Artificial Intelligence as Socio-Historical Formation: Histories, Ethics, and Critical Perspectives

Hadil Ayoub
11.03.2026

Introduction

Over the last decade, Artificial Intelligence (AI) has become a highly discussed topic, especially with the recent emergence of generative AI models such as ChatGPT and their widespread use across a variety of fields. A critical examination of the historical development of AI is urgently needed to examine its use beyond the sensationalized narratives and techno-utopian promises that often dominate public discourse. This paper examines key dimensions of AI in order to move engagement with its use beyond purely technical concerns and toward a critical investigation of its historical development, social implications, and potential opportunities. The paper begins with an examination of the term 'AI'; a term that has been increasingly abstracted to describe a wider set of technological developments, in order to highlight issues of the limitations and challenges faced in formulating the language and questions of AI discourse. The paper then proceeds with an exploration of AI's historical trajectory, including some of the foundational developments that made AI possible, such as early computing systems, with particular attention to its emergence during the Cold War era. The paper then turns to two broad categorizations of AI – often conflated or overly polarized in contemporary debates – namely symbolic and connectionist approaches. Finally, it reviews the key ethical, moral, and social concerns that shape contemporary discourses around AI. Ultimately, this paper uses these analytical entry points to argue for a more complex formulation of the questions and concerns surrounding AI. By emphasizing AI as a socio-historical formation shaped by labour regimes, capital, and geopolitical power, the paper invites a critical engagement with the frameworks used to both evaluate and critique it. This analysis contributes towards efforts to develop more grounded and politically informed approaches to AI's ongoing study, use, and discussion.

Defining (Artificial) Intelligence

A key aspect to understanding and discussing AI is establishing a common understanding of what different terms mean. Most important is defining what the "intelligence" in Artificial Intelligence means, in addition to defining terms used in conjecture with AI, such as cognition, consciousness, learning, and knowledge. What follows is an investigation of how some of these terms are defined, and for what purposes. Rather than attempting to produce a new set of definitions, this paper aims to invite a critical look into the use of language in the discussions around the topic.

"Artificial intelligence in the literal sense of the word does not exist and will not exist." These were the words of Soviet Engineer Germogen Pospelov. He advocated for the development of AI, and is considered the driving force behind the development of the Soviet AI program.[1]. For some, the history of AI and associated definitions are entangled, as the discussion around the

[1] Archive of the Russian Academy of Sciences, f. 2, op. 1, d. 1205, p. 60. as quoted in Olessia Kirtchik, "The Soviet Scientific Programme on AI: If a Machine Cannot 'Think', Can It 'Control'?," *BJHS Themes* 8 (January 2023): 111.

topic is critiqued in assuming that AI encompasses a singular comprehensive science or field. Yarden Katz for example argues that AI today "stands for a confused mix of terms—such as "big data," "machine learning," or "deep learning"—whose common denominator is the use of expensive computing power to analyze massive centralized data."[2]

In its simplest definition, Artificial Intelligence refers to the ability of machines, particularly computers, to learn and perform tasks like humans do. Today, many institutions and bodies have developed their own working definitions of AI to support their agendas and work on the topic. The AI Watch Project, which was established by the European Commission's Joint Research Centre, published a report reviewing attempts to define AI. The initiative determined a set of features in common between these attempts. These features were: "perceptions of the environment," "information processing," "decision making (including reasoning and learning)," and finally "achievement of specific goals," which is considered "the ultimate reason of AI systems."[3] The AI Watch Project ultimately adopts the definition of the European Commission's High-Level Expert Group on Artificial Intelligence as their starting point for their work on AI, citing the definition as follows:

> *Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.*[4]

Despite the definitions offered by AI Watch[5] and many others, the definition of AI remains contentious. While many agree on what 'artificial' means, i.e. an action performed by a man-made system (hardware or software), the 'intelligence' part of the definition remains evasive, and a matter of great discussion that involves discourse not only on what is considered intelligence and knowledge, but also on the nature of human thinking, and perhaps, the idea of 'humanness' itself.[6] Namely, it begs the question if human thinking can be replicated to create a man-machine? And can replicating thinking – that is, through simulation and models – be considered thinking in itself? Is intelligence the result of social factors or the result of brain chemicals and neuron activity? Fundamentally, any serious discussion on AI must grapple with these questions.

In an attempt to systematically analyse the problem of defining Artificial Intelligence, Wang argues that intelligence is necessarily anthropocentric and relating to the capabilities of the human mind. Therefore, any definition of intelligence that does not apply to the average person ought to be rejected. Furthermore, in his literature review of definitions of intelligence, he outlines five abstractions of human intelligence which the machine aims to emulate. These abstractions are structure, behaviour, capability, function and principle. He proposes his own

---

[2] Yarden Katz, "Manufacturing an Artificial Intelligence Revolution," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 27, 2017), 2.

[3] Sofia Samoiili et al., "AI WATCH. Defining Artificial Intelligence," *Publications Office of the European Union*, no. KJ-NA-30117-EN-N (online): 8.

[4] Samoiili et al., "AI WATCH. Defining Artificial Intelligence," 9.

[5] The AI Watch Project has now been expanded to become "AI Watch," a portal that presents the work of the Joint Research Committee on Artificial Intelligence.

