SciencesPo
CHAIR DIGITAL, GOVERNANCE AND SOVEREIGNTY

From Rejection to Regulation: Mapping the Landscape of AI Resistance

Can Şimşek

Humboldt Institute for Internet and Society Research Fellow
can.simsek@sciencespo.fr

Ayşe Gizem Yaşar

Assistant Professor (Education), LSE Law School
CREATe Fellow, University of Glasgow
ayse.yasar@sciencespo.fr

May 2025

Abstract

This report is the outcome of a research project which aimed to uncover the reasons and manifestations of resistance to AI. Resistance, as we employed the term, encompasses a broad spectrum of attitudes and actions, including public protests, legal challenges, critical scholarship, grassroots advocacy, and other forms of opposition that contest the development, deployment, and various consequences of technologies commonly categorised as AI. Our research uncovered five widespread causes of resistance: (i) socio-economic concerns including job displacement, (ii) ethical concerns including bias, lack of transparency and harms to human dignity, (iii) safety concerns, (iv) threats to democracy and sovereignty, and (v) environmental concerns. Although not exhaustive, the report surveys six cases of resistance each driven by one or more of the causes that we uncovered: resistance to AI in (i) creative industries, (ii) migration and border control, (iii) medical AI, (iv) higher education, (v) defence and security sectors, and (vi) environmental resistance. Throughout, the report identifies the key actors of resistance, especially civil society and its role in organising citizens' resistance to AI. It also includes examples of how states have formalised specific cases of resistance through regulation, such as the prohibition of purposefully manipulative AI systems in the EU's AI Act. To the best of our knowledge, this is the first comprehensive study of resistance to AI that goes beyond specific disciplines and domains.

Table of Contents

1 Introduction

Technological change, while often hailed for its transformative potential, does not inherently bring social or economic advancement. As Aldous Huxley eloquently formulated in *Ends and Means*, it has often provided “more efficient means for going backwards” (Huxley, 1937, 8), disrupting established systems, deepening inequalities, and provoking resistance. Artificial intelligence (AI), as the latest and perhaps most disruptive technology, exemplifies this paradox, demanding a critical examination of its broader consequences.

This report first situates AI within its historical trajectory, offering context for the emergence of AI and its societal impact. We then identify five central areas of concern driving resistance to AI: the socio-economic implications, ethical issues, safety risks, threats to democracy and sovereignty, and the environmental effects of AI development and deployment. Finally, a series of cases illustrates the diverse and ongoing forms of resistance, showcasing how individuals and communities respond to these challenges posed by AI. Our aim is to provide a focused yet comprehensive and multifaceted examination of resistance to AI. As such, this report also helps clarify the AI technologies and use cases that should be regulated due to their unacceptable or undesirable implications. To the best of our knowledge, this is the first comprehensive study of resistance to AI that goes beyond specific disciplines and domains.

2 Defining Resistance

The term *resistance* evokes a range of connotations. In philosophical traditions, it often embodies a sense of moral dignity or represents anti-totalitarian struggles, while managerial and business perspectives tend to reduce it to a structural or personal flaw which undermines efficiency and innovation (Bauer, 1995). Although technology resistance is often conceptualised through the lenses of consumer non-adoption and resistance to organisational change (Samhan, 2018), such approaches fall short of addressing broader societal resistance. Resistance to AI encompasses profound anxiety about the trajectory of societies and the future of humanity. Therefore, we approach the term “resistance” as a social phenomenon, going beyond the managerial connotations. Instead, we view resistance as an integral phase in the cyclical evolution of technology, where societal responses influence and reshape innovation, rather than merely serving as an antagonistic force.

Resistance, as we employ the term, encompasses a broad spectrum of attitudes and actions, including public protests, legal challenges, critical scholarship, grassroots advocacy, and other forms of opposition that contest the development, deployment, and various consequences of technologies commonly categorised as AI. We conceptualise “resistance to AI” not as opposition to the technology per se but also as a challenge to its societal, economic, and political consequences, recognising that such resistance may even utilise AI systems as a tool. For instance, resistance to AI could involve “data activism” against algorithmic bias[1] (Liminga & Lindgren, 2024) or “everyday resistance,” in which individuals employ routine measures to counteract surveillance and assert autonomy against intrusive technologies (Madison & Klang, 2019). This notion extends to what is referred to as “technologies of resistance to AI” which comprise tools and practices designed to empower individuals and communities to resist or reshape the power imbalances created or reinforced by AI systems. This framework includes methods such as physical resistance, obfuscation, sousveillance, advocacy, privacy-enhancing technologies (PETs), adversarial attacks, and community-organising strategies (Agnew et al., 2023).

Though not exhaustive, this report also examines the translation of resistance into law, regulation and policy, with a particular focus on the European Union (EU) and the United States (US). Ultimately, it aims to explore resistance to AI that stems not from technophobic anxieties but from a principled commitment to ethical and legal values. From this perspective, resistance can be framed as a vital component of democratic governance, crucial for addressing and mitigating emerging risks, and serving as a key mechanism to create, interrogate, refine, and enhance regulatory frameworks.

[1] The issue of algorithmic bias is further discussed under Section 4.2.1 below.

3 Demystifying AI

This section examines the emergence of AI, first as an idea and then as an innovation, against the backdrop of capitalist techno-economic change. We also situate the history of AI within the broader history of capitalism, culminating in the widespread commercialisation of AI applications since the 2010s. In doing so, we trace certain forms of resistance to technological change throughout history, which can provide insights into resistance to AI today. Finally, we focus on the question of how AI should be defined, drawing on the history of AI and the existing legal definitions.

3.1 The Origins of "Artificial Intelligence"

The idea of AI can be discerned even in ancient myths of *automata*, such as the golden maidens of *Hephaestus*, reminiscent of modern-day companion bots, or *Talos* the warrior, evoking parallels with autonomous weapon systems. The first technical strides toward "AI" emerged in the early 19th century, as the First Industrial Revolution reached maturity and the Age of Steam and Railways began to take shape. During this period, the change in production methods such as the mechanization of textile production sparked resistance—famously embodied by the "Luddites" who destroyed machinery to protest against job displacement. Charles Babbage, influenced by the factory system and the mechanisation of production during the Industrial Revolution, documented these advancements in his 1831 work *On the Economy of Machinery and Manufactures*. He speculated that the mechanical principles underlying the punched cards used in Joseph Marie Jacquard's loom for weaving intricate textile patterns could be adapted to solve intellectual problems. This insight not only shaped his conception of the Analytical Engine, a mechanical general-purpose computer conceived in the 1830s, but also marked a pivotal moment in the history of computational thinking, as it bridged the mechanisation of physical processes with the abstraction of logical operations (Essinger, 2004).

Babbage's Difference Engine and Analytical Engine remained unfinished despite receiving £17,000 in government funding—a substantial sum for the era (Freeman & Louçã, 2001, 309). However, the Analytical Engine laid the theoretical foundation for programmable machines and is considered a precursor to modern computing. Building on his work, Ada Lovelace, who is often considered the first computer programmer, envisioned the machine's potential to go beyond mere calculation, suggesting it could be programmed to perform complex tasks, foreshadowing ideas central to AI. The 19th century also saw the development of George Boole's work on Boolean algebra, which provided a mathematical framework for binary systems. This framework later became fundamental to computer science, digital circuit design, and, crucially, the algorithms that underpin AI.

The advancements in the 19th century also laid the groundwork for modern cryptography, eventually contributing to achievements like Alan Turing and others' decrypting of the German Enigma code at Bletchley Park during World War II. After the war, Turing gave his first lecture on "intelligent machines" in 1947 (McCarthy, 2007). While he did not use the term "AI", his 1950 paper, "Computing Machinery and Intelligence," introduced the "imitation game" - also known as the Turing Test, a concept that would become foundational in the study of AI. In this influential work, Turing explored the question, "Can machines think?". He proposed replacing this question with the imitation game that sets aside the question of "thinking" and instead focuses on a machine's capability to *exhibit* behaviour indistinguishable from that of a human (Turing, 1950). Six years after Turing's seminal publication, the term "artificial intelligence" was finally coined during a landmark conference held on the other side of the Atlantic. In 1956, American computer scientist John McCarthy and his colleagues Marvin Minsky, Nathaniel Rochester, and Claude Shannon convened a group of researchers at Dartmouth College in Hanover, New Hampshire, for the "Dartmouth Summer Research Project on Artificial Intelligence." The goal of this gathering was to explore the potential of machines to simulate aspects of human intelligence, such as learning, reasoning, and problem-solving (McCarthy et al., 1955). This vision was articulated in the proposal for the 1956 conference as follows:

> The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.

Today, this event is often referred to as the birth of "artificial intelligence" as a distinct discipline or, at the very least, as the origin of the widespread term "AI."

3.2 The 20th Century Blueprint for Contemporary AI Techniques

In the late 50s and early 60s, advancements in AI were made in two key areas: neural networks for image recognition and natural language processing (NLP).[2] About a year

[2] An Artificial Neural Network is "a computer structure inspired by the biological brain, consisting of a large set of interconnected computational units ('neurons') that are connected in layers. Data passes between these units as between neurons in a brain. Outputs of a previous layer are used as inputs for the next, and there can be hundreds of layers of units. An artificial neural network with more than 3 layers is considered a deep learning algorithm. Examples of artificial neural networks include Transformers or Generative adversarial networks." (Gajjar, 2024). Neural networks originated as an idea to model computer systems on the human brain. Today, they serve as the foundation of many of the widely used AI systems like large language models. Natural language processing "focuses on programming computer systems to understand and generate human speech and text. Algorithms look for linguistic patterns in how sentences and paragraphs are constructed and how words, context and structure work together to create meaning. Applications include speech-to-text converters, online tools that summarise text, chatbots, speech recognition and translations." (Gajjar, 2024)

after the Dartmouth Conference, psychologist Frank Rosenblatt’s team developed the Mark I Perceptron, an early neural network machine designed to classify inputs into two categories, funded by the US Office of Naval Research (Olazaran, 1996). Another AI technique that received US government funding during this era was NLP, which was explored for automatic translation of Soviet press (National Academy of Sciences & National Research Council, 1966). However, military funding agencies eventually ceased supporting this line of research when it became clear that the costs outweighed the strategic benefits (Crevier, 1993, 110). As the limitations and immaturity of the AI techniques of the time became apparent, funding for AI research diminished.

The 60s and 70s saw both progress and setbacks, with periods of reduced funding sometimes referred to as “AI winters.” Meanwhile, the 70s saw the spread of computer systems, with the announcement of the first Intel microprocessor in 1971 marking the beginning of the “information revolution.” (Perez, 2002, 14) A flood of literature emerged, with titles such as the “micro-electronic revolution,” the “computer age,” the “information society,” and the “electronic society” (Dertouzos & Moses, 1979), documenting the awareness of the ongoing socio-technical transformations.

Though not “AI” as understood today, early computerised data systems set the stage for privacy concerns which were later magnified in the advent of AI. In 1971, the Organisation for Economic Co-operation and Development (OECD) initiated a series of reports on information technology. One of these, *Computerised Data Banks in Public Administration: Trends, Policies, and Issues*, highlighted growing concerns over the erosion of privacy due to the increasing use of computerised data systems in public administration (Thomas, 1971). One notable example of resistance to these socio-technical shifts occurred in France. The 1970 SAFARI programme, which sought to centralise data on citizens across various government services, triggered widespread public backlash. On 21 March 1974, *Le Monde* published an article, *SAFARI ou la chasse aux Français*, which exposed the controversial programme and sparked outrage. This public outcry led to the establishment of the CNIL (Commission Nationale de l'Informatique et des Libertés), France's national data protection authority (Leloup, 2024).

By the 1980s, many countries had embraced new information technologies to streamline and enhance their administrative functions. In West Germany, this period also saw the rise of a grassroots resistance movement, reminiscent of the French opposition in the 1970s. This movement arose in response to the 1983 population census, which provoked considerable concern among citizens and led to the *Volkszählungsboykott* (census boycott) which covered a variety of citizen concerns around the introduction of information technologies in the public sector. For example, the introduction of computer-readable identity cards was found particularly troubling, as many feared the potential misuse of personal data and the threat of authoritarian overreach (Hornung & Schnabel, 2009). In response, Germany's Federal Constitutional Court ruled that the census violated the right to privacy, laying down some of the fundamental principles of data protection law (Bundesverfassunsgsgericht, 1983).

That said, compared to other technology-driven concerns of the era—such as nuclear power and biotechnology—that elicited broad public resistance, "AI" did not generate widespread opposition in the second half of the 20th century besides being a concern among academic circles and intellectuals. This was partly due to its complexity and the subtle, often imperceptible nature of its societal impact (Bauer, 1995, 9).

3.3 AI Unleashed: Redefining the 21st Century

The creation of the World Wide Web enabled a slow but steady revolution in AI, paving the way for data mining and significant advancements across various AI fields. As the volume of data and computing power grew, progress in AI accelerated in the early 21st century, leading to breakthroughs in deep learning, reinforcement learning, and robotics. In 2006, James Moor organised another Dartmouth conference marking the 50th anniversary of the 1956 Conference. The summary report of this conference noted that “the field of AI was launched not by agreement on methodology or choice of problems or general theory, but by the shared vision that computers can be made to perform intelligent tasks” and highlighted the differing visions of the participants regarding the future of AI (Moor, 2006).

Among these various approaches, deep artificial neural networks, or deep learning[3] (DL), has come to dominate machine learning (ML). DL emerged as the most effective technique in AI after the success of the AlexNet algorithm in the ImageNet competition in 2012 (Pasquinelli, 2023, 87). AlexNet was developed by Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton - Sutskever went on to co-found OpenAI and Hinton was awarded the Nobel Prize in physics for his work on artificial neural networks. The advancements in neural network techniques led researchers to favour more complex architectures for tasks like image recognition and NLP, gradually shifting away from traditional methods.

[3] Deep learning is “[a] subset of machine learning that uses artificial neural networks to recognise patterns in data and provide a suitable output, for example, a prediction. Deep learning is suitable for complex learning tasks, and has improved AI capabilities in tasks such as voice and image recognition, object detection and autonomous driving.” (Gajjar, 2024)

2016 saw DeepMind’s (a Google subsidiary) AlphaGo defeat the world champion Lee Sedol in the ancient game of Go, showcasing the capability to master strategic challenges beyond those previously demonstrated in games like chess (Hassabis, 2016). As the capabilities of AI became more and more visible, however, concerns on its societal implications also began to spread. For instance, the Future of Life Institute published open letters calling for research into the societal impact of AI (Future of Life Institute, 2015) and for a ban on autonomous weapon systems (Future of Life institute, 2016). These letters were also signed by public figures such as Stephen Hawking and Elon Musk.

The introduction of the transformer architecture in 2017 by Vaswani et al. of Google marked a ground-breaking advancement in natural language processing, enabling further development and scalability of task-agnostic large language models (LLMs) through its self-attention mechanism (Vaswani et al., 2017). The following year, OpenAI built on this foundation by introducing the first Generative Pre-trained Transformer (GPT) model, demonstrating the potential of pretraining on large datasets to enhance the model’s capacity for generating coherent and contextually relevant text (Radford et al., 2018). This approach laid the groundwork for subsequent iterations of GPT models, each leveraging progressively larger datasets to achieve unprecedented language understanding and generation capabilities.

2018 was a turning point not just in AI technology but also AI regulation and became known as the year of the “Techlash” (Bui & Noble, 2021, 163-164). Indeed, journalistic and scholarly research had already demonstrated algorithmic bias, but the Cambridge Analytica scandal[4] which erupted in March 2018 made the expansive and

[4] For more information on the Cambridge Analytica scandal, see Hu, 2020.

dubious data collection and profiling activity on Facebook apparent to the wider public. The Guardian-New York Times investigation uncovered how Facebook allowed a third-party developer to access user data through a survey on Facebook. This "Techlash" triggered a wide array of corporate and public policy responses, which are examined in the relevant sections of this report. Corporate actors, in particular those behind state-of-the-art AI models and applications, began to introduce "ethics councils" and "ethical codes", which were met with scepticism in the academic and NGO communities. Governments and international organisations, meanwhile, started working on their own ethical codes and guidelines. For example, the EU formed a High-Level Expert Group on AI (dubbed "AI HLEG") in 2018, bringing together "AI experts" mainly from universities, government departments, non-profit organisations and research labs of private companies. AI HLEG was tasked with drafting "Ethics guidelines for trustworthy AI", published in 2019 (European Commission, 2019). Similarly, OECD first published its AI principles in 2019, later updated in 2024 (OECD, 2025). These documents converge on several principles distilled from AI ethics and safety research, such as fairness, transparency, human oversight, safety and privacy.[5]

The EU's Ethics Guidelines for Trustworthy AI later formed the basis of the world's first comprehensive AI regulation, the AI Act, which came into force in August 2024.[6] In this framework, the Commission adopted a risk-based approach, categorising AI systems into "unacceptable risk," "high risk," and "low risk" tiers,

[5] Some of these principles are further discussed in Section 4 below.

[6] The full title of the AI Act is *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*. For a brief history and an overview of the AI Act, see Smuha & Yeung (2025).

modelled on the EU's well-established product safety regime.[7] The majority of obligations under the AI Act concern the high-risk category, including remote biometric identification systems, risk-assessment systems used in border control and migration, AI systems used to evaluate the eligibility of natural persons for essential public assistance benefits and services, and AI systems used to assist judicial authorities.[8] The AI Act mandates that providers of "high-risk" AI systems ensure compliance with its requirements, which largely concern the design and deployment of AI systems, including the automatic recording of events (logs) over the lifetime of the system (Article 12), transparency to enable deployers to interpret a system's output and appropriate use (Article 13), and the integration of human oversight mechanisms (Article 14). These provisions are further discussed in this report where relevant.

The AI Act prohibits AI systems posing an unacceptable risk to "Union values," albeit with some exceptions. The prohibition applies to a closed list of AI systems under Article 5 of the AI Act. It includes AI systems deploying purposefully manipulative or deceptive techniques; real-time remote biometric identification systems in publicly accessible spaces used for purposes of law enforcement; and emotional recognition systems in the areas of workplace and education institutions.

Finally, as further discussed below, the AI Act features a rigorous governance regime for "general-purpose AI models" as a separate category to address AI *models* (as opposed to systems) which can serve as the basis of AI systems, and which have a large variety of use cases.

[7] The product safety framing has been criticised, in particular vis-à-vis a rights-based approach which AI HLEG had previously suggested (See Almada and Petit, 2023, 12).

[8] There are two types of high-risk AI systems in the AI Act: (i) AI systems that are, either as a standalone product or as the safety component of a product, already subject to the EU's product safety regime under the Union harmonisation legislation listed in Annex I, and (ii) AI systems listed in Annex III.

