaeological discovery of geoglyphs (Sakai et al. 2024), neural network architectures are transforming horizons of possibility in ways that extend beyond human mimicry into more justifiable claims of 'artificial intelligence'. Only after such distinctions are identified, and made legible to and by governments, policymakers, regulators and wider publics, can actual and potential risks and benefits of AI technologies in discrete areas be properly anticipated, made visible and explored. Furthermore, it is only then that PDSD approaches to specific AI technologies and their use can know what they are working with, just as the assemblages relevant to a hammer must be distinguishable from those relevant to a rocket. Otherwise PDSD means no more than that everything is an assemblage of everything.

Anterior approaches to TSD that illuminate the 'devil in the detail' would greatly enhance the scope for PDSD to explore the ethics of ontological cross-pollination in AI-related meanings across scientific, economic, legal and social assemblages. The protein-modelling AlphaFold2 model is predicted to have significant benefits for disease pathology, targeted therapeutics, antibiotic resistance, climate change, and species extinction (Li and Gilbert 2024). From a PDSD perspective, these benefits will implicate entanglements across educational, economic, and scientific sectors, with some consequences being meaningfully traceable to AlphaFold2's TSD capability. However, AlphaFold2's successes do not necessarily correlate with outputs produced by substantially unrelated TSD processes, using different training data, like commercial generative AI services such as ChatGPT. Grounded in TSD, PDSD may thus provide critical insights into how nebulous understandings of generative AI provoke entanglements (particularly across economic, educational, or governmental arenas) that are traceable as much to economic leverage, hype, ignorance or commercial exploitation of human vulnerability to anthropomorphize AI technologies, as to the actual forms of TSD involved. The fact that AlphaFold2 was awarded a Nobel prize does not mean that many of today's students need to use ChatGPT to stay in touch with technology, or to become better equipped 'learners', or that vulnerable health populations will benefit from LLMs, especially given risks of (health) misinformation built into them (De Angelis et al. 2023). Nuanced attention to the respective TSD functionalities, and user engagement with them, may suggest the opposite (Alshurafat et al. 2024; Shin et al. 2024). Notably, early evidence suggests that the current fervor of generative AI-related economic activity is not based on accurate estimates of productivity increases (Hale 2024).

2 Limitations and future research

Adopting a sequential TSD/PDSD paradigm is one way posthumanist philosophy can be deployed to understand and explore the potentialities and problems of increasingly unsupervised AI models. Using the example of LLMs, we have argued that distinguishing between PDSD and TSD is a critical place to start, as it enables precise identification of how different AI models work and what they are in fact doing, with and without human supervision, and how human vulnerabilities towards anthropomorphism may be shaping enthusiasm for and trust in AI whilst blackboxing the actual operations of different AI technologies. Our contribution here is necessarily limited in scope. We are not able to provide comprehensive or exhaustive accounts of how our principally theoretical intervention may shape understandings of socio-ethical impacts on different groups in diverse settings. Instead we suggest that future posthumanist AI scholarship of this kind could better equip digital subjects, policy-makers and regulatory bodies to identify how and where the capabilities of AI tools in different sectors may be being overstated, and be more attentive to short term and long term risks on diverse user groups that may be unevenly impacted. Practical areas where a sequential TSD/PDSD paradigm facilitates critical interrogation would include the extent and impact of misinformation, reinforcement of bias and discrimination derived from training data, reduction in the development of critical language and development skills (especially in but not limited to education settings), compromises to work quality and integrity across domains, and risks to vulnerable (health) populations. This is a call for posthumanist scholars to take up this important work and to be better equipped to tackle the conceptual and socio-ethical challenges raised by increasingly unsupervised AI machines, and to preserve PDSD as a critical paradigm for understanding wider assemblages that feed into the ontology of particular machinic artefacts.

Acknowledgements We would like to sincerely thank the anonymous reviewers for their valuable suggestions and contributions to a previous version of this manuscript.

Author contributions Sam Cadman and Claire Tanner contributed to the research conception and design. Primary theoretical critique and development were performed by Sam Cadman and Claire Tanner, and technical discipline-specific computer science content was developed and revised by Patrick Pang. The first draft of the manuscript was written by Sam Cadman and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

Consent to participate Not applicable.

Consent to publish Not applicable.