[6] Syed Mustafa Ali et al., "Histories of Artificial Intelligence: A Genealogy of Power," *BJHS Themes* 8 (January 2023): 6.

definition of intelligence as "the capacity of an information-processing system to adapt to its environment while operating with insufficient knowledge and resources."[7] Daeyeol Lee, author of the book *Birth of Intelligence* argues that "intelligence can be defined as the ability to solve complex problems or make decisions with outcomes benefiting the actor, and has evolved in lifeforms to adapt to diverse environments for their survival and reproduction." Lee ties intelligence necessarily to reproduction and survival, and argues that only when we create a machine that is able to replicate itself and ensure its continued survival can we argue that we created artificial intelligence.[8]

Alternatively, Ali and colleagues argue that there are two strains of theories attempting to define intelligence, knowledge, and AI: one within AI and the other within the social sciences. Where "the former sought to reduce intelligence to a formalism, or the product of formal and data-driven processing. The latter insisted that knowledge is unavoidably social and embodied, requiring experiences and capacities that computers would always lack."[9] This framing of intelligence (the man-machine metaphor) was rejected, for example, by early Soviet researchers in the fields preceding the development of today's AI, such as cybernetics. This was due to two main reasons: the first is the insistence on differentiating creative acts and mechanical acts; the second was an understanding of human thinking as fundamentally "social in nature, engendered not by chemical brain processes but by the collective activity of countless generations of people – a product of socialization, not of neurology." This led them to view American cybernetics as a "reductionist approach to mind and consciousness," and, as Olessia Kirtchik argues, believing that "imitating thinking is not thinking itself."[10]

Others argue that despite an apparent basis in neurology and neuropsychology, AI's development and search for an understanding of machine intelligence have long relied on social phenomena, starting with the 1986 book *The Society of the Mind* by Marvin Minsky – who is considered as one of the founders of the field of AI. Minsky's book looks at intelligence as the product of processes that can be modelled by the logic of a 'society', and up to Frank Rosenblatt's 'perceptron,' a key algorithm in allowing the development of today's machine learning and artificial neural networks, which was modelled after Hayek's ideas of decentralized market structures.[11]

What this multitude of arguments on the nature of intelligence shows is that the question of defining and understanding what is meant by 'intelligence' is central to understanding what is meant when discussing AI. This question has been central to AI's development since the beginning, as often those attempting to develop 'AI' as such, understand the human mind through a simplistic approach; logical, reliable, and therefore, replicable through mathematical

---

[7] Pei Wang, "On Defining Artificial Intelligence," *Journal of Artificial General Intelligence* 10, no. 2 (January 1, 2019): 4, 8, 17.

[8] Annika Weder, "Q&A – What Is Intelligence?," Johns Hopkins Medicine, October 2020, https://www.hopkinsmedicine.org/news/articles/2020/10/qa--what-is-intelligence.

[9] Ali et al., "Histories of Artificial Intelligence," 5.

[10] Olessia Kirtchik, "The Soviet Scientific Programme on AI: If a Machine Cannot 'Think', Can It 'Control'?," *BJHS Themes* 8 (January 2023): 115. See also: Ekaterina Babintseva, '"Overtake and surpass": Soviet algorithmic thinking as a reinvention of Western theories during the Cold War', in Mark Solovey and Christian Dayé (eds.), Cold War Social Science: Transnational Entanglements, Cham: Palgrave Macmillan, 2021, pp. 45–71 and e Roman Abramov, 'Engineering work in the late Soviet period: routine, creativity, and project discipline', Sociology of Power (2020) 32(1), pp. 179-214.

[11] Jonnie Penn, "Animo Nullius: On AI's Origin Story and a Data Colonial Doctrine of Discovery," *BJHS Themes* 8 (January 2023): 19, 30; Matteo Pasquinelli, "How to Make a Class," *Qui Parle* 30, no. 1 (2021): 160.

models. In 1966, Soviet Marxist philosopher Evald Ilyenkov critiqued machine intelligence from the perspective of dialectical materialism. Ilyenkov and other thinkers believed that "intelligence– like culture– is dialogical (that is, intersubjective) and dialectical (driven by contradiction)."[12] Ilyenkov critique states that

> *The Western technical intelligentsia… is therefore entangled in the problem of 'man-machine' because they don't know how to formulate it properly. That is, as a social problem, as a problem of relationship between man and man, mediated by the material body of civilization, including the modern machine technology of production.*[13]

The problem is, then, not just a search for the most precise definition of AI by investigating the technical and technological development of the term – as Wang does in his paper, for example, – but rather an issue of how one 'formulates' questions about AI. This formulation embodies our own understanding of AI, and thus, is indicative of the possibilities (or limits) of the answers we might explore to further examine it. This extends to questions beyond just what intelligence means, but also to inquiries on what we mean by other widely used terms, such as cognition, consciousness, learning, and knowledge. Subject to akin challenges the use of the term 'AI' is facing, these terms, too, are frequently used in the conversation around the topic, but often without clear understanding on what we mean by their use.

These discussions on meanings and formulations in AI research are not new or restricted to the recent wave of generative AI. Taking our understanding of knowledge as an example, in 1988, Marcelo Dascal criticized the epistemological and philosophical approach of AI researchers towards 'knowledge' and its understanding in AI as representational.[14] He instead proposes a pragmatic alternative that shifts the focus from static knowledge representation to the dynamic process of justification. He suggests that true intelligence is best measured by a system's capacity to engage in social communication and defend its conclusions within a changing world. To this end, the goal of AI should not, then, be merely to provide a system with "more knowledge," but to design systems capable of rejecting justifications that do not seem reasonable and selecting the criteria for what is considered relevant in a given context. The "real Turing test," according to Dascal, is a system's ability to provide convincing justifications for its responses.[15]

Returning to Ilyenkov's critique above which urges thinking of AI as "a problem of relationship between man and man, mediated by…modern machine technology of production," I begin the quest for a better formulation of the AI problem by investigating the relationships between attempts to develop machine intelligence and production relations, by critically examining the history of AI's development, before its formulization as 'AI.'

Histories of AI: A Critical Look

The canonical story of the history of AI often begins with the coining of the term by John McCarthy in 1955 during the Dartmouth Summer Research Project on Artificial Intelligence,

[12] Penn, "Animo Nullius," 115.

[13] As cited in Penn, "Animo Nullius," 116.

[14] Dascal, "Artificial Intelligence and Philosophy: The Knowledge of Representation," *Systems Research*, ahead of print, 1989, 39, https://doi.org/10.1002/sres.3850060106.