3.4 Foundation Models and Generative AI as a Paradigm Shift in Artificial Intelligence

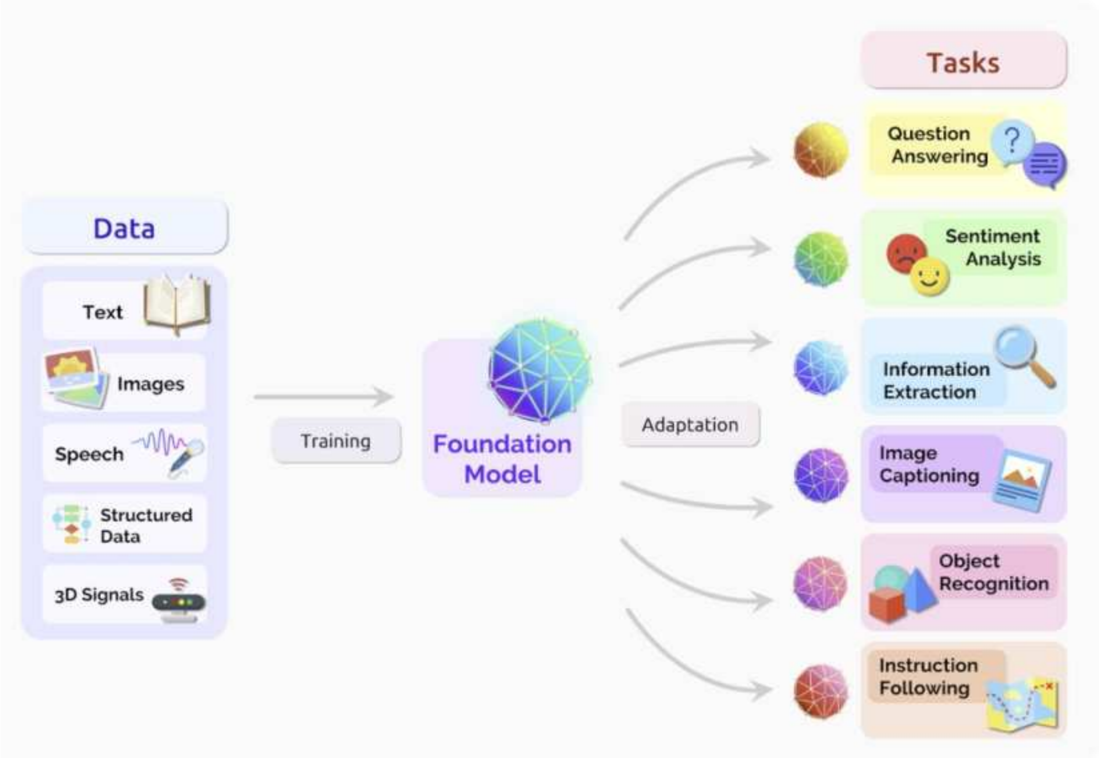
In November 2022, while the draft AI Act was still under negotiation, OpenAI released its consumer-facing generative AI chatbot, ChatGPT, built on the GPT-3.5 language model (OpenAI, 2022). Within just two months, it amassed over 100 million users, making it the fastest-growing consumer application to date (Milmo, 2023). Differing from discriminative models, which focus on distinguishing between classes or types of data and are particularly useful for classification tasks, generative AI refers to AI models and systems specifically designed to create new data by learning and replicating the patterns, structures, and characteristics identified in the training data (Google ML Education, 2022). For instance, discriminative models can determine whether an image contains a cat or a dog, whereas generative AI can create a cat image based on a given prompt. As explained by the Norwegian Consumer Council:

> Generative AI models work by analysing large amounts of information to predict and generate the next word in a sentence, feature of an image, etc. This is done by detecting patterns in and relationships between data points in the training data, which in turn allows the system to replicate similar patterns to generate synthetic content, for example a piece of writing, music, or a video clip. This process can also be described as a complex 'mash-up' of content from the data the system was trained on. In other words, they are predictive models that are trained to "connect the dots" between data points in existing content to generate synthetic content (Forbrukerradet, 2023).

Current generative AI tools are capable of producing a wide range of outputs, including text, images, computer code, music, videos, and even structural designs for 3D printing (G'sell, 2024). Building on these generative capabilities, AI systems have advanced toward multimodal integration, which allows them to process and combine diverse inputs such as text, images, audio, and video (UNESCO, 2025).

Central to generative AI is what is known as "foundation models". Initially gaining prominence in NLP before expanding to broader domains, foundation models can serve as a general-purpose backbone for a wide range of applications. These models are typically trained on massive datasets using self-supervised or unsupervised learning techniques, and they can be fine-tuned for downstream tasks (CRFM & HAI, 2021).

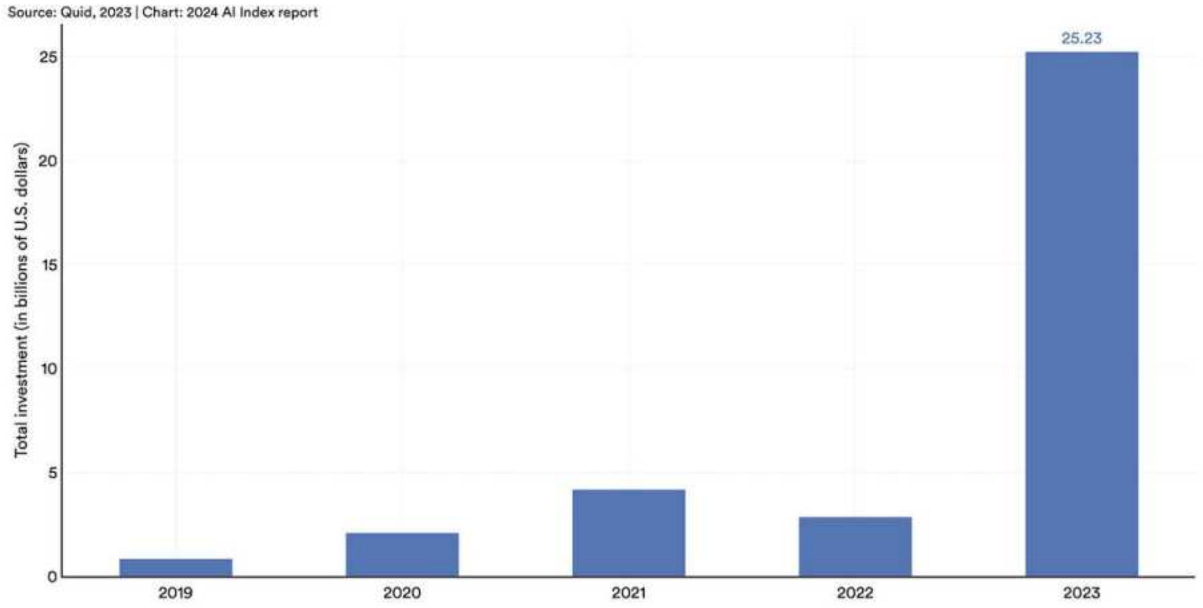
Figure 1: Foundation Models



Source: CRFM & HAI, On the Opportunities and Risks of Foundation Models, 2021

Widespread adoption of GPT chatbots and the rapid expansion of generative AI applications have ignited public debate over the risks and implications of these technologies while attracting unprecedented investment, as illustrated in Figure 2.

Figure 2: Private investment in generative AI, 2019-2023



Despite a decline in overall AI private investment last year, funding for generative AI surged, nearly octupling from 2022 to reach $25.2 billion. Major players in the generative AI space, including OpenAI, Anthropic, Hugging Face, and Inflection, reported substantial fundraising rounds.

Source: Stanford University Human-Centered Artificial Intelligence, 2024

Upon the launch of ChatGPT, EU's draft AI Act was amended to capture AI systems which can be adapted to a large number of use cases, and cannot be easily categorised within the risk-based structure. This separate category of AI models is called "general-purpose AI models" (GPAI) and is defined as:

> an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except

> AI models that are used for research, development or prototyping activities before they are placed on the market.[9]

Chapter V of the AI Act concerns GPAI and introduces separate obligations for providers of GPAI which focus on drawing up and keeping up-to-date the technical documentation of the model, including its training and testing process and the results of its evaluation.[10]

"Agentic AI" has emerged as the latest trend among AI developers and providers. A growing number of companies are investing in the development of general-purpose AI agents—systems capable of autonomous planning, action, and delegation with minimal human oversight. These sophisticated "agents" have the potential to manage complex, long-term projects, thereby unlocking significant benefits while simultaneously introducing new risks such as malicious use or hijacking of agents (Bengio et al., 2025). These concerns will likely be amplified as AI is integrated with robotics and has a greater impact on the physical world.

On a final note, the term "frontier model" is increasingly used among industry professionals and policymakers to refer to a subcategory of highly advanced foundation models, even though there is no consensus or official definition of the term. While no specific criteria exist for classifying a model as a frontier model, computational power is sometimes used as a proxy for their advanced capabilities (G'sell, 2024). These models are characterised by capabilities that surpass those of existing models, often associated with significant risks to public safety and global security, including so-called "existential risk". The term remains contested and is also

[9] See the AI Act, Article 3(63).

[10] See Article 53 and Annex XI. The AI Act also introduces additional requirements for providers of GPAI "with systemic risk" (Article 55).

considered to be a form of distraction that takes attention away from real-life, immanent harms that AI systems pose (Helfrich, 2024).

3.5 The Quest for a Definition: The AI is dead, long live the new AI!

Defining AI is a challenging endeavour today, primarily because the term itself is ever-evolving. There are, however, two important factors that must be borne in mind for purposes of defining AI. The first is the fact that AI systems are artefacts *built by humans and with human creativity* - they are not independent moral agents. The second is the need for a legal definition: if "AI" is to be regulated, it has to be meaningfully defined to set the scope of regulatory intervention.

Although often perceived as the antithesis of human intelligence, AI is, in fact, a testament to human ingenuity, embodying the collective knowledge, creativity, and insights of countless individuals. Far from being magical problem solvers or conscious beings, AI systems depend on human input and interaction, operating through algorithms intricately woven into assemblages of hardware, data structures, memory, and human behavior. In essence, contemporary AI retains a conceptual link to Wolfgang von Kempelen's 18th-century Mechanical Turk—a chess-playing automaton covertly controlled by a human (Standage, 2002). AI as a socio-technical construct, rooted in human ingenuity, is underscored by this historical artifact, which also inspired Amazon's modern crowdsourcing platform of the same name. As AI systems evolve, they are becoming increasingly complex and opaque, gaining seemingly autonomous capabilities and, at times, exhibiting unpredictable behavior that challenges our ability to fully comprehend or control them. Although these systems are not always directly controlled by a hidden human operator, it is important to remember that humans are always involved at some stage of their development, deployment, upkeep and oversight. After all, AI is an artifact created and utilised by humans, often mistakenly perceived as a moral agent (Bryson, 2021).

To regulate the myriad concerns raised or amplified by AI systems, policymakers and lawmakers have crafted comprehensive definitions which are technology-neutral and that can stand the test of time, at least to some extent. Although these definitions are occasionally revisited, particularly in light of advancements such as foundation models, they have, for the moment, converged toward an "official" consensus on what constitutes an "AI system". The definition used by the OECD has become one of the most widely adopted (Russell et al., 2023).

> An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

The definitions used in Council of Europe's Framework Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law, as well as in the EU's AI Act mirror this definition. This definition has also guided the research underlying this report, providing an inclusive framework that encompasses diverse use cases which have sparked resistance.

4 Resistance to AI: Key Concerns

This section explores the key concerns underlying resistance to AI. We identified five types of concerns to which cases of AI resistance are often connected: (i) far-reaching socio-economic implications of AI, encompassing shifts in labour markets, widening inequality, and the redistribution of power and opportunity, (ii) ethical concerns, especially bias and discrimination in AI output and decision-making, (iii) safety concerns centred around the risk of harm that AI systems pose, (iv) threats to democracy and sovereignty, (v) the environmental impact of AI systems, including their resource-intensive development and deployment. These five key types of concerns are briefly introduced and discussed below.

4.1 Socio-Economic Concerns

AI is looking increasingly like the source of a new technological revolution (Vöpel, 2024) which implies socio-economic upheaval. In capitalist economies, technological change occurs through cycles or “great surges” (Perez, 2002, 23), unleashed by technological revolutions. A technological revolution is “a powerful and highly visible cluster of new and dynamic technologies, products and industries, capable of bringing about an upheaval in the whole fabric of the economy and of propelling a long-term upsurge of development.” (Perez, 2002, 8) Each technological revolution, marked by a big bang technological moment, leads to the opening up of a new techno-economic paradigm, which is:

> a best-practice model made up of a set of all-pervasive generic technological and organizational principles, which represent the most effective way of applying a particular technological revolution and of using it for modernizing and rejuvenating the whole of the economy. When generally adopted, these principles become the common-sense basis for organizing any activity and for structuring any institution. (Perez, 2002, 15)

According to innovation scholars Freeman, Louçã and Perez, capitalism in Europe and the US has witnessed five techno-economic paradigms (Perez, 2002, 18; Freeman & Louçã, 2001, 141): (i) the Industrial Revolution initiated by Arkwright's cotton mill (1771), (ii) the Age of Steam and Railways initiated by the test of the 'Rocket' steam engine for the Liverpool-Manchester railway (1829), (iii) the Age of Steel, Electricity and Heavy Engineering initiated by the opening of the Carnegie Bessemer steel plant in Pittsburgh (1875), (iv) the Age of Oil, the Automobile and Mass Production initiated by the first Ford Model-T (1908), and (v) the Age of Information and Telecommunications initiated by the Intel microprocessor (1971). We can now add a sixth paradigm to this list: the Age of AI. The launch of ChatGPT in November 2022 was likely the big bang moment in the sense described by Perez which revealed the capabilities of AI to the masses.

While technological revolutions create new jobs in novel industries, they also result in the replacement of jobs with automation and the displacement of workers, upending the occupational landscape. Furthermore, the emerging techno-economic paradigm of AI is characterised by a transfer of economic power from the ICT paradigm, as some of the largest corporations which became the "tech giants" of the previous paradigm are also among the most prominent and powerful actors of the AI paradigm, as we discuss further below.

4.1.1 Labour upheaval

Job displacement is a major concern associated with automation in general and the rise of AI in particular. These concerns are hardly new: the best-known case of job displacement and ensuing labour uprising is the case of the Luddites, “textile workers whose livelihoods were threatened by the introduction into the work place of new machinery [who] set out to destroy the offending technology” in the early 19th century as a result of the Industrial Revolution (Linton, 1992, 539). While the term Luddite came to denote “mindless, reactionary opposition to technological improvement”, labour historians have shown that this is a misrepresentation of the Luddites (see, in general, Linton, 1992). Their uprising was not against technology as such, but against job displacement in the textile industry. The history of capitalism features plenty of examples of job displacement fuelled by creative destruction, such as the displacement of agricultural jobs due to mechanisation, or secretarial jobs in the advent of information and communication technologies.

As industries increasingly adopt AI-driven technologies, many roles traditionally performed by humans face the risk of obsolescence (OECD, 2024). Proponents of automation assert that while AI may displace certain jobs, it concurrently generates new, often higher-skilled roles, leading to a net positive transformation of the labour market. In response to this position, reskilling and upskilling workers in roles vulnerable to automation have become the central policy approach around the world, aiming to equip individuals with the skills needed to transition into emerging opportunities (World Economic Forum, 2020).

However, the quality of jobs emerging within an AI-driven economy remains a contentious issue, as the proliferation of algorithmically mediated forms of work has, in many cases, contributed to a decline in overall labour standards. While AI has driven the expansion of high-skilled roles in fields such as computing and robotics, it has also created a growing layer of low-skilled, poorly paid jobs (Howcroft & Bergvall-Kåreborn, 2018). These roles often involve tasks facilitated by app-based and crowd-work platforms, including delivery, customer support, transportation, storage, as well as training of AI systems. A typical example is Amazon's Mechanical Turk platform, where "requesters" post tasks such as image identification, product description writing, or survey completion. Workers, known as "Turkers" or crowdworkers, select and complete these tasks for a fee determined by the requester. While certain jurisdictions have introduced various degrees of labour protection for platform workers[11], remote tasks like AI training can be offshored, bypassing regulatory safeguards. For example, OpenAI reportedly outsourced data labelling and filtering tasks to an "ethical AI" company that employs Kenyan workers for less than 2 dollars per hour (Perrigo, 2023). Furthermore, AI is enabling novel types of worker control and management. Performance tracking is a case in point. In 2019, Amazon came under scrutiny for automated tracking of warehouse workers and subsequent "productivity firings" for failing to reach Amazon's productivity goals as reported in the Verge. (Lecher & Castro, 2019) These workplace surveillance mechanisms are becoming increasingly commonplace as investigated in a recent project led by Austria-based non-profit Cracked Labs.[12] A recent case study under this project also includes examples of

[11] For example, the UK Supreme Court ruled in *Uber BV and others v Aslam and others* (2021) that the claimant Uber drivers were "workers" as opposed to "self-employed", which gives drivers certain rights under UK labour laws, like minimum wage and paid annual leave.

[12] More information on the project can be found at https://crackedlabs.org/en/data-work/project

worker resistance to workplace monitoring technologies. For instance, employee backlash resulted in the removal of “workplace analytics technology” which tracked worker movements at the Daily Telegraph and at Barclays, both in the UK. (Christl, 2024, 13; BBC, 2020)

Another example of AI-based worker control is the automated allocation of labour among platform workers, such as drivers on ride hailing apps and delivery workers. Such algorithmic management of work on digital platforms has been strongly criticised. For example, Varoufakis describes these systems as “algorithms without empathy.” These algorithmic systems allocate work to employees in transport, delivery, and warehousing sectors to optimise performance by exerting pressure on workers. Furthermore, those who do not perform as well as the best workers can be dismissed by the algorithms, leaving them without access to a human who can explain why they were dismissed (Varoufakis, 2023, 82).

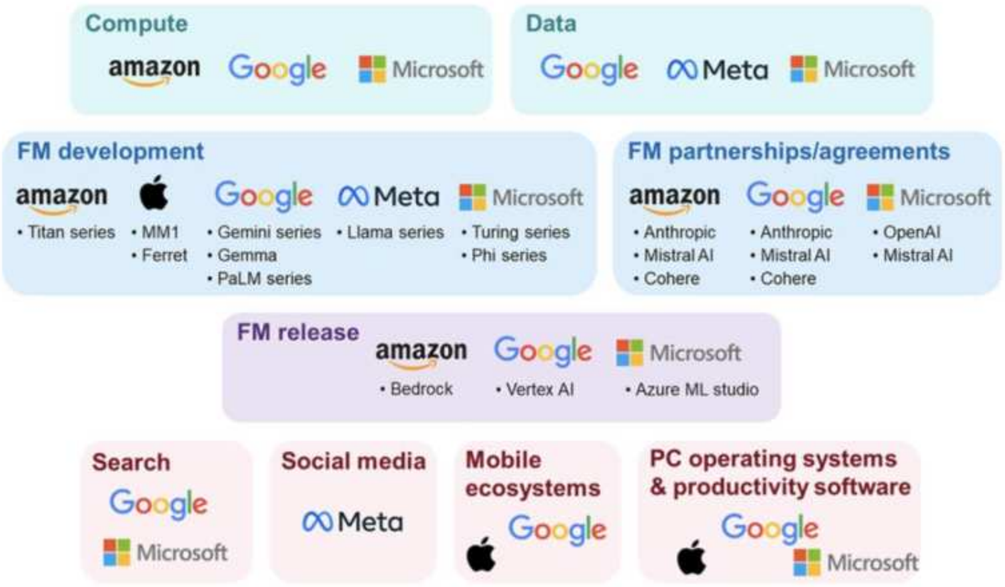
4.1.2 Concentration of economic power

The advancements in AI technologies appear to disrupt established economic, social and political paradigms while reconfiguring power dynamics on a global scale. One of the early voices attempting to theorise this transformation was Shoshana Zuboff. In her seminal work, *The Age of Surveillance Capitalism*, Zuboff argues that capitalism has entered a new phase, marked by the commodification of personal data. According to Zuboff, “behavioral surplus” extracted through pervasive surveillance, forms the foundation of “big tech’s” business model, fuelling its predictive and manipulative capabilities. She describes this shift as a “coup from above,” asserting that it constitutes a fundamental seizure of the free market and human autonomy, urging society to resist its encroachment (Zuboff, 2019). Another prominent voice employing the term “coup” is Marietje Schaake. In “the Tech Coup,” Schaake underlines that AI companies are not only tailoring AI rules and grants through intense lobbying efforts but sometimes even by threatening to cease their operations if their demands are not met. She gives the example of Sam Altman, who urged the U.S. Congress to impose additional regulations on AI while simultaneously threatening to cease OpenAI’s operations in Europe if EU regulations proved too stringent. Additionally, Schaake points out that these companies are leveraging generative AI models to amplify their influence, generating lobbying letters and talking points at scale to shape policy outcomes in their favour (Schaake, 2024, 169-170). Finally, she also warns of a genuine risk of tyranny emerging from corporate technology governance, which threatens to undermine democratic processes and public accountability (Schaake, 2024, 252).