References

Alshurafat H, Al Shbail MO, Hamdan A, Al-Dmour A, Ensour W (2024) Factors affecting accounting students' misuse of chatgpt: an application of the fraud triangle theory. J Fin Rep Account 22(2):274–288. https://doi.org/10.1108/JFRA-04-2023-0182

Anderson MM (2024) AI as philosophical ideology: a critical look back at John McCarthy's program. Phil Tech. https://doi.org/10.1007/s13347-024-00731-1

Arnon I, Kirby S (2024) Cultural evolution creates the statistical structure of language. Sci Reps 14(1):5255. https://doi.org/10.1038/s41598-024-56152-9

Barad KM (2007) Meeting the universe halfway: quantum physics and the entanglement of matter and meaning. Duke University Press, Durham

Bender E, Gebru T, McMillan-Major A, Shmitchell S & Anonymous (2021) On the dangers of stochastic parrots: can language models be too big? Conference on fairness, accountability, and transparency (FAccT '21), March 3–10, 2021, Virtual Event, Canada, 271–278. https://doi.org/10.1145/3442188.3445922

Bennett J (2010) Vibrant matter: A political ecology of things. Duke University Press, Durham

Bhabha HK (2004) The location of culture, 2nd edn. Routledge, London

Bhatia A (2024) When A.I.'s output is a threat to A.I. itself. The New York Times. https://www.nytimes.com/interactive/2024/08/26/upshot/ai-synthetic-data.html. Accessed 30 Dec 2024

Blok V (2024) Materiality versus metabolism in the hybrid world: towards a dualist concept of materialism as limit of post-humanism in the technical era. Philos Technol. https://doi.org/10.1007/s13347-024-00751-x

Bommasani R, Hudson DA, Altman R, Arora S, Sydney von A et al (2021) On the opportunities and risks of foundation models. arXiv.org. https://doi.org/10.48550/arxiv.2108.07258

Bostrom N (2013) Superintelligence: paths, dangers, strategies. Oxford University Press, London

Braidotti R (2013) The Posthuman. Polity Press, Cambridge

Braidotti R (2019) Posthuman knowledge. Polity Press, Cambridge

Burriss SK, Leander K (2024) Critical posthumanist literacy: building theory for reading, writing, and living ethically with everyday artificial intelligence. Read Res Q 59(4):560–569. https://doi.org/10.1002/rrq.565

Capaldi N (1966) Hume's rejection of "ought" as a moral category. J Philos 63(5):126–137. https://doi.org/10.2307/2023901

Caporael L (1986) Anthropomorphism and mechanomorphism: Two faces of the human machine. Comput Hum Behav 2:215–234. https://doi.org/10.1016/0747-5632(86)90004-X

Chen R, Chen C (2022) Artificial intelligence: An introduction for the Inquisitive reader. CRC, NW

Cord F (2022) Posthumanist cultural studies: taking the nonhuman seriously. Open Cultural Stud 6(1):25–37. https://doi.org/10.1515/culture-2020-0138

Dahlin E (2024) And say the AI responded? Dancing around 'autonomy' in AI/human encounters. Soc Stud Sci 54(1):59–77. https://doi.org/10.1177/03063127231193947

De Angelis L, Baglivo F, Arzilli G et al (2023) ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. Front Pub Health 11:1166120

Deckha M (2012) Toward a postcolonial, posthumanist feminist theory: centralizing race and culture in feminist work on nonhuman animals. Hypatia 27(3):527–545

Deleuze G, Guattari FL (1987) A thousand plateaus: capitalism and schizophrenia. University of Minnesota Press, Minneapolis

Demuro E, Gurney L (2024) Artificial intelligence and the ethnographic encounter: Transhuman language ontologies, or what it means "to write like a human, think like a machine." Lang Commun 96:1–12. https://doi.org/10.1016/j.langcom.2024.02.002

Dubrovsky D, Lepskiy V, Raikov A (2022) General artificial intelligence in self-developing reflective-active environments. In: Perko I, Espejo R, Lepskiy V, Novikov D (eds) World Organization of Systems and Cybernetics 18. Congress-WOSC2021 Systems Approach and Cybernetics: Engaging for the Future of Mankind. Springer Nature

Elliott A (2022) Making sense of AI: our algorithmic world. Polity Press, Cambridge

Escobar A (2018) Designs for the pluriverse: radical interdependence, autonomy, and the making of worlds. Duke University Press, Durham

Regulation 2024/1689. Regulation (EU) No 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Fairclough N (2015) Language and power. Routledge, Taylor & Francis Group, Oxfordshire

Ferrando F (2019) Philosophical posthumanism. Bloomsbury Publishing, London

Fisher SA, Howard JW, Kira B (2024) Moderating synthetic content: the challenge of generative AI. Philos Technol. https://doi.org/10.1007/s13347-024-00818-9