[15] Dascal, "Artificial Intelligence and Philosophy," 51.

where the Logic Theorist, the program considered to be the earliest AI, was presented.[16] Going back earlier, some start the history of AI with Alan Turing's paper "On Computable Numbers" in 1937, based on which the first mathematical model of a neural network was developed.[17] However, this understanding of AI's history has been described as a "myth" and a "fable".[18] While the term 'Artificial Intelligence' might have been coined by McCarthy in Dartmouth, the history of the attempts to develop and understand the possibilities of machine intelligence goes back much further than McCarthy or Turing. This section of the paper attempts to detangle some of the histories and origin stories of what is called AI beyond the common history-of-technology story.

AI can be seen as serving as "the charismatic megafauna of an entangled set of global and local histories of science, technology and economics."[19] AI's history is both "everywhere and nowhere," and there is a risk in focusing on 'a history of AI' disconnected from the broader histories of other topics which have been long-standing subjects of study, such as industrialization, militarism, colonialism, and capitalism.[20] Marie David traces the development of AI to conceptual and theoretical frameworks that started during the industrial revolution, with the rise of technological and energy development, and – perhaps more importantly – the "rise of the quantification of the world" and the establishment of the field of statistics as a discipline and "a technique for managing society." David views AI as the somewhat natural development to the replacement of physical labour by machines, moving now to the replacement of cognitive and intellectual labour. Two conceptual frameworks make that possible: the development of mechanisms to model thinking using explicit rules as well as the building of machines that can simulate such rules, thus mimicking human thought.[21] Penn similarly traces the history of artificial intelligence to capitalism, not neuropsychology.[22]

Meredith Whittaker, the president of The Signal Foundation and co-founder of the AI Now Institute takes the origin story of AI further back, starting with what is possibly the early ancestor of the computer, the Difference Engine, built in the 1820s and 30s by Charles Babbage. Rather than interrogating the technological contribution per se, Whittaker focuses on the motivations that Babbage had for inventing this engine. She argues that

> *the engines… were envisioned as tools for automating and disciplining labor. Their architectures directly encoded economist Adam Smith's theories of labor division and borrowed core functionality from technologies of labor control already in use. The engines were themselves tools for labor control, automating and disciplining not manual but mental labor.*[23]

---

[16] Rockwell Anyoha, "The History of Artificial Intelligence – Science in the News," Science in the News - Harvard Graduate School of Arts and Sciences, August 28, 2017, https://sites.harvard.edu/sitn/2017/08/28/history-artificial-intelligence/.

[17] Ali et al., "Histories of Artificial Intelligence," 1; Amirhosein Toosi et al., "A Brief History of AI: How to Prevent Another Winter (a Critical Review)," *PET Clinics* 16, no. 4 (October 2021): 5.

[18] Ali et al., "Histories of Artificial Intelligence," 1; Penn, "Animo Nullius," 23.

[19] Ali et al., "Histories of Artificial Intelligence," 1.

[20] Ali et al., "Histories of Artificial Intelligence," 1–2.

[21] Marie David, "AI and the Illusion of Human-Algorithm Complementarity," *Social Research: An International Quarterly* 86, no. 4 (2019): 889.

[22] Penn, "Animo Nullius," 19.

[23] Meredith Whittaker, "Origin Stories: Plantations, Computers, and Industrial Control," *Logic(s) Magazine*, May 17, 2023, https://logicmag.io/supa-dupa-skies/origin-stories-plantations-computers-and-industrial-control/.

This is not a unique understanding of AI's history, as Herbert Simon and Allen Newell, the inventors of the Logic Theorist, described Babbage as "the patron saint of our profession" but the real inventor of the digital computer as being Adam Smith, whose work on economics while not computational in practice is so in principle.[24] Penn sees Simon and Newell's own account as an exception to histographies of AI that often ignore the connection between the development of AI and capitalism.[25] Grzybowski et al. trace the history of the computer and the development of AI not only back to Babbage, but even further back, to the invention of the Jacquard Loom in 1804, a machine where designs for fabrics were contained on punching cards, simplifying the process of manufacturing complex textiles.[26]

Whittaker further argues that Babbage's work – both on labour theories and on engines – were in response to the anxieties the British elite felt about the management of their factories as the abolition of slavery in 1833 approached. The elite grappled with its own set of concerns on how to control the white workers and their frequent rebellions against industrialization, and the need to maintain their production pace. But preceding all of this, Whittaker argues that the theories encoded in the work of Babbage and Smith and their labour politics were not invented by them, but were "prefigured on the plantation, developed first as technologies to control enslaved people."

While industrialization was undoubtedly propelled by technological advancements – such as The Difference Engine – these advancements were inevitably linked to the labour management practices developed on the plantation (such as surveillance, written rules and regulations, regimentation, and a strictly controlled work pace). Babbage's work on labour control and his work on calculating engines cannot be divorced from one another. Rather, they can be read together as attempts to answer the singular question of developing and disciplining labour in service of capital and the empire. One of his justifications for requesting increasing funding from the government was the argument that his engines could improve navigational tables in service of the army. Furthermore, Babbage was incessant on surveillance, so much so that he was willing to sacrifice the feasibility of his engines in favour of ensuring their capacity for surveillance. He further contributed to the automation of surveillance, creating a "tell-tale" clock which recorded workers' presence and absence.[27]

But it is not only Babbage that sought and received funding from the military for his innovation. In fact, one of the aspects least discussed in the canonical history of AI mentioned earlier is who funded and supported this work, particularly in the post-war and Cold War era from 1950s forward. AI can be seen as a product of the Cold War's technological race on cybernetics and computing on one hand, but just as much as being developed within the social science Cold War.[28] The Dartmouth meeting, where the term 'AI' was born, was in fact funded by The Rockefeller Foundation, which at the time was – and arguably continues to be – a key player in American foreign-policy making. Another actor that began investing heavily in AI was IBM, an employee of which was at the Dartmouth meeting, and pitched the idea of automatic coding

---

[24] Herbert A. Simon and Allen Newell, "Heuristic Problem Solving: The Next Advance in Operations Research," *Operations Research* 6, no. 1 (1958): 2.