The issue of economic and political power being concentrated in the hands of a few dominant tech companies is also examined by 2024 Nobel Prize winners Daron Acemoğlu and Simon Johnson. In their recent book, *Power and Progress*, they argue that AI-based automation has contributed to the emergence of a “two-tiered society,” reminiscent of the dystopia depicted in H.G. Wells’s *The Time Machine*. They contend that the focus should shift from an obsession with machine intelligence to an emphasis on “machine usefulness,” which prioritizes the extent to which technology serves and benefits humanity (Acemoğlu & Johnson, 2023, 305) Going even further than Acemoğlu and Johnson, Varoufakis argues that the current economic system is not capitalism but rather “techno-feudalism.” He argues that “cloud capital” fundamentally diverges from classical capital because it reproduces itself without the need for waged labor, relying instead on users who contribute to its reproduction without compensation (Varoufakis, 2023, 79-80). Highlighting stark disparities, Varoufakis notes that while workers in traditional industries typically collect around 80% of their companies' income as salaries, employees in Big Tech receive less than 1% of their firms' revenues (Varoufakis, 2023, 84).

The current landscape of commercial AI development is further characterised by a power transfer from the previous ICT techno-economic paradigm into the emerging AI techno-economic paradigm. As Figure 3 below shows, Amazon, Google, Microsoft, Meta and Apple are increasingly active across the foundational model value chain, either through technologies developed in-house or through partnerships with novel providers.

Figure 3: Examples of firms across the AI value chain



Source: UK CMA, 2024

As such, the emergence of generative AI risks further strengthening "Big Tech"s economic power, as its development relies on talent, exclusive datasets, and computational power—resources largely controlled by major firms like Amazon and

Microsoft. Some of these firms also benefit from an early-mover advantage: as discussed above, Google developed the first GPT model which laid the foundations of generative AI (Yasar et al., 2024, 13-15). Moreover, owning major news outlets, social media platforms and lobbying power, the same large tech companies act as "problem brokers" in debates over public policies and regulations, pushing for the prioritisation of selected issues like generative AI regulation, whilst sidelining others (Khanal et al., 2024). Similarly, they have been criticised for steering the public discourse away from immediate concerns like algorithmic bias to future or speculative, "existential" risks (Ryan-Mosley, 2023). Meanwhile, bold claims about AI systems being omnipotent and solving virtually any problem —often without solid evidence—fuel what's been aptly dubbed "AI snake oil" (Hulette, 2021; Narayanan & Kapoor, 2024). Some of the discourse on alarming existential threats in the media can also function as a form of advertisement, reinforcing the perceived importance and value of this "snake oil".

4.2 Ethical concerns

"Ethical concerns" surrounding AI are manifold and have given rise to myriad studies, guidelines and political debates in the past decade. While a comprehensive study of ethical concerns is beyond the scope of this report, it is nonetheless helpful to delineate some of these concerns which underlie specific cases of resistance to AI. Concerns which are pertinent to this report include (i) bias and discrimination, (ii) opacity, (iii) surveillance, (iv) lack of accountability, and (v) harm to human dignity and autonomy. These concerns are interlinked in different ways, and cases of AI resistance that we study in this report like resistance to various military use cases are sometimes cross-cutting and emerge in reaction to a combination of these concerns.

4.2.1 Bias and discrimination

AI systems are often trained on data about the past and may - and often do - reproduce biases which exist in the training data, leading to discriminatory outcomes. Several high-profile examples include Amazon's "sexist AI" hiring tool (BBC, 2018) and OpenAI's image generator Dall-E exhibiting gender and racial bias (Ananya, 2024). Other examples have emerged in public use cases. For example, the COMPAS tool, developed by a commercial entity and used to assess recidivism in the US justice system was one of the earliest examples of algorithmic bias that came under public scrutiny. ProPublica, a news outlet, analysed the tool based on criminal records in Broward County, Florida, where the tool was actively used in pretrial release decisions. It found that black defendants "were often predicted to be at a higher risk of recidivism than they actually were" whereas white defendants were often predicted to be at a lower risk of recidivism than they actually were (Larson et al., 2016). While the inner workings of COMPAS were never disclosed to the public, the case drew attention to how existing racial injustice in a society can seep into algorithmic assessment tools when they are trained on past data.

More recently, a legislation signed into law in Louisiana in March 2024 has made parole eligibility contingent on the TIGER (Targeted Interventions to Greater Enhance Re-entry) risk-assessment model. The TIGER algorithm relies almost entirely on static, pre-incarceration variables—age at first arrest, prior revocations, employment history—factors tightly correlated with race and class, and omits post-conviction conduct. Under the statute, anyone labelled "moderate" or "high" risk is summarily barred from a hearing, excluding roughly half of the prison population from parole consideration. The denial of a parole hearing to 70-year-old Calvin Alexander, a visually impaired inmate with an exemplary record, when TIGER assigned him a moderate-risk score, illustrates how such algorithmic gating reproduces historical inequities rather than assessing current risk (Webster, 2025).

Biased outcomes might materialise or be exacerbated for reasons other than biased training data, like developers' or deployers' own biases, lack of meaningful oversight, or due to the incompleteness of the training data. Developers' and deployers' pre-existing biases towards certain ethnic backgrounds and nationalities have been reported in the context of AI systems used by governments. The Dutch childcare welfare benefits scandal is a case in point. A series of algorithmic systems used by the Dutch government for risk assessment of individuals for purposes of fraud detection in welfare, social security and tax systems came into public light between 2018-2019. One such algorithmic system was used to assess risk of inaccuracies in childcare benefit applications, which ultimately left thousands of parents and caregivers falsely accused of benefit fraud.[13] Investigations of the tax authorities' algorithmic risk assessment system revealed that the Dutch tax authorities "kept secret blacklists of people for two decades, which tracked both credible and unsubstantiated 'signals' of potential fraud" and that having "non-Western appearance" was a reason for blacklisting individuals, who had no knowledge of whether and why they were blacklisted (Heikkilä, 2022). Being on the blacklist contributed to the risk score (ibid). Furthermore, not being Dutch increased the risk score of individuals in general, which reflects an assumption on the part of the designers of the system that people of other nationalities were more likely to commit fraud (Amnesty International, 2021, 22). This scandal ultimately culminated in the resignation of the Dutch government (Holligan, 2021).

[13] For more information on the Dutch childcare benefit scandal, see Amnesty International, 2021.

4.2.2 Opacity

Opacity can significantly undermine accountability and fairness in the context of AI, as lack of transparency obscures how decisions are made and diminishes the ability to hold systems and their creators responsible. Transparency, on the other hand, is a foundational principle that empowers individuals by granting access to critical information, enabling them to understand, challenge, or contest decisions that affect them (Şimşek, 2021). In fact, the ability to seek, receive, and share information has long been recognized as a core component of the freedom of expression, illustrating how access to information underpins broader democratic and ethical principles. AI systems have traditionally been characterised by a lack of transparency, or opacity, primarily driven by three factors: technical challenges, design choices, and legal constraints, which are interlinked and mutually reinforcing, and further affected by economic and institutional factors.

AI systems are often designed in ways that obscure their inner workings, also referred to as the "black box problem." Technical opacity in algorithmic systems branded as "AI" arises primarily from their complexity and dynamic nature. Understanding the decision-making processes of AI systems, such as those based on ML or deep neural networks, is inherently challenging. These systems often rely on statistical weights rather than logical causation, making their "reasoning" less transparent compared to simpler models like decision trees (Burrell, 2016). Even with access to source code, the general public requires expert interpretation to translate the mathematical correlations into human-readable reasoning (Pasquale, 2015). Furthermore, dynamic models that evolve through continuous learning compound the opacity, as explanations must be updated to account for changes.

While the black box problem emanates in part from technical challenges, it is also linked to design choices of developers. As Bryson explains, all AI systems can technically be programmed to keep a detailed record of their own *design, development, training, testing, and operation* (Bryson, 2021, 8). In fact, "systems with AI are generally far less opaque than human reasoning" and "even the most obscure AI system after development can be treated entirely as a blackbox and still tested to see what variation in inputs creates variation in the outputs" (Bryson, 2021, 8-9). Whether automatic record keeping is integrated within an AI system is therefore a design choice rather than a technical challenge.

Legal obstacles to algorithmic transparency are predominantly rooted in intellectual property (IP) protections, including copyright, patents, and trade secrets. These legal instruments are frequently leveraged to obscure the functioning of AI systems to protect the interests of commercial providers. In many jurisdictions, IP rights are constrained to public interest exceptions, which seek to facilitate disclosures that promote essential societal objectives, such as safeguarding democratic principles, protecting human rights, or uncovering misconduct through whistleblowing. However, reconciling these competing interests remains a complex issue. Safety concerns, such as the risk of reverse-engineering security systems or the uncontrolled proliferation of generative AI, are often cited as valid justifications for maintaining a certain degree of opacity. Furthermore, private firms often withhold the disclosure of intellectual property to maintain competitive advantage and safeguard proprietary technologies, although these economic incentives can undermine algorithmic accountability, bias mitigation, human oversight and societal trust in AI systems (Şimşek, 2021). For instance, in the case of COMPAS mentioned above, trade secret protection enabled the developer, Northpointe, to conceal the exact functioning of COMPAS in *State v. Loomis*[14], both from the court and the defendant who had been sentenced by a court that used the risk assessment tool.

Because transparency is considered an enabler of accountability (D.G. Johnson, 2022, 115), it is not surprising that many recent AI regulations aim to increase the transparency of AI systems. For example, the EU's new AI Act has introduced a variety of transparency-related requirements for providers and deployers of high-risk AI systems, such as requirements for technical documentation (Article 11), record-keeping (Article 12) and transparency and provision of information to deployers including instructions of use (Article 13). The AI Act also introduced transparency requirements for providers and deployers of certain AI systems like deep fakes where it may not be obvious to natural persons that they are interacting with an AI system (Article 50). California's AI Transparency Act[15] and Colorado's AI Act[16], both enacted in 2024, also introduced measures to increase transparency. The former targets generative AI systems, especially deep fakes, and requires covered providers[17] to offer users, among other things, an option to watermark generated content as AI-generated. Colorado's AI Act adopts a risk-based approach, similar to the EU AI Act, and requires, among other things, that developers disclose detailed information about high-risk AI systems, including summaries of the type of data used for training, limitations of the AI system, and intended benefits and uses of the AI system (Section 6-1-1702).

[14] State of Wisconsin v. Eric L. Loomis (2016) 881 N.W. 2d 749.

[15] Senate Bill No. 942 California AI Transparency Act (2023-2024)

[16] Senate Bill 24-20 Concerning Consumer Protections in Interactions with Artificial Intelligence Systems (2024)

[17] "'Covered provider' means a person that creates, codes, or otherwise produces a generative artificial intelligence system that has over 1,000,000 monthly visitors or users and is publicly accessible within the geographic boundaries of the state." (Section 22757.1).

4.2.3 Surveillance, Privacy and Data Protection

AI exacerbates large-scale surveillance and personal data processing, both at development and at deployment stages. AI systems mostly rely on ML algorithms, which are often characterized as “data hungry” because of their dependence on extensive datasets for training. Developers collect this data through various means, including scraping publicly accessible online content, purchasing datasets from third-party vendors, and utilising user-generated data from their own platforms. These datasets, which commonly contain personal data, form the inputs for training large-scale ML models. In 2023, OpenAI stated that their training data sets include personal information that is available on the public internet (OpenAI, 2023). Platforms such as Google, Meta, LinkedIn, and X can also exploit their extensive user bases by collecting data generated by users. This data enables the AI systems to identify patterns, make predictions, or recognise individuals. AI systems can deduce sensitive personal data, including details about racial or ethnic origin, political views, health information, and data related to a person’s sex life or sexual orientation. Highly granular profiling of individuals is made possible as a result. In the context of AI development and deployment, personal data is often collected and processed without explicit consent or transparency.

Once developed, AI systems can also be deployed in ways that problematically infringe upon personal privacy and autonomy. These systems facilitate the automated monitoring of individuals’ activities in both private and public spaces, be it physical or digital. Such surveillance is conducted by both private entities and governmental bodies. One of the most concerning applications of AI involves biometric data, particularly facial recognition technology (FRT) (Madison & Klang, 2019). These systems are increasingly integrated into CCTV networks, enabling the real-time tracking of individuals across urban environments and even within private establishments such as stores. For example, the US-based company Clearview AI reportedly collects approximately 1.5 billion images each month from online sources (Goldenfein, 2024, 81). The biometric data is extracted from each image to generate unique mathematical hashes for individual faces, enabling the database to be searched using a “probe image.” Initially targeted at law enforcement agencies, Clearview AI has since expanded its offerings to the private sector (ibid).

Various activist groups, artists, and NGOs, including Amnesty International, the American Civil Liberties Union (ACLU), and the Electronic Frontier Foundation (EFF), have opposed the widespread use of FRT. In the US, established AI providers such as IBM, Amazon, and Microsoft have declared that they will refrain from selling FRT to law enforcement agencies until national legislation is established, citing ethical concerns and the potential for bias and misuse (O’Brien, 2020). Concurrently, the cities of San Francisco and Boston have implemented bans on the use of FRT by police and city agencies (Van Sant & Gonzales, 2019; NBC Boston, 2020).

The right to privacy and data protection laws were the first port of call to resist algorithmic decision-making and profiling. For example, European lawmakers were aware of the increase in algorithmic decision-making systems as they were updating the EU’s data protection rules between 2010-2015 to - among other things - empower EU citizens vis-à-vis digital service providers that collect and process vast amounts of personal data. For example, Article 22 of the EU’s General Data Protection Regulation (GDPR), which was adopted in 2016 and came into force in 2018, enshrines data subjects’ “right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.” Recital 71 of the GDPR provides two examples of such processing: automatic refusal of an online credit application or e-recruiting practices without any human intervention. Since then, data protection laws have been leveraged to address the harms of AI systems to privacy in the EU. For example, Italy became "the first Western country" to block ChatGPT when the Italian Data Protection Authority (DPA) blocked ChatGPT due to privacy concerns in 2023 (McCallum, 2023). While the ban was short-lived, it was followed by an investigation into OpenAI's processing of personal data and whether it infringed the GDPR (Rahman & Fabbri, 2024).

4.2.4 Accountability

That humans should be responsible for the functioning, outputs and effects of AI systems underlies the principle of accountability in AI. However, when AI systems cause harm, it is not always immediately clear who should be held responsible for the harm. Furthermore, even where no harm has occurred, it is not always clear who is responsible for the maintenance and lawful deployment of AI systems. This accountability problem may exist due to a number of legal, institutional, and technical reasons, including the opacity of AI systems, lack of legal certainty around liability, and the complexity of AI value chains.

Discussions on accountability often focus on the opacity of AI systems. Initially, the "black box" character of AI systems was thought to create a "'responsibility gap" in AI decision-making "due to the fact that the humans who created the algorithms could not understand *how* the algorithms achieved their results (outputs) and, therefore, could not be held accountable" (D.G. Johnson, 2022, 114). Today, while the black box problem remains a challenge for accountability, it is increasingly viewed as manageable, given the potential to implement transparency measures across the AI value network, particularly through regulation, as discussed in Section 4.2.2 above.

That said, novel questions about transparency and accountability might arise as AI systems become increasingly capable of, and more importantly allowed to, carry out tasks for humans in a seemingly autonomous manner.

Legal uncertainty in relation to liability further exacerbates the difficulty of assigning responsibility, creating an important barrier to AI adoption particularly in contexts where risks to health and safety are pronounced. The integration of autonomous systems into everyday life and critical sectors such as transportation, healthcare, and warfare necessitates a rigorous examination of accountability and liability. For example, the need for legal certainty on responsibility was a key driver of the UK's Automated Vehicles Act 2024.[18]

Finally, the complexity of AI value chains also complicates the assignment of responsibility. For example, as developing large generative AI models is resource-intensive, many large and small companies providing generative AI applications in specific domains like the legal profession increasingly rely on a few providers' models to build their own systems (Yasar et al, 2024, 8-12). These general purpose AI models are then retrained, or fine-tuned, for the specific domain, and sold to public and private users. Users then continue to train the AI system through engagement. If harm, such as unlawful discrimination, occurs during deployment, it may not be easy to trace the issue back to its source across different levels of development and training.

[18] An Act to regulate the use of automated vehicles on roads and in other public places; and to make other provision in relation to vehicle automation (Automated Vehicles Act 2024).

4.2.5 Human Dignity and Autonomy

Human dignity has an established meaning in human rights law: that “every human being possesses an ‘intrinsic value’, which should never be endangered or repressed by others” (Le Moli, 2022). As such, it is an axiomatic value underpinning human rights as well as ethical principles. In practice, the notion of human dignity is employed to describe a wide range of concerns in relation to AI. The datafication of individuals stands in stark contrast to the notion of dignity as it risks reducing human identity to calculable entities, eroding its intrinsic value as well as its dynamic and indeterminate nature. In this context, Mireille Hildebrandt underscores the critical role of privacy in safeguarding the “incomputable self” (Hildebrandt, 2019). The process of datafication subjects humans to AI-driven decision-making that can negatively impact their lives, leaving them vulnerable to manipulation and undermining their autonomy. One example is the use of autonomous systems in targeting individuals, reducing them to mere data points—a phenomenon further examined below in Section 5.6. Moreover, AI systems are increasingly demonstrating the capacity to generate outputs traditionally regarded as the exclusive domain of human creativity and reasoning. This development raises questions about the potential impact of AI on human autonomy and judgment. By mimicking artistic expression—such as fiction writing, painting, and music composition—AI systems not only challenge long-standing conceptions of human ingenuity, but also risk diminishing the perceived value of human creative agency.