Forum OD (2023) Cheat sheet: mastering temperature and top_p in chatgpt apI. https://community.openai.com/t/cheat-sheet-mastering-temperature-and-top-p-in-chatgpt-api/172683. Accessed 3 Dec 2024

Foucault M (2005) The order of things: an archaeology of the human sciences. Routledge, Oxfordshire

Fowler R (2007) Language in the news: discourse and ideology in the press. Routledge, Oxfordshire

Gong L (2008) How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. Comput Hum Behav 24(4):1494–1509. https://doi.org/10.1016/j.chb.2007.05.007

Grassini S, Koivisto M (2024) Understanding how personality traits, experiences, and attitudes shape negative bias toward AI-generated artworks. Sci Reps 14(1):4113. https://doi.org/10.1038/s41598-024-54294-4

Grose J (2024) What teachers told me about A.I. in school. The New York Times. https://www.nytimes.com/2024/08/14/opinion/ai-schools-teachers-students.html. Accessed 6 Feb 2025

Gross M, McGoey L (2022) Routledge international handbook of ignorance studies, 2nd edn. Taylor & Francis Group, Abingdon

Grynbaum MM, Mac R (2023) The times sues OpenAI and microsoft over AI use of copyrighted work. The New York Times. https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html. Accessed 5 Jan 2025

Guattari FL (1995) Chaosmosis: an ethico-aesthetic paradigm. Indiana University Press, Bloomington

Hale C (2024) Using an Ai Pc may actually make users less productive—for now. Tech Radar https://www.techradar.com/pro/using-an-ai-pc-may-actually-make-users-less-productive-for-now. Accessed 9 Feb 2025

Haraway D (2006) A cyborg manifesto: Science, technology, and socialist-feminism in the late 20th century. In: Weiss J, Nolan J, Hunsinger, Trifonas (eds) The international handbook of virtual learning environments. Springer, Netherlands, pp 117–158

Herbrechter S (2022) Palgrave handbook of critical posthumanism. Springer International Publishing, New York. https://doi.org/10.1007/978-3-031-04958-3

Innerarity D (2021) Making the black box society transparent. AI Soc 36:975–981. https://doi.org/10.1007/s00146-020-01130-8

Islam M (2016) Posthumanism: through the postcolonial lens. In: Banerji D, Paranjape M (eds) Critical posthumanism and planetary futures. Springer, New York, pp 115–129

Jecker NS (2024) Extremely relational robots: Implications for law and ethics. Philos Technol. https://doi.org/10.1007/s13347-024-00735-x

Kalpokiene J, Kalpokas I (2023) Creative encounters of a posthuman kind–anthropocentric law, artificial intelligence, and art. Technol Soc. https://doi.org/10.1016/j.techsoc.2023.102197

Kaplan A, Haenlein M (2019) Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Bus Horiz 62(1):15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Köves A, Feher K, Vicsek L, Fischer M (2024) Entangled AI: artificial intelligence that serves the future. AI & Soc. https://doi.org/10.1007/s00146-024-02037-4

Kriegeskorte N, Golan T (2019) Neural network models and deep learning. Curr Biol 29(7):R231–R236. https://doi.org/10.1016/j.cub.2019.02.034

Latham and Watkins LLP (2024) The New York Times Company v. Microsoft Corporation and Ors [Memorandum of Law in Support of OpenAI Defendants' Motion to Dismiss]. (S.D.N.Y. Case 1:23-cv-11195)

Latour B (2002) Morality and technology: the end of the means. Theory Cult Soc 19(5–6):247–260. https://doi.org/10.1177/026327602761899246

Lau F-LA (2024a) Nanonetworks: The future of communication and computation, 1st edn. John Wiley & Sons, New Jersey

Lau M (2024b) Is ChatGPT taking over the language classroom? How language ideologies of large language models impact teaching and learning. Working papers in Applied Linguistics & Linguistics at York 4, 1–11. https://wally.journals.yorku.ca/index.php/default/article/download/36/34. Accessed 9 Feb 2025

Lévi-Strauss C (1966) The savage mind. University of Chicago Press, Chicago

Li B, Gilbert S (2024) Artificial Intelligence awarded two Nobel Prizes for innovations that will shape the future of medicine. NPJ Digit Med 7(1):336. https://doi.org/10.1038/s41746-024-01345-9