[25] Penn, "Animo Nullius," 27.

[26] Andrzej Grzybowski, Katarzyna Pawlikowska–Łagód, and W. Clark Lambert, "A History of Artificial Intelligence," *Clinics in Dermatology*, Dermatology and Artificial Intelligence, 42, no. 3 (May 1, 2024): 221.

[27] Whittaker, "Origin Stories."

[28] Ali et al., "Histories of Artificial Intelligence," 7; David Hounshell, "The Cold War, RAND, and the Generation of Knowledge, 1946-1962," *Historical Studies in the Physical and Biological Sciences* 27, no. 2 (January 1, 1997): 224.

to IBM "as a patriotic act" after IBM re-started its military work in response to the Korean War. Additionally, John McCarthy was invited to work at IBM-funded MIT Computation Center, and he was to convince his peers in academia of the ideas of computing.[29] Another important actor was RAND Corporation (Research ANd Development), a project initially founded and funded by the United States Air Force until 1962, and described as a "pure cold war" institution, synonymous with the idea of a "think tank" and the first institution to be called that.[30] It was within RAND that the Logic Theory machine was developed by Newell and Simon, and introduced in the Dartmouth meeting.

While the different participants in the Dartmouth conference (namely McCarthy, Minsky, Simon, Newell, and Shaw) are referred to as 'the founding fathers' of artificial intelligence, Penn points out that these narratives of history never consider where the support for these developments came from. Penn argues that institutional support and their entanglements in capital are seen as extraneous to the story of AI's history, rather than a fundamental part of it. For example, in the mid-1950s, RAND "possessed perhaps the largest computing facility in the world. This and other corporate affordances to proto-AI, however, are treated as incidental, not constitutive. Neither IBM nor RAND figure in the list of AI's 'founding fathers.'"[31] He argues that there has always been an alignment between "mind-as-computer research and state industrial aims" that have often gone unexplored in research on AI's history.[32]

These challenges to normative narratives of AI histories invite a more critical look at the topic beyond the often-represented timeline of AI summers and winters. The vagueness and complexity of what AI itself means challenges the possibility of establishing its history of development. The history of the development of the machine, of cybernetics, and of computation are all parts of the history of AI, and these are parts of the history of industrial, labour, and production relations that cannot be ignored when attempting to understand AI.

Technical Categories of AI: Symbolic vs Connectionist

While there are many ways to categorize AI, historically, there have been two broad categories regarding how to establish a machine that learns, performs tasks, and is thus intelligent: symbolic and connectionist. It is important to distinguish between these two types of AI in order to understand the particular concerns and the drive behind modern AI systems. Both of these paradigms in AI have the previously critiqued assumption that the brain followed a reliable logic that can be mathematically modelled. However, their approach to doing that differed.

The first type is the one that is most familiar and oldest, which is Symbolic (also called classic or traditional AI). This method relies on what today is called simply 'coding', that is, modelling a machine based on an established set of logical rules coded into the machine, which would surpass human thought and tendency to bias and errors.[33] This type of AI is ideal for expert systems, such as medical diagnostics, as it has a set of pre-defined rules that it must follow, imitating the thought process a doctor might have, and following the knowledge available to

[29] Penn, "Animo Nullius," 25.

[30] Hounshell, "The Cold War, RAND, and the Generation of Knowledge, 1946-1962," 239–40.

[31] Penn, "Animo Nullius," 27.

[32] Penn, "Animo Nullius," 26.

[33] David, "AI and the Illusion of Human-Algorithm Complementarity," 889–90.

the AI system.[34] Symbolic AI understands cognition as a set of rules, that can be modelled using a language to communicate with the machine.[35] This AI had many limitations, as it required for each operation to be translated manually into punching cards used for coding the massive early IBM computers.

The second type of AI is connectionist AI, which is based on Artificial Neural Networks (ANN), or what is more often referred to today as 'deep learning.' The work on neural networks and connectionist thinking in AI started in 1957 when psychologist Frank Rosenblatt built an analogue neural network that could learn through trial and error, named the Perceptron.[36] The idea of neural networks is to loosely imitate the way neurons in the brain function, but they are not replicas (or attempts to replicate) the human neural system exactly. Neural nets are organized into layers, each layer containing numerous processing nodes that are connected. Nodes in each layer are connected with nodes in other layers. These layers are 'feed-forward,' meaning that they send data through them in one direction only. A simple neural network model would have an input layer, a hidden layer, and an output layer. When data comes through the input layer, it is assigned a number, and then processed through a hidden layer, where the number is multiplied by a weight that this data is assigned.

The results of all this processing through the different nodes and layers are then shown in the output layer if it meets a certain threshold. When training neural networks, all weights and thresholds are assigned random values. Then, training data is fed through the input layer, and the weights and thresholds are repeatedly adjusted until the outputs are consistent.[37] Initial models like the Perceptron had only a single 'hidden' layer between the input and output layers, however, thanks to a variety of technological developments, such as the graphics processing units (GPUs). Modern networks can have 50 layers of processing, thus achieving the name of 'deep learning.'[38] Hidden layers represent the brain of the neural network. The layers process input data using functions called 'activation functions.' It is in these functions that the 'weight' and 'bias' aspects are included. The illustration below shows a simplification of how a neural network might function.[39]

[34] Haoyi Xiong et al., "Converging Paradigms: The Synergy of Symbolic and Connectionist AI in LLM-Empowered Autonomous Agents," arXiv:2407.08516, preprint, arXiv, October 14, 2024, 2, https://doi.org/10.48550/arXiv.2407.08516.

[35] P. Smolensky, "Connectionist AI, Symbolic AI, and the Brain," *Artificial Intelligence Review* 1, no. 2 (1987): 99–100.