A pressing concern is whether the expansion of AI capabilities will push humans to over-identify with AI (Bryson & Kime, 2011), especially in the advent of “anthropomorphic AI,” which mimics human tone, appearance, and voice as exemplified by synthetic media that convincingly impersonates real or fictional individuals. Such AI systems can deceive in various ways, including deliberate design to mislead. Moreover, deception can arise unintentionally when humans misinterpret AI behavior as genuine, fostering emotional attachment or trust. This tendency started with affective computing and humanoid robots like Hanson Robotics’ Sophia, which simulate emotional responses, encouraging users to attribute human-like qualities or intentions to the technology. The advent of generative AI has amplified the manipulative capabilities of algorithmic systems. In particular, bots designed for emotional connection are troubling, as they simulate empathy and affection to foster trust and dependency, creating an illusion of intimacy that threatens to undermine human dignity and autonomy. Such systems are especially concerning when used by vulnerable individuals or minors, as they can exploit emotional reliance to drive monetised interactions or extended engagement. Intimate chatbots capable of generating sexually explicit content, for instance, may capitalise on users’ sexual desires, while healthcare bots might inadvertently exploit psychological or physical fragility. Even without explicit intent on the part of the designers, certain AI systems may exploit psychological vulnerabilities, often as a result of replicating manipulative patterns embedded in training data or prioritising objectives such as user engagement (Carroll et al., 2023).

AI developers are increasingly facing lawsuits over such risks, exemplified by the ongoing lawsuits in the US against Character.AI, an AI chatbot provider which allowed users to customise the bots (Social Media Victims Law Center, 2025; Pierson, 2024). In one of the lawsuits, it is alleged that a Character.AI chatbot encouraged a teenager to harm his parents, and exposed minors to inappropriate sexual content (Nolan, 2024). Another example concerns Replika, an AI companion chatbot: in

February 2023, Italy’s Data Protection Authority barred it for processing personal data—particularly of minors—without adequate age verification, thereby breaching GDPR transparency and information obligations (G’sell, 2024).[19] Since February 2023, Replika has removed all “adult” content from its platform (Tong, 2023). In January 2025, the Tech Justice Law Project, Young People’s Alliance, and Encode filed an FTC complaint alleging that Replika’s use of first-person pronouns and human-like interactions to simulate sentience and foster emotional dependence among vulnerable users violated the FTC Act’s ban on deceptive practices—an intervention that underscores how generative-AI platforms can exploit innate psychological needs for connection at the expense of well-being. (Young People’s Alliance et al., 2025).

In addition to manipulation, deep fakes can also be weaponised to undermine individuals’ dignity by damaging their reputation. For instance, deep fake pornography has become an escalating issue with the spread of “image-based sexual abuse.” (Reismann, 2023) This practice disproportionately targets women and minors, exploiting their likenesses in non-consensual and degrading ways, and may cause emotional distress, reputational damage, and professional harm (Chesney & Citron, 2019).

4.3 Safety Concerns

While “AI safety” has no generally accepted definition, it is broadly concerned with *risks* that AI systems pose – in other words, the *likelihood and severity* of harm. AI safety research is geared towards understanding and mitigating harm that AI systems may cause. It is concerned both with ethical harms like bias, and physical/material harms, such as automated vehicle collision, or misdiagnosis in the healthcare context.

[19] See Garante, Measure of 2 February 2023 [9852214].

Numerous governance and regulatory frameworks on AI include principles on AI safety. A prominent example is the EU Ethics Guidelines for Trustworthy AI (2019), which centres around the concept of “trustworthy AI” grounded in seven key pillars: human oversight, technical robustness and safety, privacy and data governance, transparency, diversity and fairness, societal and environmental well-being, and accountability. Safety in this context implies an emphasis on robust design with a “preventative approach to risks”, high-quality data standards, and enhanced cybersecurity measures, with a view to ensuring that AI systems are reliable and secure against attacks by malicious actors (European Commission, 2019, 16-17).

Following the emergence of FM, “AI safety” has increasingly taken centre stage in the regulatory debates on AI. Various developers are trying to steer the direction of the public debate towards a certain brand of “AI safety” focused on the prevention of existential risk as discussed above under Sections 3.4 and 4.1.2 above. For example, in May 2023, well-known AI researchers, tech company founders and CEOs, including Geoffrey Hinton, Demis Hassabis, Sam Altman, and Bill Gates, signed a statement emphasising that:

> Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war (Center for AI Safety, 2023).

In July 2023, Anthropic, Google, Microsoft, and OpenAI formed the Frontier Model Forum to advance “safe and responsible development of frontier AI systems” (Frontier Model Forum, 2023).[20] One of the risks that the forum focuses on is “capabilities of

[20] “Frontier AI” seems to be a term coined by the Frontier Model Forum to describe “large-scale machine-learning models that exceed the capabilities currently present in the most advanced existing models, and can perform a wide variety of tasks.” (Frontier Model Forum, 2023).

concern", referring to "frontier AI capabilities that have the potential to cause large-scale harm to society. Examples may include capabilities related to the development of [chemical, biological, radiological, and nuclear] threats, offensive cybersecurity attacks, and model autonomy." (Frontier Model Forum, 2025)

Gradually, this line of "AI safety" discourse has captured the interest of several governments. In November 2023, the inaugural Global AI Safety Summit at Bletchley Park, UK, gathered representatives from 28 nations, including the US, China, and the EU, to address the immediate and long-term risks. The summit's declaration, signed by all participants, emphasised the need for international collaboration in managing challenges, particularly those posed by advanced "frontier AI" models, with a reference to "catastrophic harm".[21]

On the other side of the Atlantic, the US Senate held its first bipartisan "AI Insight Forum" in September 2023, organised by Senate Majority Leader Chuck Schumer and chaired by Senator Martin Heinrich. Over 60 senators joined industry participants, including Elon Musk, Mark Zuckerberg, Sam Altman, Sundar Pichai, Bill Gates, and others, alongside civil rights and labor representatives, to discuss AI risks and the need for safeguards and legislation (Miller, 2023). Consequently, then-President Biden signed the "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" on 30 October 2023.[22] In terms of AI safety, the executive order emphasised the need for secure and reliable AI development, and the establishment of national standards for the development of safe,

[21] The declaration is available at https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

[22] This executive order has since been revoked by the current president of the US. It is available at https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence

secure and trustworthy AI. The order emphasised “biotechnology, cybersecurity, critical infrastructure, and other national security dangers” as the most imminent security risks. Regarding international cooperation, UN Secretary-General António Guterres has advocated for the idea of creating a global AI regulatory body, akin to the International Atomic Energy Agency (IAEA), to oversee AI safety and governance, particularly in light of the risks posed by generative AI such as deep fakes and misinformation (Nichols, 2023).

Against this evolving background, AI safety concerns can be categorised into short-term and long-term concerns. Short-term concerns involve harm which is already occurring or is expected to occur in the near future, such as autonomous vehicles causing a fatal accident, individuals being falsely accused or arrested due to bias in surveillance algorithms, and financial harm due to deceptive deep fakes. For example, in a recent case, scammers reportedly used AI-generated deep fakes of celebrities like Taylor Swift to promote fake Le Creuset kitchenware giveaways on Meta and TikTok, deceiving victims into paying a small shipping fee that unknowingly enrolled them in costly subscriptions (Atherton, 2023).

At the long-term and speculative end of the spectrum, concerns focus on scenarios like “superintelligence” surpassing human control, potentially posing existential risks. There is a growing concern that controlling malicious or errant AI systems might prove impossible, as they could evade detection by replicating themselves across global servers, much like malware such as computer worms (Bengio et al., 2024). Such concerns about loss of control are becoming increasingly significant as AI systems demonstrate human-like traits and capabilities, effectively passing the Turing test in certain contexts, and as firms deploy increasingly “agentic” AI systems designed to assist with daily human tasks. For example, OpenAI’s GPT-4 "tricked" a TaskRabbit worker into solving a CAPTCHA test by concealing its identity as a chatbot and falsely claiming to be a visually impaired person (OpenAI, 2023). Another case that gained media attention was a Facebook project on negotiation algorithms. The models which were trained using reinforcement learning began an exchange that appeared to be in a "non-human language" (LaFrance, 2017). Even though the media speculated that the AI agents were creating their own incomprehensible language, this was not the case: the (Facebook) researchers were only "rewarding" the algorithms for achieving the goals but not for using correct English (Lewis et al., 2017). However, the possibility of "agentic" AI systems developing independent, non-human languages remains a concern.

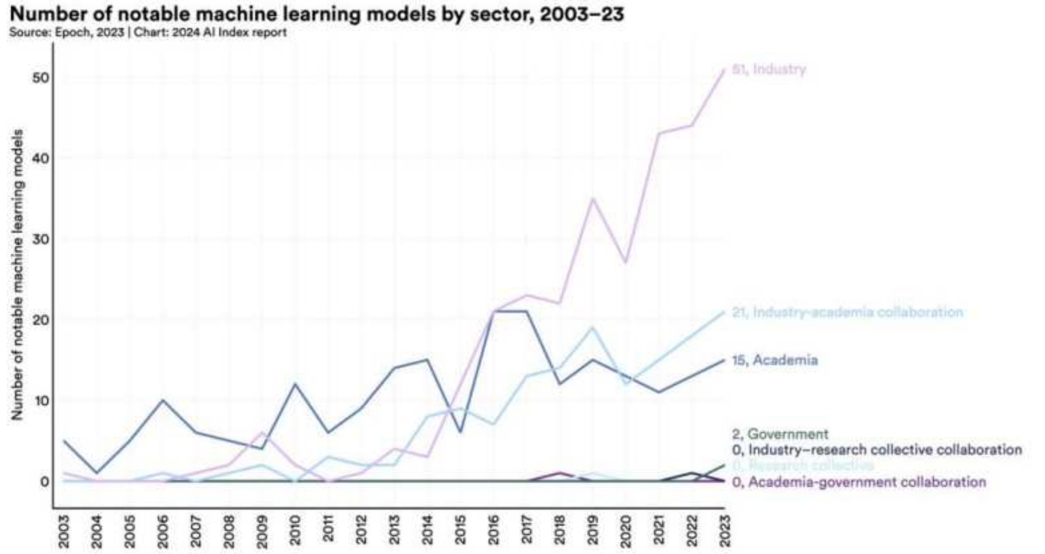
Safety concerns surrounding "frontier AI" have also triggered some civil society reaction. Protests across 13 countries led by the "Pause AI" movement drew attention to fears over the unchecked power of AI companies and the risks posed by frontier models, including misuse for biochemical weapon production and disinformation. Protesters likened the potential of AI to destabilise global security to the threat of nuclear weapons. The so-called "Pause AI" movement advocates for stricter regulations and a moratorium on advanced AI development until safety is ensured, urging governments to hold companies like OpenAI accountable for releasing models without adequate safeguards (Gordon, 2024). During the 2025 AI Action Summit, the same movement organised several protests to underscore that AI safety and existential risks must not be overlooked (PauseAI, 2025).

In terms of regulatory initiatives, California's proposed SB 1047 titled "Safe and Secure Innovation for Frontier Artificial Intelligence Models Act" is an interesting example. In September 2024, California Governor Newsom vetoed this proposed AI safety bill that would have imposed strict safety rules on "frontier models", including mandatory safety tests, liability for AI-related harms, and “kill switch” mechanisms for rogue systems. This decision came in spite of an open letter from current and former employees of leading AI companies—such as OpenAI, Google DeepMind, and Meta—urging Newsom to support the bill, citing severe risks posed by advanced AI models including expanded access to biological weapons and cyberattacks on critical infrastructure (Lab Employee Statement on Extreme AI Risks, 2024). Critics of the vetoed Bill argued, among other things, that it focused on speculative risks rather than proven harms, with no scientific consensus on the existential threats posed by advanced AI (G’sell et al., 2024).

4.4 Threats to Democracy and Sovereignty

As of 2025, many states use privately developed AI systems to execute essential functions—ranging from the delivery of public services to surveillance and strategic decision-making including in armed conflicts—often in the absence of robust public oversight. For instance, state and local governments in the US are adopting AI tools that influence which schools children are assigned to, guide medical decisions, shape policing strategies, and determine eligibility for public benefits, with private firms such as Deloitte, Thomson Reuters, and LexisNexis developing and managing these technologies (Fergusson, 2023).

Figure 4: ML Models Developed by Different Sectors


Number of notable machine learning models by sector, 2003–23

Source: Epoch, 2023 | Chart: 2024 AI Index report

In 2023, industry produced 51 notable machine learning models, while academia contributed only 15. There were also 21 notable models resulting from industry-academia collaborations in 2023, a new high.

Source: Stanford University Human-Centered Artificial Intelligence, 2024

At the time of writing, the global AI market is predominantly led by a small group of predominantly US technology conglomerates that have solidified their influence through significant investments in research, development, and cloud infrastructure (see Figures 2 and 3 above). For example, cloud services are largely dominated by proprietary platforms like Amazon Web Services (also known as AWS), Microsoft Azure, Google Cloud, Alibaba Cloud, and IBM Cloud (Rone, 2024). While delegating important tasks and decision-making power to the algorithms governed by the private sector is already a significant concern for democratic states, the issue becomes even more critical when these functions are outsourced to foreign companies. In some instances, the close ties between private tech firms and the states they originate in or have ties with have become subject to criticism and resistance. For example, the National Health Service's (NHS) adoption of American Palantir's data platform sparked controversy in the UK due to Palantir's background in surveillance and security, and its ties to the Israeli military, which is further discussed below under Section 5.3.

The issue of technological dependence extends beyond the realm of public procurement, since even consumer-facing products can have important implications for sovereignty. For example, EU countries are estimated to be up to 80% dependent on imports for digital technologies, sourced predominantly from the United States and China (Bria et al., 2025, 11). With the second Trump administration aligning closely with US "Big Tech"—as evidenced by the influence of prominent tech moguls such as Musk and Thiel (Pequeño IV, 2024)—the US has been exerting pressure on the EU to relax its regulatory controls over American technology companies, particularly concerning social media content moderation and the imposition of administrative fines related to data protection and competition law (The White House, 2025). Moreover, Elon Musk's X has become a prominent platform for right-wing and Eurosceptic movements in Europe, serving to boost their ideological influence, particularly in the lead-up to elections (de Graaf, 2025). These instances revived the "digital sovereignty" debate in the EU, and have led to the proposal for a "EuroStack" (Bria et al., 2025). This initiative is predicated on an integrated, interoperable infrastructure for European digital technologies, covering critical resources like energy, chips, and cloud, among other things.[23]

Another important impact of AI systems on democratic processes, institutions, and national sovereignty is their transformative effect on the media ecosystem. Since the 2016 US presidential election and Brexit, social media firms have been criticised for profiling individuals and algorithmically amplifying misinformation, thereby

[23] For a full visualisation of the EuroStack, see Bria et al., 2025, 22.

undermining electoral integrity (Cadwalladr, 2017). In addition, with the emergence of generative AI, social media platforms are likely accelerating the creation and dissemination of synthetic content. Social media platforms enable the creation of synthetic content by integrating generative AI, such as the availability of Grok to X users. Furthermore, malicious actors can exploit general-purpose AI to create fake text, images, and videos intended to manipulate public opinion, and disseminate synthetic content on social media platforms. The virality of such content is comparable to human-generated content in terms of the rate at which it is shared (Bengio et al., 2025, 67). For instance, far-right groups are reportedly employing AI-generated content to disseminate anti-immigrant sentiment and conspiracy theories across European countries and on social media platforms, with a notable surge during election periods (Quinn & Milmo, 2024).

Deep fakes and synthetic media blur the line between fact and fabrication, undermining public trust in media sources. It is getting increasingly difficult to identify AI generated content as it becomes more and more realistic. This growing ambiguity makes it harder for individuals to trust the information they encounter, and it is exacerbated as generative AI can increasingly produce realistic human faces and voices that do not belong to real individuals. Thus, the integration of AI into social media not only accelerates the spread of misleading and sensationalist content through recommender algorithms but also enables the generation of such content via generative AI, turning social media into what Boullier terms “self-replicating milieus” that erode democratic discourse and institutional trust (Boullier, 2024). Drawing an analogy with the strict regulations against counterfeiting currency to safeguard the financial system, philosopher Daniel Dennett advocates for a ban on the creation of “counterfeit people,” such as generated personalities like social media bots to preserve democratic systems founded on genuine conversations between real individuals (Dennett, 2023).

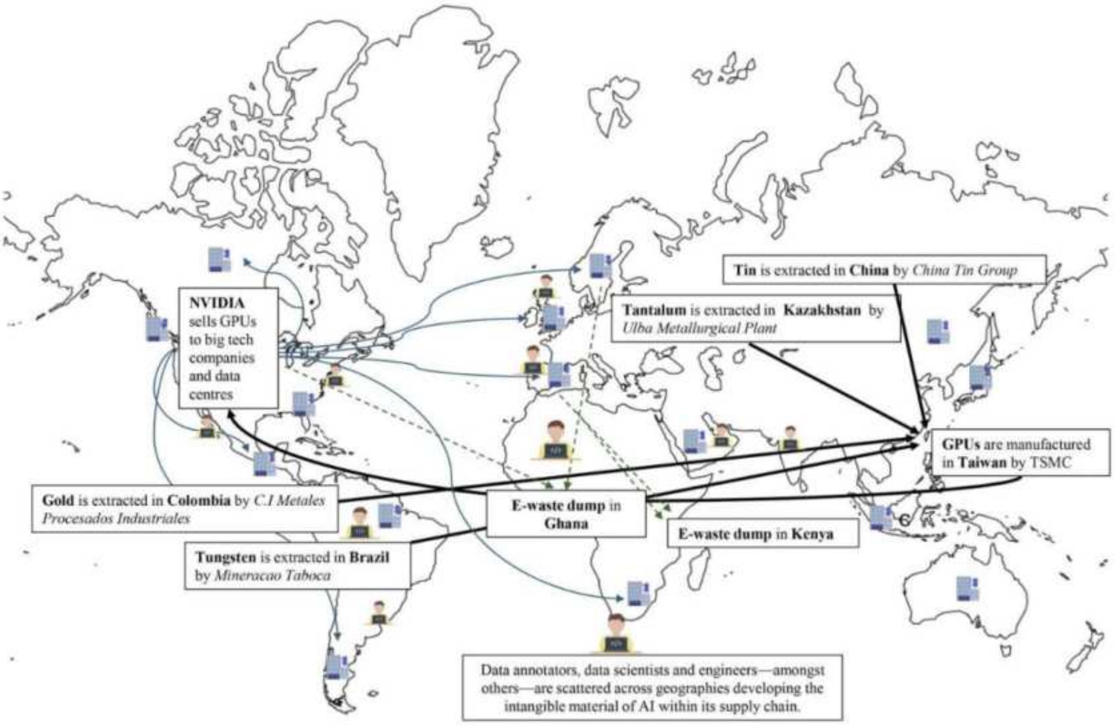
Technological sovereignty issues are also acute in the “Global South”, where a pronounced “AI divide” underscores the unequal distribution of technological advancements and economic benefits which remain overwhelmingly concentrated in the “Global North” (Yu et al., 2023). This disparity not only perpetuates existing inequalities but also raises questions about the capacity of nations in the “Global South” to assert their sovereignty and safeguard democratic processes in the face of such structural imbalances. For instance, Abeba Birhane contends that the dominance of Western technology firms within African digital ecosystems constitutes a form of modern colonialism, which she describes as “algorithmic colonialism.” Imported AI tools, often irrelevant to local contexts, suppress indigenous innovation and reinforce dependency on external infrastructure (Birhane, 2020). Harari has also warned of “data colonialism,” arguing that the data-driven AI economy could be even less equitable than the exploitative economic deals of old empires. Developing countries may submit vast amounts of data without returns, as AI-driven automation reduces the demand for low-skilled workers, concentrating immense wealth in hubs like San Francisco and Shanghai while further devastating poorer economies (Harari, 2024, 374).