Luo Y, Cheong-Iao P, Chang S (2024) Enhancing exploratory learning through exploratory search with the emergence of large language models. arXiv.org. https://arxiv.org/abs/2408.08894. Accessed 4 Dec 2024

Maasen S, Dickel S, Schneider C (2020) Technosciencesociety: technological reconfigurations of science and society. Springer, Switzerland

Mahdawi A (2025) AI-generated 'slop' is slowly killing the internet, so why is nobody trying to stop it? The Guardian https://www.theguardian.com/global/commentisfree/2025/jan/08/ai-generated-slop-slowly-killing-internet-nobody-trying-to-stop-it. Accessed 6 Feb 2025

Maples B, Cerit M, Vishwanath A, Pea R (2024) Loneliness and suicide mitigation for students using Gpt3-enabled Chatbots. Npj Ment Health Res 3(1):4. https://doi.org/10.1038/s44184-023-00047-6

Mariani MM, Perez-Vega R, Wirtz J (2021) AI in marketing, consumer research and psychology: a systematic literature review and research agenda. Psychol Mark 39(4):755–776. https://doi.org/10.1002/mar.21619

Markov I, Kharitonova K, Grigorenko EL (2023) Language: its origin and ongoing evolution. J Intell. https://doi.org/10.3390/jintelligence11040061

Mauthner NS (2019) Toward a posthumanist ethics of qualitative research in a big data era. Am Behav Sci 63(6):669–698. https://doi.org/10.1177/0002764218792701

McCarthy J, Minsky ML, Rochester N, Shannon CE (2006) A proposal for the Dartmouth summer research project on artificial intelligence: August 31, 1955. AI Mag 27(4):12–14. https://doi.org/10.1609/aimag.v27i4.1904

McLean S, Read GJM, Thompson J, Baber C, Stanton NA, Salmon PM (2021) The risks associated with artificial general intelligence: a systematic review. J ExpTheor Artif Intell 35(5):649–663. https://doi.org/10.1080/0952813x.2021.1964003

Millet K, Buehler F, Du G, Kokkoris MD (2023) Defending humankind: anthropocentric bias in the appreciation of AI art. Comput Hum Behav. https://doi.org/10.1016/j.chb.2023.107707

Mitchell M (2024) Debates on the nature of artificial general intelligence. Sci 383(6689):7069. https://doi.org/10.1126/science.ado7069

Nath R, Manna R (2023) From posthumanism to ethics of artificial intelligence. AI & Soc 38:185–196. https://doi.org/10.1007/s00146-021-01274-1

Ooi K, Tan G, Al-Emran M, Al-Sharafi M, Capatina A, Chakraborty A, Dwivedi Y, Huang T, Kar A, Lee V, Loh X, Micu A, Mikalef P, Mogaji E, Pandey N, Raman R, Rana N, Sarker P, Sharma A et al (2023) The potential of generative artificial intelligence across disciplines: perspectives and future directions. J Comput Inf Syst 65:76–107. https://doi.org/10.1080/08874417.2023.2261010

OpenAI (2023) Planning for AGI and beyond. https://openai.com/index/planning-for-agi-and-beyond/. Accessed 18 Jan 2025

OpenAI (2024a) API reference - chat - create chat completion. Retrieved 3 December 2024, from https://platform.openai.