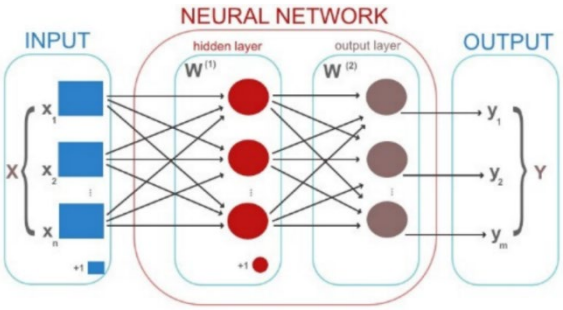
[36] Toosi et al., "A Brief History of AI," 7.

[37] Toosi et al., "A Brief History of AI," 10.

[38] Larry Hardesty, "Explained: Neural Networks," MIT News | Massachusetts Institute of Technology, April 14, 2017, https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414.

[39] Pedro Figueirinhas et al., "Development of an Artificial Neural Network for the Detection of Supporting Hindlimb Lameness: A Pilot Study in Working Dogs," *Animals* 12, no. 14 (January 2022): 3.

Figure 1: Simple representation of a neural network (Figueirinhas et al., 2022).



One issue with neural networks is their issue of interpretability. As the layers of neural networks are exposed to more learning data, they become more complex. It then becomes more difficult to interpret and understand why AI systems make certain decisions, predictions, or other kinds of outputs. In fact, the relationship between a machine-learning model's performance, and the ability to explain and interpret its results is a trade-off, leading to deep-learning models creating a 'black-box' problem, where there is great ambiguity regarding how these models operate and make decisions. This issue is deeply troubling when these models are being applied in sensitive fields such as healthcare and the military. [40] The opacity of these models raises issues regarding accountability, and assigning responsibility for when errors occur. This ambiguity makes AI systems seem unsettling, shrouding them in mystery, and making them difficult to trust.

Despite diverging in important methodological ways, both AI paradigms still hold the same assumption about the brain: that "complex mental processes were susceptible to mathematical formalization, that digital electronic computers were the appropriate tool for that job, and that the product of this enterprise would be an abstract 'language'."[41] In addition, neural networks and symbolic AI have the possibility to converge through hybrid approaches, making use of the strengths of each paradigm.[42]

Since the early 2010s, the resurgence of AI – and particularly connectionist AI – is often characterized as an 'AI revolution,' but in his article titled *Manufacturing the Artificial Intelligence Revolution*, Katz argues that this was not a sudden and organic breakthrough in the history of computer science development. Rather, Katz highlights that there was a push by tech giants to rebrand work on 'big data' as AI, by enlisting the help of academics to reshape academic fields. He notes that many of the leading academics in the AI field hold corporate positions in companies such as Google, Twitter/X, and Uber.[43] Leaning on Hannah Arendt's work, he deems "thoughtlessness" as a most appropriate defining feature of the "so-called age of AI."[44]

Driving this push for rebranding AI is the desire by tech conglomerates to re-position the US as a safe haven for large-scale investments, particularly following the collapse of real-estate investments during the financial crisis of 2007-2008. To meet shareholders' demands for returns, cloud providers, such as Amazon, Google, and Microsoft, pursued budgetary

[40] Pantelis Linardatos, Vasilis Papastefanopoulos, and Sotiris Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," *Entropy* 23, no. 1 (January 2021): 2.

[41] Penn, "Animo Nullius," 23.

[42] Xiong et al., "Converging Paradigms," 1.

[43] Katz, "Manufacturing an Artificial Intelligence Revolution," 1, 4.

[44] Katz, "Manufacturing an Artificial Intelligence Revolution," 18.

"organizational capture" from hospitals, universities, governments, and smaller businesses. To encourage long-term reliance on their proprietary systems, tech giants invested heavily in machine-learning advancements, effectively enclosing and privatizing areas of common online knowledge. Blackwell describes this process as "institutionalized plagiarism."[45]

Selected Issues in AI Discourses

The concerns pertaining to modern AI span a number of topics: from issues of ethics and morality, problems of impact on the workforce and the hidden labour of AI, issues of misinformation and hallucinations; concerns from the perspective of human rights and discriminatory biases and prejudices embedded within the data that feeds and trains AI models; concerns about the climate and broader environmental impact of running AI servers; issues of how to regulate AI (which are related to how to define AI); consequences of military applications and accountability; and lastly, its impact on education, plagiarism, and academic integrity. However, while these concerns are certainly important, situating and grounding these concerns in a proper understanding of what AI is and is not, and the history of its development and deployment remains an essential – and often missing – aspect of these discourses.

In 2024, a group of researchers conducted a systematic literature review and analysis of 309 articles covering articles focusing on the most recurrent ethics issues in these articles. Ethical issues include aspects such as the design, use, and deployment of AI systems; issues of responsibility where AI systems cannot be held accountable for their errors, and challenges of human-machine interactions.[46] Articles were reviewed since the beginning of publishing on the topic of ethics and AI (1989) until 2021. The table below presents a general summary of the findings of the analysis.[47]

Table 1: Summary of ethical concerns in AI in published literature, summarized from the paper by Giarmoleo et al. (2024)

Broad Category	Sub-categories (with percent of papers concerned with this issue)	Detailed themes
Ethical Concerns about the design of AI	Algorithm and Data (>20%)	Data bias and algorithm fairness (12.3%), algorithm opacity (7.8%)
	Balancing AI's risks (>16%)	Design faults and unpredictability (9.2%), military and security purposes (3.8%), emergency procedures (1.9%), AI takeover (1.7%)
	Threats to human institutions and life (>11%)	Threats to law and democratic values (10.6%), transhumanism (0.5%)
	Uniformity in the AI field (2%)	Western centrality and cultural difference (1.3%), unequal participation (0.9%)

[45] Alan Blackwell, Moral Codes: Designing Software without Surrender to AI, Cambridge, MA: MIT Press, 2022, p. 11. As cited in Penn, "Animo Nullius," 31.