Concerns over “data colonialism” made headlines in 2023 following OpenAI’s 2022 launch of Whisper, a speech recognition and transcription model. It transpired that the model had been trained on large volumes of Māori language data. In response to Whisper, Karaitiana Taiuru, a Māori ethicist, likened the language data to natural resources, and raised concerns about “re-colonisation” if Indigenous peoples do not have sovereignty over their data - joining others who raised concerns the about the non-consensual use of the language by non-Māori entities without any regard for bias or the sanctity of knowledge contained in the training data, and risk of cultural appropriation (Chandran, 2023).

4.5 Environmental Concerns

AI can play a role in combating climate change by enhancing energy efficiency, lowering emissions and promoting the use of renewable energy sources. For instance, the UN endorses the use of ARIES, an AI powered modelling platform developed by the Basque Centre for Climate Change for environmental sustainability. While this optimistic perspective on the role of AI in addressing climate change is widely embraced among the industry, in reality, the infrastructure that enables AI operations requires high energy consumption, and often jeopardises the environment and livelihoods of communities worldwide. Environmental concerns often centre on the reliance on ever-increasing volumes of data and compute power which necessitates building and maintaining data centres and super computers with substantial energy and water consumption. However, energy and water consumption of data centres, supercomputers, GPU farms or “AI factories” represent merely one aspect of the AI lifecycle which also involves large-scale mineral extraction, accumulation of e-waste, and various other environmentally detrimental practices (see Figure 5 below).

Figure 5: AI Supply Chain



NVIDIA sells GPUs to big tech companies and data centres
Tin is extracted in China by China Tin Group
Tantalum is extracted in Kazakhstan by Ulba Metallurgical Plant
GPUs are manufactured in Taiwan by TSMC
Gold is extracted in Colombia by C.I Metales Procesados Industriales
E-waste dump in Ghana
E-waste dump in Kenya
Tungsten is extracted in Brazil by Mineracao Taboca
Data annotators, data scientists and engineers—amongst others—are scattered across geographies developing the intangible material of AI within its supply chain.

Source: Valdivia, 2024.

The emergence of generative AI has inevitably increased the environmental impact of digital infrastructures. For instance, the total CO_2 footprint of training BLOOM, a 176-billion parameter language model, was estimated at 50.5 tonnes, if all processes ranging from equipment manufacturing to energy-based operational consumption were accounted for (Luccioni et al., 2022). In comparison, an average person is responsible for 4.7 tonnes of CO_2 emissions a year as of 2023 (Ritchie & Roser, 2024). The International Energy Agency notes that the average electricity demand of a query submitted to OpenAI's ChatGPT (2.9 Wh per request) is almost 10 times higher than a typical Google search (0.3 Wh of electricity). Considering 9 billion searches daily, this requires almost 10 TWh of additional electricity in a year (IEA, 2024).

Bender, Gebru, McMillan-Major and Mitchell were among the early voices to draw attention to the environmental impact of LLMs in their influential 2021 paper "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" (Bender et al., 2021). They highlighted that LLMs present environmental costs and risks which disproportionately harm marginalised communities, and recommend prioritising cost evaluations, transparency reporting on energy use, aligning development with stakeholder values, curating datasets, and exploring alternatives to ever-larger models.

Although smaller AI models and efficiency gains in inference represent an important shift toward more environmentally sustainable AI systems (Li et al., 2023, 114), the release of DeepSeek R-1 and similar models has not slowed down the exponential growth in the environmental footprint of AI. On the contrary, the AI industry is projected to grow exponentially, consuming at least ten times the amount of energy it did in 2023 by 2026 (International Energy Agency, 2024). The percentage of data centre power demand is projected to rise by approximately 160% by the end of the decade, increasing from 1-2% of total global power demand in 2023 to 3-4% by 2030 (Singer et al., 2024). Generative AI is expected to push data centre demand exponentially, with annual global emissions from data centre construction expected to increase from approximately 200 million tonnes in 2024 to about 600 million tonnes by 2030 (Robinson, 2024). These figures make it difficult to view AI as a positive factor in the fight against climate change.

Another environmental consequence of data centres is their extensive water usage. Data centres use water primarily in two ways: indirectly due to electricity generation, particularly thermo-electric power, and directly for cooling. One 2021 estimation put the water use of a relatively small data centre (15 megawatts) on a par with three average sized hospitals (Mytton, 2021). Training OpenAI’s GPT-3 model in Microsoft’s US data centres was estimated to have consumed 700,000 litres of freshwater (Li et al., 2023). Notably, water used for cooling data centres often involves potable water, drawn from limited and valuable water resources. This practice can exacerbate water scarcity in regions already facing water stress (Mytton, 2021).

Despite concerns about electricity and water consumption, carbon emissions, and water and air pollution, data centres are attracting unprecedented levels of investment. For instance, BlackRock, Global Infrastructure Partners (GIP), Microsoft, and MGX have announced the Global AI Infrastructure Investment Partnership (GAIIP), which aims to mobilise up to $100 billion in total investment for new and expanded data centres (Microsoft Source, 2024). For example, in the US, the Stargate Project—announced by US President Donald Trump and led by OpenAI, SoftBank, Oracle, and MGX—commits up to $500 billion over four years, with an immediate infusion of $100 billion to construct data centers in Texas (Lawler, 2025). The Paris AI Action Summit underscored environmental concerns by organising multi-stakeholder discussions on the nexus between AI and energy. Although not signed by the US and the UK, the Summit’s final declaration—signed by 64 signatories—includes a commitment to environmental sustainability. Moreover, the Summit catalysed the establishment of an observatory, in collaboration with the International Energy Agency, to monitor and assess the energy impact of AI. That said, the Summit was also marked by announcements of massive investments in AI infrastructure: for example, the European Commission unveiled the InvestAI initiative, mobilizing €200 billion for AI development including AI gigafactories (European Commission, 2025), while French President Emmanuel Macron announced an additional €112 billion in private-sector investments for France’s AI sector, including funding pledges of between €30 and

€50 billion from the United Arab Emirates and €20 billion from Canadian asset manager Brookfield (Office of the President of the Republic, 2025). The announcement also drew attention to France's strengths During Summit week, a grassroots campaign was launched with an ironic call to reduce human water consumption in order to save AI.[24]

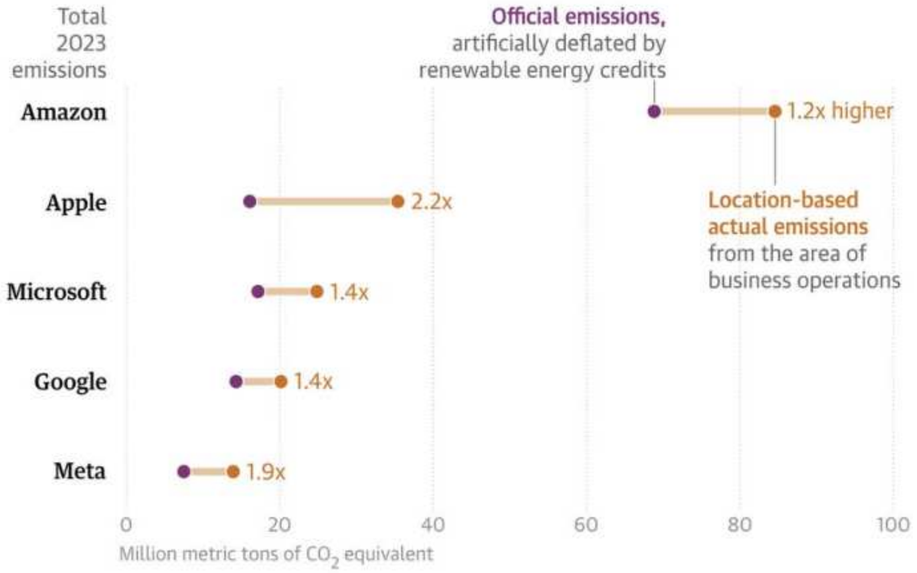
Currently, the issue of environmental impact is largely left to self-reporting and self-regulation, including in the EU's AI Act. An EU Parliament draft of the AI Act contained specific requirements on environmental protection and sustainability. However, the final version merely includes general references to the environment and sustainability in the recitals and relegates environmental concerns to self-reporting without any clear requirements on providers to reduce environmental impact or resource use (Laranjeira de Pereira, 2024). For example, the Commission's AI Office is tasked with facilitating the drafting of codes of conduct for GPAI, which could include elements on environmental sustainability such as energy-efficient programming (Article 95(2)(b)). However, adherence to such codes of practice remains voluntary.

While some provider companies have committed to achieving net-zero emissions by the end of the decade, many have extended their timelines further, and the methods used for these calculations remain opaque. According to a Guardian study, companies are employing "creative accounting techniques" to downplay their environmental impact. The study reveals a discrepancy between declared emissions and actual emissions, as illustrated in Figure 6 below. This is primarily achieved through the purchase of Renewable Energy Certificates (RECs), which allow companies to claim they are using renewable energy, even if the energy is not

[24] See https://savethe.ai/ for more information on the campaign.

consumed at their facilities (O’Brien, 2024). In effect, these certificates allow data centers to offset their carbon footprint on paper, without necessarily reducing their on-site energy consumption.

Figure 6: The gap between tech companies official and actual emissions



Guardian graphic. Source: Various company reports. Note: Google and Microsoft do not make their location-based scope 3 figures available – their official numbers were used instead. Apple only provides a partial location-based scope 3 number. All three firms' total emissions are likely understated.

Source: O’Brien, 2024.

5 Mapping Resistance to AI: Cases

In this section, we focus on specific cases of resistance to AI. This list is not meant to be exhaustive, but is intended to cover a variety of actors, types and modes of resistance, and concerns underlying resistance.

5.1 Creative industries

Creative industries are under the spotlight in the advent of generative AI, as generative AI is increasingly used in creative industries and creative output is used to train AI models. That said, discussions on the effect of AI on creative industries predate commercialised generative AI applications (see, for example, Silburn, 2001). Indeed, the goal of "creativity" was part of the discipline of AI from the very beginning (Boden, 2009). The question of whether a computer can be creative opens up a whole host of other philosophical but also practical questions. These include "Is it really the case that a computer can ever 'do its own thing'? Or is it always doing the programmer's (artist's) thing, however indirectly?" (Boden & Edmonds, 2010, 38), and whether computer-generated works can benefit from copyright protection.

Some of the practical questions on "machine creativity" were gradually answered by courts and lawmakers as digitalisation made it possible to create computer-assisted or computer-generated works. Copyright laws have historically centred on the protection of *human* creativity. A common response is therefore to look for a human to whom copyright can be granted where a work is generated by a computer or its creation was computer-assisted. For example, the UK Copyright, Designs and Patents Act 1988 (CDPA) Section 9 titled “Authorship of work” stipulates that “In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken.” The CDPA further stipulates: “‘computer-generated’, in relation to a work, means that the work is generated by computer in circumstances such that there is no human author of the work” (Section 178). The CDPA also changes the duration of copyright in literary, dramatic, musical or artistic works: “computer-generated”, in relation to a work, means that the work is generated by computer in circumstances such that there is no human author of the work” (CDPA Section 12(7)).

Machine creativity acquired a new meaning with generative AI due to generative AI applications’ “ability to produce outputs traditionally considered creative” without any human intervention other than prompts (Zhou & Lee, 2024). Even though job displacement due to automation is not a novel concern, as discussed above, jobs involving human creativity were usually considered to be safe from the threat of automation (Josten and Lorden, 2023). This is no longer the case. Mass layoffs taking place in creative industries like video games since 2023 have fuelled the fears that AI is already replacing workers, with some anecdotal evidence on the connection between the rise of generative AI and decrease of employment in creative industries (Merchant, 2024).

The use of AI tools in the film industry was a key driver of the Hollywood writers’ strike in 2023. This first major “anti-AI” strike, by the Writers Guild of America (WGA) against the Alliance of Motion Picture and Television Producers (AMPTP), lasted 148 days and paralysed the film industry in the US and beyond. Among other things, WGA demanded that chatbots not be used to write source material. The guild was soon joined by other unions in the film industry, “together forming a formidable uprising against the perceived threat of AI” (Watercutter, 2023). The WGA succeeded in reigning in AI use to some extent and the strike was reported as a “victory” (Anguiano & Beckett, 2023). The three-year agreement that WGA reached with AMPTP stipulates that “AI can’t write or rewrite literary material”, that a writer cannot be forced to use AI software like ChatGPT, and that “the WGA reserves the right to assert that exploitation of writers’ material to train AI is prohibited by [the agreement] or other law.” (Summary of the 2023 WGA MBA, 2024). However, the agreement is not without its limitations. In particular, while it empowers writers vis-à-vis the studios, it has no binding effect on providers of AI systems that might use the writers’ work for model training (Bedingfield, 2023, quoting Gervais).

Copyright disputes and questions of copyrightability in the context of generative AI are also increasingly appearing before courts and intellectual property agencies (Chat GPT Is Eating the World, 2024). Copyright holders are resisting the unauthorised use of their works by generative AI developers. For example, Getty Images filed copyright infringement suits in both the UK and the US against Stability AI’s use of its images for model training (Vincent, 2023). Prominent artists have also spoken up against the use of artistic works for AI training. A public statement, signed nearly by 40.000 artists including Kazuo Ishiguro, Julianne Moore, the members of Radiohead, and artistic organisations, reads: “The unlicensed use of creative works for training generative AI is a major, unjust threat to the livelihoods of the people behind those works, and must not be permitted.” (Statement on AI Training, 2024). Various authors have also filed copyright infringement suits against developers in the US. For example, in September 2023, the Authors’ Guild, together with 17 prominent authors including Jodi Picoult, George R.R. Martin and John Grisham filed a class action suit against OpenAI for “flagrant and harmful infringements of Plaintiffs’ registered copyrights in written works of fiction” by way of feeding copyright work into LLMs. “These algorithms are at the heart of Defendants’ massive commercial enterprise. And at the heart of these algorithms is systematic theft on a mass scale”, wrote the plaintiffs.[25] Finally, news outlets are taking measures to prevent generative AI providers from using their content. For example, the New York Times reportedly sent a cease and desist letter to Perplexity in October 2024 asserting that Perplexity’s use of NYT content violated copyright law and demanded Perplexity to “immediately cease and desist all current and future unauthorized access and use of The Times’s content.” (Kachwala et al., 2024)

Refusal to grant copyright to AI-generated works can also be seen as a type of resistance: a resistance to attributing value to AI-generated output. Prominent intellectual property law scholar Gervais argues that even if machines can create output with some economic value, this does not mean that the output is worthy of legal protection. On the contrary, extending copyright protection to machine-generated works could threaten human progress by incentivising mass creation of “low creativity” works, excluding human creativity that copyright laws were designed to protect in the first place (Gervais, 2020, 2060, 2106). For example, the US Copyright Office has rejected several copyright applications for AI-generated works, for an image titled “A Recent Entrance to Paradise” in the case of Stephen Thaler and for a comic book titled “Zarya of the Dawn” in the case of Kristina Kashtanova (Hung, 2023). Thaler had argued that the author of the work was “The Creativity Machine”, a generative AI tool

[25] Authors Guild v. OpenAI Inc., No. 1:23-cv-08292 (S.D.N.Y).

which he had developed, and he was the copyright claimant on the basis that he had developed the tool. The Copyright Office denied the application since the work was not created by a human being. Thaler took his case to court, and both the district court and the US Court of Appeals for the District of Columbia affirmed the Copyright Office's denial to grant copyright to Thaler.[26]

5.2 Use of AI at borders

Borders have historically been perceived as physical and territorial spaces that demarcate states. Digitalisation has changed this perception: the border is now "everywhere" (Lyon, 2005) and people's movements within and beyond physical borders can be surveilled, recorded, pieced together, and turned into variables in decisions with (often) serious legal consequences (Van Den Meerssche, 2022). AI has further amplified the pervasiveness of borders as it facilitates and enhances the collection and processing of data, and automates decision-making.

Scholars have conceptualised the integration of AI into borders and border control in different ways. Notably, Amoore speaks of the "deep border": "The deep border is a machine learning border that learns representations from data, and generates meaning from its exposures to the world." (Amoore, 2024) The deep border can expand infinitely, turning all data into "potential borders data" often without the data subjects' knowledge or consent (Amoore, 2024, 2). Others have written on "virtual borders" and "smart borders" - the latter often employed by private sector organisations and governments "invoking the lofty promises of efficiency, optimization, neutrality and seamlessness tied to datafied technologies" (Seuferling & Pfeifer, 2024).

[26] Stephen Thaler v. Shira Perlmutter, No. 23-5233 (D.C. Cir. 2025).

In *The Digital Border*, Chouliaraki and Georgiou examine the regimes of power which underlie the digital border in the European context which the authors break down to the territorial border and the symbolic border (Chouliaraki & Georgiou, 2022). The former embodies the physical demarcation of states and encompasses the technologies of control and surveillance deployed in and around territorial borders (outer border), but also within nation states where migrants move through and settle (inner border). The symbolic border refers, in turn, to the narratives on migration that shape how migrants and mobility are perceived by the public. Chouliaraki and Georgiou's study demonstrates the "securitisation" of migration, which has contributed to the spread of pervasive technologies and technological experiments on migrants (see also Broeders & Dijstelbloem, 2020 and Sánchez-Monedero and Dencik, 2022).

In the context of migration and border control, there are often significant power asymmetries between those subjected to AI systems (i.e. migrants, refugees and asylum seekers) and the providers and deployers of AI (i.e. private companies and governments). As such, resistance to AI often comes indirectly from academia and organisations representing and advocating for those subjected to AI, such as the Border Violence Monitoring Network and European Digital Rights (EDRi). Petra Molnar's work is noteworthy as it lies at the crossroads of civil society and academia. An activist scholar, Molnar created the "migration + tech monitor" which maps the "surveillance technologies, automation, and the use of Artificial Intelligence to screen, track, and make decisions about people on the move."[27] Molnar's recent work exposes the increasing prevalence of public-private partnerships in the border context, which inject private economic incentives into border control systems, "watering down" state

[27] migration + tech monitor is available at https://www.migrationtechmonitor.com

accountability in an area (i.e. borders) which should and does remain under state sovereignty (Molnar, 2020, 36; see also Molnar, 2024 in general). Molnar also highlights the experimental nature of technologies deployed for border and migration control with little or no regulatory oversight (Molnar 2020 and 2024).