com/docs/api-reference/chat/create#chat-create-temperature. Accessed 18 Jan 2025
OpenAI (2024b) Clarifications on setting temperature = 0. https://community.openai.com/t/clarifications-on-setting-temperature-0/886447. Accessed 18 Jan 2025
OpenAI, Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Florencia Leoni A, Almeida D et al (2024) GPT-4 technical report. https://arxiv.org/abs/2303.08774. Accessed 18 Jan 2025
Pleyer M, Hartmann S (2024) Cognitive linguistics and language evolution. Cambridge University Press, Cambridge
Rau J (2025) Is AI a boom, bubble, or con? Here's what the evidence suggests. Forbes https://www.forbes.com/sites/johnrau/2025/01/03/is-ai-a-boom-bubble-or-con-heres-what-the-evidence-suggests/. Accessed 6 Feb 2025
Rosenberg S (2024) Microsoft's game-changing Super Bowl ad. Axios https://www.axios.com/2024/02/13/microsoft-copilot-super-bowl-ad. Accessed 6 Feb 2025
Russell Group (2023) Russell Group Principles. https://russellgroup.ac.uk/media/6137/rg_ai_principles-final.pdf. Accessed 10 Dec 2024
Sakai M, Sakurai A, Lu S, Olano J, Albrecht CM, Hamann HF, Freitag M (2024) AI-accelerated Nazca survey nearly doubles the number of known figurative geoglyphs and sheds light on their purpose. Proc Natl Acad Sci USA 121(40):e2407652121. https://doi.org/10.1073/pnas.2407652121
Shin D, Koerber A, Lim JS (2024) Impact of misinformation from generative AI on user information processing: how people understand misinformation from generative AI. New Media Soc. https://doi.org/10.1177/14614448241234040
Shoemaker E (2024) Is AI art theft? The moral foundations of copyright law in the context of AI image generation. Philos Technol. https://doi.org/10.1007/s13347-024-00797-x
Shumailov I, Shumaylov Z, Zhao Y, Papernot N, Anderson R, Gal Y (2024) AI models collapse when trained on recursively generated data. Nat 631(8022):755–759. https://doi.org/10.1038/s41586-024-07566-y
Starmer K (2025) PM speech on AI opportunities action plan: 13 January 2025. https://www.gov.uk/government/speeches/pm-speech-on-ai-opportunities-action-plan-13-january-2025. Accessed 9 Feb 2025
Stevis-Gridneff M (2024) Major Canadian news outlets sue OpenAI in new copyright case. The New York Times. https://www.nytimes.com/2024/11/29/world/canada/canada-openai-lawsuit-copyright.html. Accessed 10 Dec 2024
Susman Godfrey LLP (2023) The New York Times Company v. Microsoft Corporation and Ors [Complaint] (S.D.N.Y. Case 1:23-cv-11195)
Svensson J (2023) Artificial intelligence is an oxymoron. AI Soc 38(1):363–372. https://doi.org/10.1007/s00146-021-01311-z
Tahir (2025) DeepSeek R1 explained: chain of thought, reinforcement learning, and model distillation. Medium. https://medium.com/@tahirbalarabe2/deepseek-r1-explained-chain-of-thought-reinforcement-learning-and-model-distillation-0eb165d928c9. Accessed 10 Feb 2025
Thórisson KR, Isaev P, Sheikhlar A (2024) Artificial general intelligence:17th international conference, AGI 2024, Seattle, WA, USA, August 13–16, 2024, Proceedings (1st 2024. ed.). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-65572-2
Togelius J (2024) Artificial general intelligence. MIT Press, Cambridge
Troshani I, Rao Hill S, Sherman C, Arthur D (2020) Do we trust in AI? Role of anthropomorphism and intelligence. J Comp Info Syst 61(5):481–491. https://doi.org/10.1080/08874417.2020.1788473
Trotta A, Ziosi M, Lomonaco V (2023) The future of ethics in AI: challenges and opportunities. AI Soc 38(2):439–441. https://doi.org/10.1007/s00146-023-01644-x
Turing AM (1950) Computing machinery and intelligence. Mind: Q Rev Psychol Philos 59(236):1–28
United States (2023) Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/. Accessed 10 Dec 2024
Uysal E, Alavi S, Bezençon V (2023) Anthropomorphism in artificial intelligence: a review of empirical work across domains and insights for future research. In: Sudhir K, Toubia O (eds) Artificial Intelligence in marketing. Emerald Publishing Limited, Bingley. https://doi.org/10.1108/s1548-643520230000020015
Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention Is All You Need. arXiv.org. https://doi.org/10.48550/arxiv.1706.03762
Verbeek P-P (2011) Moralizing technology understanding and designing the morality of things. University of Chicago Press, Chicago
Wang J, Chen Y (2023) Introduction to transfer learning: algorithms and practice. Springer Nature, Singapore. https://doi.org/10.1007/978-981-19-7584-4
Weil K (2012) Thinking animals: why animal studies now? Columbia University Press, New York
Wolfram S (2023) What is ChatGPT ng…and why does it work? https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-ng-and-why-does-it-work/. Accessed 10 Dec 2024
Wörsdörfer M (2024) Biden's executive order on AI and the E.U.'s AI act: a comparative computer-ethical analysis. Philos Technol. https://doi.org/10.1007/s13347-024-00765-5
Yan D (2024) Posthuman creativity: unveiling cyborg subjectivity through ChatGPT. Qual Inquiry. https://doi.org/10.1177/10778004241231923
Zhang Y (2025) From zero to reasoning hero: how DeepSeek-R1 leverages reinforcement learning to master complex reasoning Hugging Face. https://huggingface.co/blog/NormalUhr/deepseek-r1-explained. Accessed 10 Feb 2025