[46] Francesco Vincenzo Giarmoleo et al., "What Ethics Can Say on Artificial Intelligence: Insights from a Systematic Literature Review," *Business and Society Review* 129, no. 2 (June 2024): 260.

[47] Giarmoleo et al., "What Ethics Can Say on Artificial Intelligence," 266–75.

Ethical concerns about human-AI interactions	Building a human-AI environment (17%)	Impact on business (10.1%), impact on jobs (5.7%), accessible AI (1.1%)
	Privacy protection (14%)	Privacy threats to citizens (10.5%), privacy threats to customers (3.3%)
	Building an AI able to adapt to humans (9%)	Effective human-AI interaction (6.6%), dialogue systems (2.3%)
	Attributing the responsibility for AI's failures (8%)	AI moral agency and legal status (5.1%), responsibility gap (2.7%)
	Human unethical conducts (2.5%)	Instrumental and perfunctory use of ethics (1.4%), outsourcing human specificities (1.2%)

The ethical concerns reviewed in this paper broadly reflect the themes across many topics. For example, one can see issues of bias and the responsibility gap come up with an AI tool used for medical transcriptions. The tool 'Nabla' was used to transcribe about 7 million medical visits in the US. The model that Nabla is based on, OpenAI's Whisper, has been found to 'hallucinate' – that is, add in entirely new data not existing in the audio into the transcript – in every one of 26,000 transcripts created by one developer to test the tool. Some additional information included mentioning of race where it was not mentioned, admitting to criminal acts, and inventing new types of non-existent medication. In the case of Nabla, which is built off of Whisper, it is not possible to even review the original transcripts to know the extent of the hallucination, as Nabla is built as a medical AI which deletes audio-recordings after transcription to protect patient data.[48]

The issue of the workforce and labour dynamics dimension of AI has multiple aspects. On the one hand, there is the threat of mass layoffs as AI will both replace as well as create employment opportunities. But the gap in skills will mean that these employment opportunities will likely not be accessible for those who suffer layoffs.[49] Additionally, while proponents of AI's widespread deployment promote AI as a way to replace or reduce human labour, AI relies primarily on massive amounts of data that needs to be precisely labelled, filtered, categorized, and annotated.[50] This is labour that precedes the role of the machine, and is performed by human intelligence and effort. These tasks are often invisibilized through outsourcing labour to workers on digital platforms, or to Artificial Intelligence-Business Process Outsourcing companies. These companies take complex tasks and break them down into micro-tasks that they then offer small payments for, often outsourced to Global South countries. [51]

[48] Garance Burke and Hilke Schellmann, "AI-Powered Transcription Tool Used in Hospitals Reportedly Invents Things No One Ever Said - National | Globalnews.Ca," Global News, October 26, 2024, https://globalnews.ca/news/10832303/ai-transcription-medical-errors/.

[49] World Economic Forum, *The Future of Jobs Report 2023* (World Economic Forum, 2023), 6, https://www.weforum.org/publications/the-future-of-jobs-report-2023/.

[50] Kate Crawford and Vladan Joler, "Anatomy of an AI System," *Virtual Creativity* 9, no. 1 (2019): 117–20, https://doi.org/10.1386/vcr_00008_7.

[51] Uma Rani and Rishabh Kumar Dhir, "The Artificial Intelligence Illusion: How Invisible Workers Fuel the 'Automated' Economy | International Labour Organization," International Labour Organization, December 10, 2024, https://www.ilo.org/resource/article/artificial-intelligence-illusion-how-invisible-workers-fuel-automated; Alex Hern, "TechScape: How Cheap, Outsourced Labour in Africa Is Shaping AI English," Technology, *The Guardian*, April 16, 2024, https://www.theguardian.com/technology/2024/apr/16/techscape-ai-gadgest-humane-ai-pin-chatgpt.

This model of devaluing labour connects back directly to Babbage's work and theories on the division of labour, and motivation for the development of the computing machine. Babbage's work focused on reducing the value of labour by reducing the skill level needed to achieve tasks, thus tying value to skill level, and allowing for lower wages.[52] It is precisely this model of devaluing labour that Whittaker and David are concerned about, in what Whittaker describes as 'deskilling.'[53] David is equally concerned on the topic, viewing AI as a deskilling project. With increased reliance on and replacement with AI, the skills and experience needed to perform certain tasks, even creative tasks, might be lost. She argues that "what is jeopardized by AI is a way of learning through ongoing contact with reality, an engagement with the concreteness of the world and in the context of a community of peers. What is jeopardized is expertise itself."[54]

This issue of skill can be observed clearly in the context of education. AI's growing presence in education has raised concerns about plagiarism and academic integrity. This is precisely what Blackwell referred to as AI enabling "institutionalized plagiarism." Tools like OpenAI's GPT models do not, in fact, generate new or unique knowledge or texts, and their output is better understood as a "pastiche" of already existing results, often times from data across the web.[55] But ChatGPT, for example, is not capable of accurately citing its sources when producing material. This means that reliance on tools like ChatGPT will almost always result in content that is somewhat plagiarized. Not only that, but research has shown that students who use ChatGPT to assist in essay writing underperform on neural, linguistic, and behavioural levels in comparison to a control group, in what researchers called "the accumulation of cognitive debt."[56] This raises the question of what exactly are we giving up in exchange for the efficiency and the ease that AI systems offer? And at what point do these trade-offs become too costly and a threat to our basic human cognitive capacities?

Another issue becoming increasingly more alarming is the environmental impact of AI. Not only does AI have the potential for a massive carbon-footprint, but its water consumption has also been found to be much higher than expected. Training large AI models requires significant computing power, which in turn consumes an enormous amount of energy and may emit large amounts of emissions that contribute to climate change. Developing and teaching an AI model emits five times as much carbon dioxide as an American car over its entire life cycle, including production emissions.[57] In addition, data centres containing the servers that help train and run AI models require constant cooling, as they constantly emit high heat. Sending a 100-word email with ChatGPT consumes a 500 ml water bottle. Training GPT-3 consumed the same

---

[52] Charles Babbage, *On the Economy of Machinery and Manufactures*, Cambridge Library Collection - History of Printing, Publishing and Libraries (Cambridge: Cambridge University Press, 2010), 148.