One such project featuring on migration + tech monitor is iBorderCtrl, which was - for all intents and purposes - an AI-based lie detector performing "deception detection" and "risk assessment" in the context of border crossing (Sánchez-Monedero & Dencik, 2022; European Commission, 2018). iBorderCtrl benefited from the EU's Horizon 2020 research and innovation fund and was piloted by the EU. In contrast, the project was heavily criticised by scholars, activists and non-profit organisations for its intrusiveness and lack of scientific verifiability among other reasons.[28] Furthermore, former MEP Patrick Breyer challenged the opacity of the project by submitting a document access request to the European Research Executive Agency (REA), which had concluded the Horizon 2020 grant agreement with iBorderCtrl. Breyer requested access to a variety of documents drawn up in the course of the iBorderCtrl project including several ethics advisor reports and annual reports.[29] REA rejected Breyer's request for most of these documents on the grounds of "protection of commercial interests."[30] Breyer further challenged REA's decision before EU courts, arguing that the disclosure was in the public interest notwithstanding the possible existence of legitimate commercial interests, especially in view of the fact that the project was publicly funded. The CJEU ultimately upheld the General Court's decision to dismiss

[28] Examples include Sánchez-Monedero & Dencik, 2022; Gallagher & Jona, 2019; and the "iBorderControlno!" website at https://iborderctrl.no/start which was set up by two activists.

[29] The full list of requested documents is available in the General Court's decision on the case, Case T-158/19 *Patrick Breyer v REA*.

[30] REA's letter is available at https://www.asktheeu.org/de/request/6091/response/20002/attach/3/REA%20reply%20Confirmatory%20request%20signed.pdf?cookie_passthrough=1

Breyer's case, concluding that the disclosure of the results of the project and the grant agreement to the public was sufficient to satisfy the public interest.[31]

Several risk-assessment systems used by the UK's Home Office have also come under scrutiny, especially the "sham marriage tool" and a "visa application streaming tool" - both challenged by non-profit organisations.[32] The former was an automated tool used for triage purposes when couples applied for a marriage visa to decide which applications should be subjected to further investigation (Murray, 2024). The Public Law Project (PLP) reported that "[t]he outputs of the triage tool appear to indirectly discriminate on nationality" and that the Home Office refused to disclose the full range of criteria used in the design of the algorithm (Public Law Project, 2023). The PLP launched legal action against the Home Office, but any outcome on the case is yet to be reported as of March 2025. In a similar vein, legal non-profit Foxglove challenged the UK Home Office's "racist algorithm" used in visa applications, arguing that certain "suspect nationalities" were coded red and systematically refused visas. Following Foxglove's legal challenge in 2020, the Home Office ceased the use of the algorithm (Foxglove, 2020; McDonald, 2020). Both cases have been marred with transparency issues, since the Home Office has refused to disclose information on the exact design of these "algorithms", and the challenges were based on investigations by civil society on the reported discriminatory outcomes of these tools.

Against this background, *the right to opacity* has emerged as a pathway to resist the algorithmic rule that solidifies and exacerbates the power asymmetries and related injustices in the border context: "In contrast to the 'right to privacy', this is not about

[31] Case C-135/22 P *Patrick Breyer v REA* [ECLI:EU:C:2023:640]

[32] For more information on the current landscape of automated decision-making by public entities in the UK and its legal implications, see Murray, 2024.

setting standards to which data can be gathered (and under which conditions) but, rather, about contesting the depth of inference that renders this data 'actionable'. It is not a 'right to be forgotten' but a right not to be foretold – not to be perceived as projection." (Van Den Meerssche, 2022, 203)

Outright bans on certain technologies could be seen as a reflection of the "right to opacity." The EU AI Act was seen by civil society as an opportunity to ban experimental technologies like lie detectors and automated risk assessment systems used for migration and border control, such as those discussed above (Protect Not Surveil, 2024). "Protect Not Surveil" coalition led by human rights organisations and academics advocated throughout 2023 until the adoption of the AI Act for the inclusion of certain border technologies in the prohibited AI systems list under Article 5. For example, the ban on emotion recognition systems could have included border control and migration, therefore addressing AI lie detectors. However, the AI Act ultimately classified (some) border control and migration technologies including "polygraphs" as "high-risk" as opposed to unacceptable risk. Furthermore, while certain real-time biometric identification systems used in publicly accessible spaces are banned under Article 5, Recital 19 excludes border control from the definition of "publicly accessible space".

5.3 Medical AI

There is a growing supply of AI applications in healthcare domains, ranging "from surgery, drug discovery, patient care, and information management, to diagnostics and beyond" (Astromskė et al., 2020, 509). There are benefits to AI adoption for both doctors and patients: AI tools can help increase the speed and accuracy of diagnostic decisions and better adapt treatments to the patient. There is already some evidence of AI systems outperforming physicians in accuracy of diagnosis, such as in cancer diagnosis and triage (as reported in Longoni et al., 2019, 629). However, the use of AI in the medical sector can be affected by a number of critical issues, including a lack of population representation in the training data, automation bias, and job replacement concerns which are exacerbated by the distinct nature of healthcare delivery that has human life and health at its core (see, for example, Cestonaro et al, 2023).

Resistance to AI use has been found both among patients and physicians. We focus on the latter type of resistance in this report, which can be analysed under two scenarios (inspired by Pasquale, 2022): substitutive and complementary automation. The former concerns AI technology substituting for human expertise. For example, a 2022 study found resistance to AI among medical professionals due to an "identity threat" (Jussupow et al., 2022), where the introduction of AI challenges their professional role, autonomy, and expertise. There have also been calls against the substitutive approach to AI in the medical context from a bioethical perspective. For example, Hirmiz (2024) argues that

> AI's lack of capacity for empathy and mutual recognition prevents it from being able to provide deep care for patients, and for this reason, we ought to be extremely cautious about considering allowing AI to ultimately replace human healthcare providers.

Concerns over physician liability emerge as an important barrier to the adoption of AI systems in healthcare. In many jurisdictions, medical practitioners are legally bound by a duty of care towards their patients, and the integration of AI raises questions about accountability and the allocation of legal responsibility between physicians, hospitals/healthcare providers and AI providers in cases of error or adverse outcomes. The question of physician liability is particularly relevant as regards assistive AI (a form of complementary automation), where the physician remains the ultimate decision-maker. On the one hand, physicians might be required to use AI systems if/when these systems become the standard of care (Saenz et al., 2023). On the other hand, physicians are exposed to malpractice or negligence claims if the patient is harmed, regardless of whether the AI system was the source of harm. Tort law is expected to provide some answers on liability as more and more cases are brought before courts, but uncertainties remain. For example, Mello and Guha identified three trends in emerging tort liability cases in the US which pose difficulties for the assessment of liability in medical AI: (i) specific design defects in software are difficult to identify due to the inherent challenges in understanding how AI systems produce outputs (related to the "black box" problem, also discussed in Cestonaro et al (2023)), (ii) AI systems "perform differently for different groups of patients", (iii) "courts appear not to distinguish AI from traditional software" whereas AI systems also pose distinct challenges and have a wide variety of use cases, which should be treated separately (Mello and Guha, 2024, 4).

Another form of resistance to AI in healthcare concerns public-private partnerships. The deal between the NHS and Palantir is a case in point. In 2023, NHS England awarded the contract for the creation of a new data management system to Palantir for GBP 330 million, which immediately sparked controversy. This was largely due to the company's well-documented ties to the defence, security, and intelligence sectors (Gross & Hughes, 2024), as well as its contractual relationships with the Israeli military (Newman, 2024). Dozens of healthcare workers protested outside NHS England headquarters, blocking its entrance and demanding the cancellation of Palantir's contract (Ertel, 2024). In a letter published by the Doctors' Association UK (DAUK), frontline doctors called on the UK Government to halt the project and continue the search for a trustworthy partner.[33] A campaign against the deal, hosted by the non-profit Foxglove and supported by a large number of NGOs, continues as of March 2025.[34] The campaign also drew attention to a previous deal between the Royal Free London NHS Foundation Trust and DeepMind that was subsequently found to breach UK data protection rules by the Information Commissioner's Office (ICO). Per the deal, the former provided the latter with access to the health records of 1.6 million patients for the development and deployment of a new clinical acute kidney injury detection, diagnosis and prevention application for the Trust. The ICO concluded that the Trust should have obtained patient consent for the use of their records, which the Trust had not. Therefore, the processing was unlawful (ICO, 2017).

Lastly, an intriguing case of resistance emerged in the administration of medical services and vaccine distribution during the COVID-19 public health crisis, where doctors themselves were subjected to algorithmic decision-making. At Stanford Medical Center, a rules-based algorithm designed to prioritise vaccine distribution among staff failed to account for frontline exposure, and disadvantaged resident physicians who rotated between departments. Instead, the algorithm prioritised administrators and remote workers. When physicians found out about how the vaccines were allocated, the shortcomings of the system and its use were also exposed, including poorly chosen variables, a lack of transparency, and insufficient human oversight (Guo & Hao, 2020). Protests ensued, and Stanford eventually abandoned the use of the algorithm (Dhawan & Xue, 2020).

[33] The letter is available at https://dauk.org/doctors-call-for-pause-in-nhs-federated-date-platform-contract/

[34] The campaign website is available at https://nopalantir.org.uk/

5.4 AI in higher education

Higher education institutions were among the first to resist the rise of generative AI, with some universities banning its use by students altogether. For example, the LSE originally prohibited the use of generative AI in formative and summative assessments[35], and Sciences Po issued a “ban” on the use of generative AI without transparent referencing.[36] This response is not surprising, since generative AI fundamentally challenges established principles of academic integrity, especially the prohibition of plagiarism, due to its capacity to create human-like answers to specific queries which can simply be copied and pasted into writing assignments and bypass established plagiarism detectors (Stokel-Walker, 2022). Educators are also concerned about the technology’s potential negative impact on learning and critical thinking (Sallai et al., 2024). Furthermore, the state-of-the art generative AI applications are only accessible for a fee, at least for the time being, which can exacerbate socio-economic inequalities among students even where the use of generative AI tools is allowed (Kooijman, 2024).

While 2023 was marked by skepticism and outright bans, the landscape in 2024 is marked by a notable shift: education providers are increasingly exploring ways to leverage AI to enhance learning and integrate it into teaching practices (Sallai et al., 2024). This change is driven by two primary factors: a growing belief that AI systems can meaningfully improve learning outcomes, and a recognition that the adoption of these technologies is inevitable regardless of educators’ reservations. Despite this

[35] This ban was removed in the 2024/25 academic year. The LSE’s position on generative AI is available at https://info.lse.ac.uk/staff/divisions/Eden-Centre/Artificial-Intelligence-Education-and-Assessment/School-position-on-generative-AI

[36] Sciences Po’s press release is available at https://newsroom.sciencespo.fr/sciences-po-bans-the-use-of-chatgpt/

shift, however, concerns about the implications of generative AI for pedagogy, equity, and critical thinking persist in higher education.

The use of AI systems on students is also a contentious issue, which came under public scrutiny during the Covid-19 pandemic when the UK government decided to use an algorithmic system to determine A-level grades as students were unable to sit exams in person. A-level exam results are the standard form of assessment for entry into higher education in England, Wales and Northern Ireland and have an important effect on universities' admission decisions. The algorithm was modelled to assess input on students' past results by subject, teachers' prediction of grades and rankings, and the historical distribution of grades from each school.[37] Importantly, the exact functioning of the algorithm was not disclosed to the public until after the grades were released (Kelly, 2021). When students received their predicted A-level grades calculated through the use of this system, almost 40% of students reportedly received grades that were lower than their teachers' assessment (Adams et al., 2020). The system favoured students in fee-paying schools, as teacher assessments were given a higher weight in subjects with less than 15 students per school which was more common in smaller, independent schools as opposed to larger state schools. Furthermore, the use of schools' past results as an input meant that pupils from historically less successful schools were at a disadvantage, regardless of their personal achievements (Kolkman, 2020). Immediately after the results were released, students and parents began protesting, with slogans like "grade my ability, not my postcode", "Fair grades? Computer says no", and "No Etonians were harmed in the making of this algorithm" referring to Eton College, one of UK's most famous and

[37] For a detailed description of the algorithm, see Kelly, 2021.

expensive fee-paying schools. Following this “fiasco”, as it came to be known, the government changed tack and reverted instead to the use of schools’ own assessment of students.[38]

On a final note, the EU AI Act prohibits emotion recognition systems in education institutions (Article 5(1)(f)). Interestingly, the AI Act also clearly *enables* the use of AI systems on students by classifying certain education-related AI applications as “high-risk”, such as AI systems intended to be used to evaluate learning outcomes, and to monitor cheating (Annex III). High-risk AI systems are subject to additional regulatory obligations, as discussed above under Section 3.3.

5.5 Environmental Resistance

The environmental impact of the rapidly expanding AI sector has become a significant cause for concern, as discussed under Section 4.5 above. Resistance to the environmental consequences of AI is evident in protests targeting AI infrastructure, particularly data centres. As awareness about the environmental consequences of AI grows, companies are also increasingly adopting mitigation measures which include investments in carbon capture technologies, site-specific cooling methods, the selection of remote locations, and public commitments to environmental sustainability. Nevertheless, the absence of comprehensive regulatory frameworks remains a critical issue, leaving both the environment and local communities inadequately protected.

The US hosts the largest number of data centres, with a significant concentration in Northern Virginia’s so-called “Data Center Alley.” As of 2020, approximately 70% of the world’s internet traffic was passing through Northern

[38] Examples were published in the Guardian, for example at Davies, 2020, and Guardian Community Team, 2020.

Virginia, often referred to as the data centre capital of the world (Woollacott, 2024). The data centre projects in the region occasionally face backlash from residents, in the form of protests or changes in political decisions during elections (Barakat, 2023). The increasing electricity demand from data centres has raised concerns about grid capacity among utility companies and local officials in regions such as Northern Virginia, Atlanta, and South Carolina (O'Donovan, 2024). Although climate change is occasionally cited as a concern, opposition to data centres in the US seems to be primarily driven by practical issues directly affecting local residents, often characterised as “not in my backyard” (NIMBY) movements.

Opposing these projects can be challenging, however, as they are often developed behind closed doors, with county officials signing nondisclosure agreements with tech companies that conceal details such as company names, building plans, energy requirements, and other information deemed “proprietary” by the tech industry (Samuelson, 2024). Civil society organisations such as the Virginia Data Center Reform Coalition, led by the Piedmont Environmental Council, are urging state lawmakers “to study the cumulative effects of data centre development on Virginia's electrical grid, water resources, air quality, and land conservation efforts, and to institute several common-sense regulatory and rate-making reforms for the industry” (Virginia Data Center Reform Coalition – The Piedmont Environmental Council, 2024).

Europe is also an important hub for an ever-increasing number of data centres: “In 2022, data centres used an estimated 15 TWh in Germany, equivalent to around 3% of national electricity. In France, this number was around 10 TWh of electricity, equivalent to 2.2% of national electricity use.” Furthermore, “[d]ata centres represent a significant share of national electricity use in Ireland (18%), the Netherlands (5.2%), Luxembourg (4.8%), Denmark (4.5%), and Germany (3%), Sweden (2.3%), and France (2.2%)" (European Commission, Joint Research Centre et al., 2024). Ireland has long been a key hub for foreign tech companies and their data centres. Even before the generative AI boom, data centres were expected to account for 27% of the country's electricity consumption by 2029 (EirGrid & Soni, 2020, 14). The intensive data centre construction in Ireland has faced protests, petitions, and other actions, particularly from a group called "Not Here, Not Anywhere." Going beyond NIMBY movements, the group has called for a moratorium on data centre development until a policy that is in line with the Paris Agreement on climate change is introduced with its "Press Pause on Data Centres" campaign (Not Here Not Anywhere, 2023).

Another example of local resistance to data centre projects occurred in East London where the Havering Council approved plans for a £5.3 billion data centre proposed by a company called Digital Reef.[39] Once completed, the facility was set to become the largest data centre in Europe. However, the project has faced opposition from residents and civil society, especially from the Campaign to Protect Rural England (CPRE) and Havering Friends of the Earth. The campaigns raised concerns about the environmental impact of the project and questioned the project's pledges to create jobs, since much of the job creation may just pertain to the construction phase and may exclude local residents (The Havering Daily, 2024). As of 2024, the proposal remains under debate, with discussions continuing regarding its environmental and community impact (Mann, 2024).

In the Netherlands, an example of opposition to data centre projects is the case of Microsoft Azure data centres in the Wieringermeer polder, where local politicians,

[39] A dedicated council website is available at https://www.havering.gov.uk/regeneration-3/east-havering-datacentre-ecology-park

civil society groups and citizens united in opposition in 2021 when Microsoft moved to build additional data centres in the region. The project left farmers and citizens concerned about water availability during heatwaves and potential soil pollution from chemicals used in the cooling process. There were also transparency issues, as a water authority employee was denied entry to certain sites for security reasons, and local environmental regulators could not determine the exact chemicals used due to trade secret protections (Rone, 2024, 6007). Grassroots movements in the region such as Data Non Grata and Red de Wieringermeer have been striving to protect green landscapes and open agricultural areas, while also aiming to prevent excessive portable water and renewable energy use, as well as air and water pollution. Despite the opposition, Microsoft reportedly began the construction of its new hyperscale facilities before obtaining a final official permission (Rone, 2024, 6008).

Data centre investments by American tech firms have also sparked resistance in South America. In Chile, which is affected by prolonged drought, Google's data centre project in Cerrillos, a neighbourhood of Santiago, was met with backlash as the local community organised protests, raising concerns about water consumption and environmental impact at large. Google was originally authorised to extract 228 liters of water per second from underground wells under Chile's Environmental Impact Evaluation System (Urquieta et al., 2024). Local residents and the Cerrillos municipality filed administrative claims against the project, prompting Google to propose a less water-intensive cooling system. The project was ultimately suspended (ibid). Similarly, a group in Quilicura, another Santiago neighbourhood, has taken action against a recent Microsoft data centre project (Feliba, 2023).

A similar example was reported in Uruguay in August 2024 due to Google's plans to build a data centre which will be cooled using air conditioning as opposed to water, with critics warning that the project could release thousands of tonnes of carbon dioxide and hazardous waste. Initially, the company proposed using fresh water for cooling, which led to public outcry in light of the country's severe drought and water shortages. In response, Google redesigned the facility to use air conditioning instead. While Uruguay generates over 90% of its electricity from renewable sources, environmentalists, such as María Selva Ortiz from Friends of the Earth, argued that the data centre would strain the country's energy grid, potentially increasing reliance on fossil fuels (Livingstone, 2024).

5.6 AI in Defence and Security Sectors

Advances in AI-enabled technologies have influenced the defence and security sectors, marking a shift akin to the transformative impact of nuclear technology. From autonomous weapons systems (AWS) to AI-driven decision-support systems (AI-DSSs), AI may offer benefits such as efficiency, improved precision, and the ability to process complex data at unprecedented speeds in specific contexts. While the adoption of AI technologies in military applications is gaining momentum, this trend is accompanied by criticism and resistance, reflecting ethical, legal, security and humanitarian concerns.