[53] Whittaker, "Origin Stories."

[54] David, "AI and the Illusion of Human-Algorithm Complementarity," 899.

[55] Ali et al., "Histories of Artificial Intelligence," 1.

[56] Nataliya Kosmyna et al., "Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task," arXiv:2506.08872, preprint, arXiv, December 31, 2025, https://doi.org/10.48550/arXiv.2506.08872.

[57] Karen Hao, "Training a Single AI Model Can Emit as Much Carbon as Five Cars in Their Lifetimes," MIT Technology Review, June 6, 2019, https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/.

amount of water as 26 4-member UK households would consume in a year.[58] These environmental demands will inevitably increase as the complexity, size, and deployment of AI systems expand.

Lastly, AI is increasingly being integrated into modern warfare, raising serious ethical and security concerns. Autonomous drones, AI-enhanced surveillance systems, and predictive analytics are transforming military strategies as AI helps analyse vast amounts of data in real-time, including satellite imagery and intercepted communications, allegedly to identify potential targets. However, the ethical implications of delegating life-and-death decisions to machines are highly controversial. The case of Israel's use of AI during its genocide against Gaza is exemplary. Israel has deployed AI-powered systems for intelligence gathering, surveillance, and targeting during its strikes.[59] This use of AI in warfare has been criticized for increasing the risk of civilian casualties and raising questions about accountability in conflicts where decisions are partially made by machines.[60]

Nonetheless, despite the significance of all these concerns, discussions on these matters without the necessary understanding of what AI is, how it came to be, how it is being developed, by whom, and for what purposes, limits both our ability to benefit from AI, as well as our capacity to critique it. While the symptomatic aspects of AI are urgent and pressing, associated assessments and critiques are incomplete without both qualifying what we mean by AI, and in the absence of critical inquiries regarding the systems that brought AI to life and that allow it to prosper.

Conclusion

AI cannot be understood as a purely technical phenomenon; it is a socio-historical formation shaped by labour, capital, and geopolitical imperatives. Its development, from Babbage to modern deep learning, has been intertwined with labour control, institutional funding, and strategic agendas. Symbolic and connectionist approaches embody assumptions about cognition and intelligence, but their application raises complex ethical, social, and environmental concerns. A critical engagement with AI requires attention not only to technical mechanisms but also to historical contexts, labour relations, corporate strategies, and geopolitical power structures. Recognizing AI's entanglement with these broader themes is essential for developing informed, responsible, and politically grounded approaches to its study, deployment, and governance. An important part of that is being critical of the ways that our own questions and ideas about AI are formulated.

Bibliography

Abraham, Yuval. "'Lavender': The AI Machine Directing Israel's Bombing Spree in Gaza." +972 Magazine, April 3, 2024. https://www.972mag.com/lavender-ai-israeli-army-gaza/.

[58] Mark Sellman and Adam Vaughn, "'Thirsty' ChatGPT Uses Four Times More Water than Previously Thought," October 4, 2024, https://www.thetimes.com/uk/technology-uk/article/thirsty-chatgpt-uses-four-times-more-water-than-previously-thought-bc0pqswdr.

[59] Yuval Abraham, "'Lavender': The AI Machine Directing Israel's Bombing Spree in Gaza," +972 Magazine, April 3, 2024, https://www.972mag.com/lavender-ai-israeli-army-gaza/.

[60] Al-Jazeera, "'AI-Assisted Genocide': Israel Reportedly Used Database for Gaza Kill Lists," Aljazeera, April 4, 2024, https://www.aljazeera.com/news/2024/4/4/ai-assisted-genocide-israel-reportedly-used-database-for-gaza-kill-lists.

Ali, Syed Mustafa, Stephanie Dick, Sarah Dillon, Matthew L. Jones, Jonnie Penn, and Richard Staley. “Histories of Artificial Intelligence: A Genealogy of Power.” *BJHS Themes* 8 (January 2023): 1–18. https://doi.org/10.1017/bjt.2023.15.

Al-Jazeera. “‘AI-Assisted Genocide’: Israel Reportedly Used Database for Gaza Kill Lists.” Aljazeera, April 4, 2024. https://www.aljazeera.com/news/2024/4/4/ai-assisted-genocide-israel-reportedly-used-database-for-gaza-kill-lists.

Anyoha, Rockwell. “The History of Artificial Intelligence – Science in the News.” Science in the News - Harvard Graduate School of Arts and Sciences, August 28, 2017. https://sites.harvard.edu/sitn/2017/08/28/history-artificial-intelligence/.

Babbage, Charles. *On the Economy of Machinery and Manufactures*. Cambridge Library Collection - History of Printing, Publishing and Libraries. Cambridge University Press, 2010. https://doi.org/10.1017/CBO9780511696374.

Burke, Garance, and Hilke Schellmann. “AI-Powered Transcription Tool Used in Hospitals Reportedly Invents Things No One Ever Said - National | Globalnews.Ca.” Global News, October 26, 2024. https://globalnews.ca/news/10832303/ai-transcription-medical-errors/.

Crawford, Kate, and Vladan Joler. “Anatomy of an AI System.” *Virtual Creativity* 9, no. 1 (2019): 117–20. https://doi.org/10.1386/vcr_00008_7.

Dascal. “Artificial Intelligence and Philosophy: The Knowledge of Representation.” *Systems Research*, ahead of print, 1989. https://doi.org/10.1002/sres.3850060106.

David, Marie. “AI and the Illusion of Human-Algorithm Complementarity.” *Social Research: An International Quarterly* 86, no. 4 (2019): 887–908.

Figueirinhas, Pedro, Adrián Sanchez, Oliver Rodríguez, et al. “Development of an Artificial Neural Network for the Detection of Supporting Hindlimb Lameness: A Pilot Study in Working Dogs.” *Animals* 12, no. 14 (2022): 14. https://doi.org/10.3390/ani12141755.

Giarmoleo, Francesco Vincenzo, Ignacio Ferrero, Marta Rocchi, and Massimiliano Matteo Pellegrini. “What Ethics Can Say on Artificial Intelligence: Insights from a Systematic Literature Review.” *Business and Society Review* 129, no. 2 (2024): 258–92. https://doi.org/10.1111/basr.12336.

Grzybowski, Andrzej, Katarzyna Pawlikowska–Łagód, and W. Clark Lambert. “A History of Artificial Intelligence.” *Clinics in Dermatology*, Dermatology and Artificial Intelligence, vol. 42, no. 3 (2024): 221–29. https://doi.org/10.1016/j.clindermatol.2023.12.016.

Hao, Karen. “Training a Single AI Model Can Emit as Much Carbon as Five Cars in Their Lifetimes.” MIT Technology Review, June 6, 2019. https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/.

Hardesty, Larry. “Explained: Neural Networks.” MIT News | Massachusetts Institute of Technology, April 14, 2017. https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414.

Hern, Alex. “TechScape: How Cheap, Outsourced Labour in Africa Is Shaping AI English.” Technology. *The Guardian*, April 16, 2024. https://www.theguardian.com/technology/2024/apr/16/techscape-ai-gadgest-humane-ai-pin-chatgpt.

Hounshell, David. “The Cold War, RAND, and the Generation of Knowledge, 1946-1962.” *Historical Studies in the Physical and Biological Sciences* 27, no. 2 (1997): 237–67. https://doi.org/10.2307/27757779.

Katz, Yarden. “Manufacturing an Artificial Intelligence Revolution.” SSRN Scholarly Paper No. 3078224. Social Science Research Network, November 27, 2017. https://doi.org/10.2139/ssrn.3078224.

Kirtchik, Olessia. “The Soviet Scientific Programme on AI: If a Machine Cannot ‘Think’, Can It ‘Control’?” *BJHS Themes* 8 (January 2023): 111–25. https://doi.org/10.1017/bjt.2023.4.

Kosmyna, Nataliya, Eugene Hauptmann, Ye Tong Yuan, et al. “Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task.” arXiv:2506.08872. Preprint, arXiv, December 31, 2025. https://doi.org/10.48550/arXiv.2506.08872.

Linardatos, Pantelis, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. “Explainable AI: A Review of Machine Learning Interpretability Methods.” *Entropy* 23, no. 1 (2021): 1. https://doi.org/10.3390/e23010018.

Pasquinelli, Matteo. “How to Make a Class.” *Qui Parle* 30, no. 1 (2021): 159.

Penn, Jonnie. “Animo Nullius: On AI’s Origin Story and a Data Colonial Doctrine of Discovery.” *BJHS Themes* 8 (January 2023): 19–34. https://doi.org/10.1017/bjt.2023.14.

Rani, Uma, and Rishabh Kumar Dhir. “The Artificial Intelligence Illusion: How Invisible Workers Fuel the ‘Automated’ Economy | International Labour Organization.” International Labour Organization, December 10, 2024. https://www.ilo.org/resource/article/artificial-intelligence-illusion-how-invisible-workers-fuel-automated.

Samoiili, Sofia, M. Lopez Cobo, G. De Prato, F. Martinez-Plumed, and B. Delipetrev. “AI WATCH. Defining Artificial Intelligence.” *Publications Office of the European Union* (Luxembourg (Luxembourg)), no. KJ-NA-30117-EN-N (online) (February 2020). https://doi.org/10.2760/382730 (online).

Sellman, Mark, and Adam Vaughn. “‘Thirsty’ ChatGPT Uses Four Times More Water than Previously Thought.” October 4, 2024. https://www.thetimes.com/uk/technology-uk/article/thirsty-chatgpt-uses-four-times-more-water-than-previously-thought-bc0pqswdr.

Simon, Herbert A., and Allen Newell. “Heuristic Problem Solving: The Next Advance in Operations Research.” *Operations Research* 6, no. 1 (1958): 1–10.

Smolensky, P. “Connectionist AI, Symbolic AI, and the Brain.” *Artificial Intelligence Review* 1, no. 2 (1987): 95–109. https://doi.org/10.1007/bf00130011.

Toosi, Amirhosein, Andrea Bottino, Babak Saboury, Eliot Siegel, and Arman Rahmim. “A Brief History of AI: How to Prevent Another Winter (a Critical Review).” *PET Clinics* 16, no. 4 (2021): 449–69. https://doi.org/10.1016/j.cpet.2021.07.001.

Wang, Pei. “On Defining Artificial Intelligence.” *Journal of Artificial General Intelligence* 10, no. 2 (2019): 1–37. https://doi.org/10.2478/jagi-2019-0002.

Weder, Annika. “Q&A – What Is Intelligence?” Johns Hopkins Medicine, October 2020. https://www.hopkinsmedicine.org/news/articles/2020/10/qa--what-is-intelligence.

Whittaker, Meredith. “Origin Stories: Plantations, Computers, and Industrial Control.” *Logic(s) Magazine*, May 17, 2023. https://logicmag.io/supa-dupa-skies/origin-stories-plantations-computers-and-industrial-control/.

World Economic Forum. *The Future of Jobs Report 2023*. World Economic Forum, 2023. https://www.weforum.org/publications/the-future-of-jobs-report-2023/.

Xiong, Haoyi, Zhiyuan Wang, Xuhong Li, et al. “Converging Paradigms: The Synergy of Symbolic and Connectionist AI in LLM-Empowered Autonomous Agents.” arXiv:2407.08516. Preprint, arXiv, October 14, 2024. https://doi.org/10.48550/arXiv.2407.08516.