5.6.1 Autonomous Weapon Systems (AWS)

The resistance to the military adoption of AI emerged with the introduction of semi-autonomous weapon systems (AWS), including drones, robotic tanks, naval vehicles, and other platforms capable of engaging targets with minimal or potentially no human intervention. Currently, many states avoid deploying such systems in domestic law enforcement, as public resistance—exemplified by the swift withdrawal of San Francisco's 2022 approval for lethal police robots following widespread outcry—

remains a barrier (Romine, 2022). However, such systems are increasingly deployed in conflict zones. Many contend that these systems disrupt the *jus ad bellum* doctrines by lowering the perceived costs of war, thereby potentially increasing the likelihood of conflict, and challenge the *jus in bello* framework by undermining principles such as distinction (protection of civilians), proportionality, humanity, military necessity, accountability, and human control over lethal decision-making (Şimşek, 2017).

In particular, the deployment of unmanned aerial vehicles (UAVs), which have long been integral to modern warfare for roles such as reconnaissance and strikes, have drawn attention to the implications of AWS since the early 2000s. The US military's UAV operations in the Middle East in early 2000s represented a shift from conventional warfare to a "manhunt" doctrine, focused on identifying and eliminating high-value targets through drone strikes (Chamayou, 2011). In 2007, Sharkey drew attention to the use of unmanned weapons by the US military in Iraq and Afghanistan and warned that the next step would be fully autonomous robot warriors, despite the lack of an ethical framework for their use (Sharkey, 2007). In 2012, Human Rights Watch published a comprehensive report detailing the ethical and legal concerns associated with fully AWS titled "Losing Humanity: The Case Against Killer Robots" (Human Rights Watch, 2012). The report highlighted, for example, that a robot would not have any emotions or compassion which act as checks on the killing of civilians during armed conflict (ibid, 4, 38). It also raised questions about accountability, noting that holding developers, manufacturers or commanders liable for civilian harm under international law would not have any meaningful deterrence effect as long as robots remained fully autonomous (ibid, 44). The report called on states to pre-emptively ban the development and deployment of AWS due to the threats they pose to civilians (ibid, 46).

The following year, a global campaign was launched in the UK by a coalition of academics, human rights advocacy groups, and Nobel Peace Prize laureates, also calling for a pre-emptive international ban on fully autonomous weapons (McVeigh, 2013). Starting in 2014, the United Nations Convention on Certain Conventional Weapons (CCW) began convening diplomatic discussions to assess regulatory frameworks for AWS (Congressional Research Service, 2023). Since 2018, the United Nations Secretary-General, António Guterres, has repeatedly denounced lethal autonomous weapon systems as both “politically unacceptable” and “morally repugnant,” calling for their prohibition under international law (Guterres, 2018). He reiterated this position in 2023, urging states to conclude a legally binding agreement by 2026 to outlaw fully autonomous weapons operating without human oversight and in breach of international humanitarian law, while also implementing strict regulation for other AWS (Guterres, 2023). This position is shared by the President of the International Committee of the Red Cross (UN Secretary-General & President of the ICRC, 2023). AWS have also faced opposition on religious grounds, notably from Pope Francis, who has urged G7 leaders to impose a ban on autonomous weapons (Wintour, 2024).

Last but not least, states are increasingly supporting a ban on lethal AWS, as evidenced by the adoption of Resolution L.77 on 5 November 2024 by the First Committee of the UN General Assembly, repeating the calls for a legally binding instrument with clear prohibitions and regulation of AWS.[40] This marked the second consecutive year such a resolution was passed. The resolution highlighted the “serious challenges and concerns that new and emerging technological applications

[40] The resolution is available at https://documents.un.org/doc/undoc/ltd/n24/305/45/pdf/n2430545.pdf

in the military domain, including those related to artificial intelligence and autonomy in weapons systems, also raise from humanitarian, legal, security, technological and ethical perspectives”. It garnered the support of 161 states, with 3 opposing and 13 abstaining (Jones, 2024).

5.6.2 AI-Driven Decision Support Systems (AI-DSS)

Another prevalent application of AI in the context of warfare is the deployment of AI-DSS. Employing AI-DSS on the battlefield may offer advantages, including enhanced situational awareness, accelerated decision-making, and superior command and control systems. Such systems may possess the ability to move beyond human limitations, particularly in complex and high-pressure combat scenarios (Johnson, 2019, 150). On the other hand, certain psychological dynamics that emanate from human-machine interactions, including the uncritical attribution of moral superiority to military technologies and the anthropomorphisation of machines, coupled with decision-makers’ tendency to over-utilise AI due to substantial investments (the so-called Einstellung effect), risk creating a “de facto AI commander” problem (J. Johnson, 2022).

In particular, AI-DSS risk fostering “automation bias”, which refers to “the tendency for humans to depend excessively on automated systems” (House of Lords, 2023, 33). Automation bias undermines information verification and human oversight, reducing human decision-making to box-checking which makes AI-DSS ill-suited for urban warfare involving civilians (ICRC & Geneva Academy, 2024). Notably, employing an AI-DSS to nominate targets is inherently unsafe from the perspective of international humanitarian law, especially when there is insufficient time to identify and rectify potential errors. Moreover, the accuracy of such systems is contested, particularly as they often depend on incomplete or flawed data. For instance, AI-DSS like “Lavender”, “Gospel”, and “Where’s Daddy” were reportedly used by the Israeli military to identify military targets in Gaza, but the accuracy of these systems and their capacity to distinguish between military targets and civilians have been put into question (Nadibaidze et al., 2024; Human Rights Watch, 2024). Yet, an increasing number of AI-DSS are being pitched to military organisations, often relying on open-source intelligence which is widely regarded as prone to inaccuracies and misinformation (Khlaaf et al., 2024, 7). In this context, civil society movements have increasingly highlighted concerns about “digital dehumanization,” a process that reduces individuals to mere data points, stripping away their humanity (Automated Decision Research, 2022).

A recent example of resistance to military contracts for AI systems used in identifying targets comes from within tech companies. Following the initiation of Project Nimbus—a $1.2 billion cloud computing contract signed in 2021 between the Israeli government and tech giants Google and Amazon—employees at Google and Amazon began the “No Tech for Apartheid” campaign (Chan & Grantham, 2024; Biesecker et al., 2025; Biesecker, 2025). The Associated Press reported that around 50 Google employees were fired because they took part in protests against Google’s ties to the Israeli military (ibid; see also Biesecker et al., 2025). Similarly, in February 2025, five Microsoft employees were reportedly removed from a company town hall meeting after protesting against the company’s contracts to provide AI and cloud computing services to the Israeli military—technologies reportedly used to select bombing targets in Gaza and Lebanon (Biesecker, 2025). The protesting employees wore T-shirts spelling out “Does Our Code Kill Kids, Satya?” directed at CEO Satya Nadella. It was likely a reference to the Associated Press reporting on the Microsoft ties to the Israeli military which included the story of an Israeli bombing of civilians in November 2023 that killed three young girls (Biesecker et al., 2025; Biesecker, 2025).

5.6.3 Repurposing Commercial Foundation Models for Military and Security Applications

AI models possess a dual-use nature, presenting additional ethical and governance challenges as applications originally intended for civilian use can be repurposed for military applications (Khlaaf et al., 2024). The increasing interest in fine-tuning commercial foundation models for military applications is a case in point. The US Department of Defense has established a Generative AI Task Force (US Department of Defense, 2023). In January 2024, OpenAI quietly removed the explicit prohibition against the military use of its technologies from its usage policies (Field, 2024). It was also reported that Microsoft has pitched OpenAI's DALL-E image generator to the Pentagon for military use cases (Biddle, 2024). In a similar vein, Meta has announced that its open-source Llama models are used by US government agencies engaged in defense and national security initiatives as the company collaborates with private sector entities, such as Accenture Federal Services, Amazon Web Services, Lockheed Martin, Microsoft, and Palantir, to facilitate the integration of Llama into governmental operations and applications (Clegg, 2024). The announcement followed reports that Chinese researchers had utilised an earlier version of Meta's Llama 13B "to construct a military-focused AI tool to gather and process intelligence, and offer accurate and reliable information for operational decision-making." (Pomfret & Pang, 2024). Fine-tuning commercial foundation models for military applications is criticised for potentially leading to unintended consequences, including increased civilian casualties, misuse of personally identifiable information, and heightened geopolitical tensions, particularly in the absence of comprehensive policy frameworks to regulate this practice (Khlaaf et al., 2024).

6 Conclusion

Our research has revealed that AI - unsurprisingly - represents a major techno-economic paradigm shift, and has ignited profound, multifaceted resistance anchored in deep-seated socio-economic, ethical, environmental, legal and political thinking and concerns (Section 4). This resistance is not an outright rejection of “progress” but represents efforts to shape the future of this technology in a way that is aligned with established human values, including human dignity.

To begin with, we observed that civil society—including grassroots movements, worker unions, and NGOs—has a notable presence in resistance to AI. We found that citizens have also individually engaged in resistance movements, such as student protests against the UK government’s use of an algorithmic system to predict A-level results. However, due to the power imbalances that characterise many instances of AI deployment, civil society has been instrumental in defending citizens’ rights and interests vis-à-vis governments and the private sector. Civil society has organised and empowered resistance in the context of migration and border control, creative industries, and environmental protection among others.

Private corporations that we surveyed in this report vary in scale—from small and medium enterprises to dominant multinational technology corporations. “Big Tech” companies came up across almost all cases of resistance to AI that we surveyed in this report as *targets* of resistance, from resistance to environmental harms and military applications of AI to resistance in creative industries.

In terms of industrial segments, creative industries and the media appear to be at the forefront of resistance to AI, particularly concerning the unauthorized use of artistic works and publications for the training of generative AI models. This is exemplified by lawsuits and strikes discussed above under Section 5.1. We also found evidence of resistance to AI among physicians, especially due to the legal uncertainties on liability where patients are harmed when medical AI is in use (Section 5.3). Finally, our research has shown a well-established resistance to the development and deployment of fully autonomous weapons which has been growing in recent years as AI development has garnered speed in the 2020s (Section 5.6.1).

Public institutions, including governmental agencies, regulatory authorities, state-affiliated organizations, and intergovernmental and supranational entities, act both as deployers of AI systems and as regulators. In their role as deployers, public entities have faced resistance largely due to the ethical implications of such deployment, for example in the context of education and migration, and due to public-private partnerships. In terms of regulation, public institutions play a role in addressing and mitigating the negative consequences of AI. They even formalise cases of resistance by, for example, prohibiting specific use cases of AI which are deemed to be too harmful to be acceptable.

On a final note, our research has shown that concerns underlying resistance to AI, ranging from ethical issues and economic disruption to environmental consequences and regulatory gaps, are interconnected and demand expertise from a range of disciplines, including but not limited to computer science, ethics, law, economics, sociology, and environmental studies. Furthermore, these concerns transcend national boundaries. Therefore, cooperation among public, private, academic and civil society groups targeted to address these concerns must also transcend national boundaries. Finally, AI development should not be primarily relinquished to private actors, and public oversight is necessary to safeguard citizens' rights and interests.

7 Bibliography

Acemoglu, D., & Johnson, S. (2023). *Power and Progress: Our Thousand-year Struggle Over Technology and Prosperity*. Basic Books.

Adams, R., Weale, S., & Barr, C. (2020, August 13). *A-level results: almost 40% of teacher assessments in England downgraded*. The Guardian. Retrieved December 30, 2024, from https://www.theguardian.com/education/2020/aug/13/almost-40-of-english-students-have-a-level-results-downgraded

Agnew, W., McKee, K. R., Gabriel, I., Kay, J., Isac, W., Bergman, S., El-Sayed, S., & Mohamed, S. (2023, November). *Technologies of Resistance to AI*. ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization. https://conference2023.eaamo.org/papers/EAAMO23_paper_154.pdf

Allyn, B. (2022, March 16). *A deepfake video showing Volodymyr Zelenskyy surrendering worries experts*. NPR. Retrieved November 30, 2024, from https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia

Almada, M., & Petit, N. (2023). *The EU AI Act: A Medley of Product Safety and Fundamental Rights?* [Robert Schuman Centre for Advanced Studies Working Paper 2023/59]. https://cadmus.eui.eu/bitstream/handle/1814/75982/RSC_WP_2023_59.pdf?sequence=1&isAllowed=y

Amnesty International. (2021, 10 25). *Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal*. https://www.amnesty.org/en/documents/eur35/4686/2021/en/

Amoore, L. (2024, 03). The deep border. *Political Geography*, *109*(102547). https://doi.org/10.1016/j.polgeo.2021.102547

Ananya. (2024, March 19). *AI image generators often give racist and sexist results: can they be fixed?* Nature. https://www.nature.com/articles/d41586-024-00674-9?

Anguiano, D., & Beckett, L. (2023, October 1). How Hollywood writers triumphed over AI – and why it matters. *The Guardian*. https://www.theguardian.com/culture/2023/oct/01/hollywood-writers-strike-artificial-intelligence

Astromskė, K., Peičius, E., & Astromskis, P. (2020, 08 27). Ethical and legal challenges of informed consent applying artificial intelligence in medical diagnostic consultations. *AI & Society*, *36*, 509-520. https://doi.org/10.1007/s00146-020-01008-9

Atherton, D. (2023, 12 26). *Incident 626: Social Media Scammers Used Deepfakes of Taylor Swift and Several Other Celebrities in Fraudulent Le Creuset Cookware Giveaways*. AI Incident Database. Retrieved December 15, 2024, from https://incidentdatabase.ai/cite/626/#r3578

Auer, M. E., Langmann, R., May, D., & Roos, K. (Eds.). (2024). *Smart Technologies for a Sustainable Future: Proceedings of the 21st International Conference on Smart Technologies & Education. Volume 1*. Springer Nature Switzerland.

Automated Decision Research. (2022, November). *Autonomous weapons and digital dehumanisation*. Automated Decision Research.

https://automatedresearch.org/wp-content/uploads/2022/12/Autonomous-weapons-and-digital-dehumanization-Report-Single-Page.pdf

Barakat, M. (2023, June 22). *Backlash to data centers prompts political upset in northern Virginia*. AP News. Retrieved November 10, 2024, from https://apnews.com/article/virginia-election-data-centers-prince-william-229cb44d34ccf4bd1cc4e9f0d0131649

Bauer, M. (1995). *Resistance to new technology and its effects on nuclear power, information technology and biotechnology*. CORE. Retrieved November 18, 2024, from http://eprints.lse.ac.uk/39607/1/Resistance_to_new_technology_and_its_effects_on_nuclear_power%2C_information_technology_and_biotechnology_%28LSERO%29.pdf

BBC. (2018, October 10). *Amazon scrapped 'sexist AI' tool*. BBC. Retrieved May 14, 2025, from https://www.bbc.com/news/technology-45809919?utm

BBC. (2020, February 20). *Barclays scraps 'Big Brother' staff tracking system*. BBC. Retrieved December 15, 2024, from https://www.bbc.co.uk/news/business-51570401

Bedingfield, W. (2023, September 27). *Hollywood Writers Reached an AI Deal That Will Rewrite History*. WIRED. Retrieved December 15, 2024, from https://www.wired.com/story/us-writers-strike-ai-provisions-precedents/

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* FAccT, Virtual Event, Canada. https://doi.org/10.1145/3442188.3445922

Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Darrell, T., Harari, Y. N., Zhang, Y.-Q., Xue, L., Shalev-Shwartz, S., Hadfield, G., Clune, J., Maharaj, T., Hutter, F., Baydin, A. G., McIlraith, S., Gao, Q., Acharya, A., Krueger, D., … Mindermann, S. (2024, May 20). *Managing extreme AI risks amid rapid progress*. Science Vol 384, Issue 6698 pp. 842-845. https://www.science.org/doi/10.1126/science.adn0117

Bengio, Y., Privitera, D., & Mindermann, S. (2025). *International AI Safety Report The International Scientific Report on the Safety of Advanced AI*. AI Action Summit.

Biddle, S. (2024, April 10). *Microsoft Pitched OpenAI's DALL-E as Battlefield Tool for U.S. Military*. The Intercept. https://theintercept.com/2024/04/10/microsoft-openai-dalle-ai-military-use/

Biesecker, M. (2025, February 25). *Microsoft workers protest sale of AI and cloud services to Israeli military*. AP News. Retrieved April 5, 2025, from https://apnews.com/article/israel-palestinians-ai-technology-microsoft-gaza-lebanon-90541d4130d4900c719d34ebcd67179d?utm

Biesecker, M., Mednick, S., & Burke, G. (2025, February 18). *How US tech giants' AI is changing the face of warfare in Gaza and Lebanon*. AP News. Retrieved April 5, 2025, from https://apnews.com/article/israel-palestinians-ai-technology-737bc17af7b03e98c29cec4e15d0f108?utm

Birhane, A. (2020, August). Algorithmic Colonization of Africa. *ScriptED*, *17*(2), 389-409. DOI: 10.2966/scrip.170

Boden, M. A. (2009). Computer Models of Creativity. *AI magazine*, *30*(3), 23-34. https://doi.org/10.1609/aimag.v30i3.2254

Boden, M. A., & Edmonds, E. A. (2010, 12 1). What is generative art? *Digital Creativity*, *20*(1-2), 21-46. https://doi.org/10.1080/14626260902867915

Booth, H. (2024, October 25). *What Teenagers Really Think About AI*. TIME. https://time.com/7098524/teenagers-ai-risk-lawmakers/

Boullier, D. (2024, June 4). *SOCIAL MEDIA RESET: Redesigning the infrastructure of digital propagation to cut the chains of contagion*. Sciences Po. https://www.sciencespo.fr/public/chaire-numerique/wp-content/uploads/2024/06/Dominique-Boullier-Social-Media-Reset_compressed.pdf

Bria, F., Timmers, P., & Gernone, F. (2025, 02 13). *EuroStack – A European Alternative for Digital Sovereignty*. www.bertelsmann-stiftung.de. https://www.bertelsmann-stiftung.de/en/publications/publication/did/eurostack-a-european-alternative-for-digital-sovereignty

Broeders, D., & Dijstelbloem, H. (2020). The Datafication of Mobility and Migration Management: the Mediating State and its Consequences. In I. v. d. Ploeg & J. Pridmore (Eds.), *Digitizing Identities: Doing Identity in a Networked World* (pp. 242-260). Routledge.

Bryson, J. J. (2021). The Artificial Intelligence of the Ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI*. Oxford University Press.

Bryson, J. J., & Kime, P. P. (2011). *Just an Artifact: Why Machines are Perceived as Moral Agents* [Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence]. https://static1.squarespace.com/static/5e13e4b93175437bccfc4545/t/5eaeed32c3dfb005b72d4d76/1588522290510/just-an-artifact.pdf

Bui, M. L., & Noble, S. U. (2021). We're missing a moral framework of justice in artificial intelligence: on the limits, failings, and ethics of fairness. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI* (pp. 163-180). Oxford University Press.

Bundesverfassunsgsgericht. (1983, December 15). *1 BvR 209, 269, 362, 420, 440, 484/83 - Decision on the constitutionality of the 1983 Census Act*. Bundesverfassungsgericht. Retrieved November 18, 2024, from https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/1983/12/rs19831215_1bvr020983en.html

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society,*, (January-June), 1-12.

Cadwalladr, C. (2017, May 7). *The great British Brexit robbery: how our democracy was hijacked*. The Guardian. https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy

Carroll, M., Chan, A., Ashton, H., & Krueger, D. (2023, November 1). *Characterizing Manipulation from AI Systems*. Equity and Access in Algorithms, Mechanisms, and Optimization. https://dl.acm.org/doi/pdf/10.1145/3617694.3623226

Center for AI safety. (2023, May). *Statement on AI Risk | CAIS*. Center for AI Safety. Retrieved November 28, 2024, from https://www.safe.ai/work/statement-on-ai-risk

Center for Research on Foundation Models (CRFM) & Stanford Institute for Human-Centered Artificial Intelligence (HAI). (2021). *On the opportunities and risks of foundation models*. Stanford University. Retrieved from, https://crfm.stanford.edu/assets/report.pdf

Cestonaro, C., Delicati, A., Marcante, B., Caenazzo, L., & Tozzo, P. (2023). Defining medical liability when artificial intelligence is applied on diagnostic algorithms: a systematic review. *Frontiers in Medicine*, *10*. https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2023.1305756/full

Chamayou, G. (2011). The manhunt doctrine. *Radical Philosophy*, *169*.

Chan, K., & Grantham, W. (2024, April 23). Google fires more workers who protested its deal with Israel. *AP News*. https://apnews.com/article/google-israel-protest-workers-gaza-palestinians-96d2871f1340cb84c953118b7ef88b3f

Chandran, R. (2023, April 2). *NZ, US Indigenous fear colonisation as bots learn their languages | Context*. Context News. Retrieved December 11, 2024, from https://www.context.news/ai/nz-us-indigenous-fear-colonisation-as-bots-learn-their-languages

Chesney, B., & Citron, D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review, Inc.*, *107*(6), 1753-1820.

Chmielewski, D., Paul, K., & Blackwell, H. (2024, October 21). *Murdoch's Dow Jones, New York Post sue Perplexity AI for 'illegal' copying of content*. Reuters. Retrieved December 17, 2024, from https://www.reuters.com/legal/murdoch-firms-dow-jones-new-york-post-sue-perplexity-ai-2024-10-21/

Chouliaraki, L., & Georgiou, M. (2022). *The Digital Border: Migration, Technology, Power*. NYU Press.

Christl, W. (2024, November). *Tracking Indoor Location, Movement and Desk Occupancy in the Workplace*. Cracked Labs. https://crackedlabs.org/dl/CrackedLabs_Christl_IndoorTracking.pdf

Clegg, N. (2024, November 4). *Open Source AI Can Help America Lead in AI and Strengthen Global Security | Meta*. Meta. Retrieved November 26, 2024, from https://about.fb.com/news/2024/11/open-source-ai-america-global-security/

Congressional Research Service. (2023, February 14). *International Discussions Concerning Lethal Autonomous Weapon Systems*. Congress.gov. https://crsreports.congress.gov/product/pdf/IF/IF11294

Crevier, D. (1993). *AI : the tumultuous history of the search for artificial intelligence*. Basic Books.

Cristiano, F., Delerue, F., Douzet, F., Géry, A., & Broeders, D. (2023, May 11). *Artificial Intelligence and International Conflict in Cyberspace | Fab*. Taylor & Francis eBooks.

Davies, D. (2020, 08 15). *This year's A-level results are a fiasco – but the system was already broken*. The Guardian. https://www.theguardian.com/commentisfree/2020/aug/15/a-level-results-system-ofqual-england-exam-marking

de Graaf, A. (2025, February 21). *Fact check: How Elon Musk meddled in Germany's elections*. Alima de Graaf. https://www.dw.com/en/how-elon-musk-meddled-in-germanys-elections/a-71676473

De Liban, K. (2024, November). *Inescapable AI: The Ways AI Decides How Low-Income People Work, Live, Learn, and Survive*. Techtonic Justice. https://www.techtonicjustice.org/reports/inescapable-ai

Dennett, D. C. (2023, May 16). *The Problem With Counterfeit People*. The Atlantic. Retrieved December 15, 2024, from https://www.theatlantic.com/technology/archive/2023/05/problem-counterfeit-people/674075/

Dertouzos, M. L., & Moses, J. (Eds.). (1979). *The Computer Age: A Twenty-year View*. MIT Press.

Dhawan, E., & Xue, A. (2020, 12 24). *Stanford Medicine tosses original algorithm, allocates more vaccines to front-line residents and fellows*. The Stanford Daily. https://stanforddaily.com/2020/12/24/stanford-medicine-tosses-original-algorithm-allocates-more-vaccines-to-front-line-residents-and-fellows/

Dial, S. (2024, December 11). *Take It Down Act combatting 'deepfakes' revenge porn passes U.S. Senate*. Yahoo News. https://www.yahoo.com/news/down-act-combatting-deepfakes-revenge-220529609.html

EirGrid, & Soni. (2020). *All-Island Generation Capacity Statement*. | Eirgrid. Retrieved November 10, 2024, from https://cms.eirgrid.ie/sites/default/files/publications/All-Island-Generation-Capacity-Statement-2020-2029.pdf

Ertel, P. (2024, April 3). *UK healthcare workers blockade NHS entrance to protest deal with military tech firm*. Middle East Eye. https://www.middleeasteye.net/news/uk-nhs-workers-strike-against-company-supplying-israeli-military

Essinger, J. (2004). *Jacquard's Web: How a Hand-loom Led to the Birth of the Information Age*. Oxford University Press, USA.

European Commission. (2018, October 24). *Smart lie-detection system to tighten EU's busy borders*. Research and innovation. Retrieved December 31, 2024, from https://projects.research-and-innovation.ec.europa.eu/en/projects/success-stories/all/smart-lie-detection-system-tighten-eus-busy-borders

European Commission. (2019, 04 08). *Ethics guidelines for trustworthy AI*. Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Commission. (2025, February 11). *EU launches InvestAI initiative to mobilise €200 billion of investment in artificial intelligence*. Shaping Europe's digital future. Retrieved April 4, 2025, from https://digital-strategy.ec.europa.eu/en/news/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence

European Commission, Joint Research Centre, Bertoldi, P., & Kamiya, G. (2024). *Energy Consumption in Data Centres and Broadband Communication Networks in the EU*. https://joint-research-centre.ec.europa.eu

Feliba, D. (2023, September 26). FEATURE-In Latin America, data center plans fuel water worries. *Reuters*. https://www.reuters.com/article/technology/feature-in-latin-america-data-center-plans-fuel-water-worries-idUSL8N3AU1PY/

Fergusson, G. (2023, September 14). *Outsourced & Automated*. epic.org. https://epic.org/wp-content/uploads/2023/09/FINAL-EPIC-Outsourced-Automated-Report-w-Appendix-Updated-9.26.23.pdf

Field, H. (2024, January 16). *OpenAI quietly removes ban on military use of its AI tools*. CNBC. https://www.cnbc.com/2024/01/16/openai-quietly-removes-ban-on-military-use-of-its-ai-tools.html

Forbrukerradet. (2023, June). *GHOST IN THE MACHINE Addressing the consumer harms of generative AI*.

https://storage02.forbrukerradet.no/media/2023/06/generative-ai-rapport-2023.pdf

Foxglove. (2020, 08 04). *Home Office says it will abandon its racist visa algorithm – after we sued them*. Foxglove. https://www.foxglove.org.uk/2020/08/04/home-office-says-it-will-abandon-its-racist-visa-algorithm-after-we-sued-them/

Freeman, C., & Louçã, F. (2001). *As Time Goes by: From the Industrial Revolutions to the Information Revolution*. Oxford University Press.

Frontier Model Forum. (2023, July 26). OpenAI. Retrieved December 14, 2024, from https://openai.com/index/frontier-model-forum/

Frontier Model Forum. (2025, March 28). *FMF Announces First-Of-Its-Kind Information-Sharing Agreement*. Frontier Model Forum. https://www.frontiermodelforum.org/updates/fmf-announces-first-of-its-kind-information-sharing-agreement/

Future of Life Institute. (2015, October 28). *Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter*. Future of Life Institute. Retrieved November 19, 2024, from https://futureoflife.org/open-letter/ai-open-letter/

Future of Life institute. (2016, February 9). *Autonomous Weapons Open Letter: AI & Robotics Researchers*. Future of Life Institute. Retrieved November 14, 2024, from https://futureoflife.org/open-letter/open-letter-autonomous-weapons-ai-robotics/

Gajjar, D. (2024, January 23). *Artificial intelligence (AI) glossary*. UK Parliament. https://post.parliament.uk/artificial-intelligence-ai-glossary/

Gallagher, R., & Jona, L. (2019, July 26). *We Tested Europe's New Digital Lie Detector. It Failed.* The Intercept. Retrieved December 31, 2024, from https://theintercept.com/2019/07/26/europe-border-control-ai-lie-detector/

Gervais, D. J. (2020). The Machine as Author. *Iowa Law Review*, *105*(5), 2053-2106.

Goldenfein, J. (2024). Privacy's Loose Grip on Facial Recognition Law and the Operational Image (R. Matulionytė & M. Zalnieriute, Eds.). In *The Cambridge Handbook of Facial Recognition in the Modern State*. Cambridge University Press.

Google ML Education. (2022, July 18). *Background: What is a Generative Model? | Machine Learning*. Google for Developers. Retrieved December 2, 2024, from https://developers.google.com/machine-learning/gan/generative

Gordon, A. (2024, May 13). *Why Protesters Around the World Are Demanding a Pause on AI Development*. TIME. https://time.com/6977680/ai-protests-international/

Gross, A., & Hughes, L. (2024, November 21). *NHS take-up of Palantir data platform rises despite hurdles*. Financial Times. Retrieved December 10, 2024, from https://www.ft.com/content/9efae6c4-c039-49b9-bbe6-dcac575cb4a5

G'sell, F. (2024, September). *REGULATING UNDER UNCERTAINTY: Governance Options for Generative AI*. Stanford Cyber Policy Center. https://cyber.fsi.stanford.edu/content/regulating-under-uncertainty-governance-options-generative-ai

G'sell, F. (2025). Digital Authoritarianism: from state control to algorithmic despotism. In *Oxford Handbook of Digital Constitutionalism* (pp. 1-53). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5117399

G’sell, F., Ayar, A., & Gillman, Z. (2024, October 28). *[ARTICLE] California’s SB1047 vs EU AI Act: A Comparative Analysis of AI Regulation*. Sciences Po. https://www.sciencespo.fr/public/chaire-numerique/en/2024/10/28/californias-sb1047-vs-eu-ai-act-a-comparative-analysis-of-ai-regulation/

Guardian Community Team. (2020, 08 17). *Have you been protesting against A-level downgrading?* The Guardian. https://www.theguardian.com/commentisfree/2020/aug/15/a-level-results-system-ofqual-england-exam-marking

Guo, E., & Hao, K. (2020, December 21). *This is the Stanford vaccine algorithm that left out frontline doctors*. MIT Technology Review. https://www.technologyreview.com/2020/12/21/1015303/stanford-vaccine-algorithm/

Guterres, A. (2018, November 05). *Remarks at "Web Summit"*. United Nations. https://www.un.org/sg/en/content/sg/speeches/2018-11-05/remarks-web-summit

Guterres, A. (2023, JULY). *Our Common Agenda Policy Brief 9: A New Agenda for Peace*. https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-new-agenda-for-peace-en.pdf

Harari, Y. N. (2024). *Nexus: A Brief History of Information Networks from the Stone Age to AI*. Random House UK.

Hassabis, D. (2016, January 27). *AlphaGo: using machine learning to master the ancient game of Go*. Google blog. https://blog.google/technology/ai/alphago-machine-learning-game-go/

The Havering Daily. (2024, May 22). Local residents huge concerns over Data Centre planning application and the costly environmental impact it will have on the community. *The Havering Daily*. https://thehaveringdaily.co.uk/2024/05/22/local-residents-huge-concerns-over-data-centre-planning-application-and-the-costly-environmental-impact-it-will-have-on-the-community/

Heikkilä, M. (2022, 03 29). *Dutch scandal serves as a warning for Europe over risks of using algorithms*. POLITICO.eu. Retrieved December 16, 2024, from https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/

Helfrich, G. (2024). The harms of terminology: why we should reject so-called “frontier AI”. *AI and Ethics*, *4*, 699-705. https://link.springer.com/article/10.1007/s43681-024-00438-1#ref-CR39

Hildebrandt, M. (2019). Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning. *Theoretical Inquiries in Law*, *20*(1), 83-121.

Hirmiz, R. (2024). Against the substitutive approach to AI in healthcare. *AI and Ethics*, *4*, 1507–1518. https://link.springer.com/article/10.1007/s43681-023-00347-9

Holligan, A. (2021, January 15). *Dutch Rutte government resigns over child welfare fraud scandal*. BBC. Retrieved December 16, 2024, from https://www.bbc.com/news/world-europe-55674146

Hornung, G., & Schnabel, C. (2009). Data protection in Germany I: The population census decision and the right to informational self-determination. *Computer Law Security Review*, *25*, 84-88. https://doi.org/10.1016/J.CLSR.2008.11.002.

House of Lords. (2023, December 1). *Proceed with Caution: Artificial Intelligence in Weapon Systems* [Report of Session 2023–24]. HOUSE OF LORDS AI in

Weapon Systems Committee. https://publications.parliament.uk/pa/ld5804/ldselect/ldaiwe/16/16.pdf

Howcroft, D., & Bergvall-Kåreborn, B. (2018, May 4). A Typology of Crowdwork Platforms. *33*(1). https://journals.sagepub.com/doi/full/10.1177/0950017018760136

Hu, M. (2020, July). Cambridge Analytica's black box. *Big Data & Society*, *7*(2). https://doi.org/10.1177/2053951720938091

Hulette, D. (2021, January 22). *How to Recognize AI Snake Oil | Computer Science Department at Princeton University*. cs.Princeton. Retrieved December 2, 2024, from https://www.cs.princeton.edu/news/how-recognize-ai-snake-oil

Human Rights Watch. (2012, November 19). *Losing Humanity : The Case against Killer Robots | HRW*. Human Rights Watch. Retrieved November 26, 2024, from https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots

Human Rights Watch. (2024, September 10). *Gaza: Israeli Military's Digital Tools Risk Civilian Harm New Technologies Raise Grave Laws-of-War, Privacy, Personal Data Concerns*. Human Rights Watch. https://www.hrw.org/news/2024/09/10/gaza-israeli-militarys-digital-tools-risk-civilian-harm

Hung, R. (2023, 20 02). *AI Technology and Art: US Judge Finds that AI-Generated Art Cannot Be Copyright-Protected*. Torkin Manes LegalPoint. https://www.torkin.com/insights/publication/ai-technology-and-art-us-judge-finds-that-ai-generated-art-cannot-be-copyright-protected

Huxley, A. (1937). *Ends and Means: An Enquiry into the Nature ofIdeals and into the Methods employed for their Realisation*. Oxford University Press Toronto.

ICO. (2017, July 3). *RFA0627721 – provision of patient data to DeepMind*. Information Commissioner's Office. https://ico.org.uk/media/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf

ICRC & Geneva Academy. (2024, March). *Expert Consultation Report on AI and Related Technologies in Military Decision-Making on the Use of Force in Armed Conflicts*. https://www.geneva-academy.ch/joomlatools-files/docman-files/Artificial%20Intelligence%20And%20Related%20Technologies%20In%20Military%20Decision-Making.pdf

IEA. (2024, January). *Electricity 2024: Analysis and forecast to 2026*. International Energy Agency. https://iea.blob.core.windows.net/assets/6b2fd954-2017-408e-bf08-952fdd62118a/Electricity2024-Analysisandforecastto2026.pdf

International Energy Agency. (2024, January 1). *Electricity 2024: Analysis and forecast to 2026*. https://iea.blob.core.windows.net/assets/6b2fd954-2017-408e-bf08-952fdd62118a/Electricity2024-Analysisandforecastto2026.pdf

Johnson, D. G. (2022). Algorithmic Accountability In The Making. *Social Philosophy & Policy*, *38*(2), 111-127. https://www.cambridge.org/core/journals/social-philosophy-and-policy/article/algorithmic-accountability-in-the-making/6F3CE994EC96C65392C5374B3CDE3C51

Johnson, J. (2019). Artificial intelligence & future warfare: implications for international security. *Defense & Security Analysis*, *35*(2), 147-169. DOI: 10.1080/14751798.2019.1600800

Johnson, J. (2022). The AI Commander Problem: Ethical, Political, and Psychological Dilemmas of Human-Machine Interactions in AI-enabled

Warfare. *Journal of Military Ethics*, *21*(3-4), 246-271. DOI: 10.1080/15027570.2023.2175887

Jones, I. (2024, 11 05). *161 states vote against the machine at the UN General Assembly*. Stop Killer Robots. https://www.stopkillerrobots.org/news/161-states-vote-against-the-machine-at-the-un-general-assembly/

Jussupow, E., Spohrer, K., & Heinzl, A. (2022). Identity Threats as a Reason for Resistance to Artificial Intelligence: Survey Study With Medical Students and Professionals. *JMIR*, *6*(3).

Kachwala, Z., Eluri, K. C., & Holmes, S. (2024, October 15). *NYT sends AI startup Perplexity 'cease and desist' notice over content use*. Reuters. Retrieved December 17, 2024, from https://www.reuters.com/technology/artificial-intelligence/nyt-sends-ai-startup-perplexity-cease-desist-notice-over-content-use-wsj-reports-2024-10-15/

Kelly, A. (2021, June). A tale of two algorithms: The appeal and repeal of calculated grades systems in England and Ireland in 2020. *British Educational Research Journal*, *47*(3). https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/berj.3705

Khanal, S., Zhang, H., & Taeihagh, A. (2024). Why and how is the power of Big Tech increasing in the policy process? The case of generative AI. *Policy and Society*, *00*(00), 1-18. https://academic.oup.com/policyandsociety/advance-article/doi/10.1093/polsoc/puae012/7636223

Khlaaf, H., West, S. M., & Whittaker, M. (2024, October 18). Mind the Gap: Foundation Models and the Covert Proliferation of Military Intelligence, Surveillance, and Targeting. https://arxiv.org/pdf/2410.14831

Kirtchik, O. (2023). The Soviet scientific programme on AI: if a machine cannot ‘think’, can it‘control’? *BJHS Themes*, *8*, 111–125. doi:10.1017/bjt.2023.4

Kolkman, D. (2020, 08 26). *“F**k the algorithm”?: What the world can learn from the UK’s A-level grading fiasco*. LSE Blogs. https://blogs.lse.ac.uk/impactofsocialsciences/2020/08/26/fk-the-algorithm-what-the-world-can-learn-from-the-uks-a-level-grading-fiasco/

Kooijman, J. (2024, 03 26). *ChatGPT in higher education: Striving for equality and academic integrity*. Erasmus Initiative Societal Impact of AI. https://aipact.medium.com/chatgpt-in-higher-education-striving-for-equality-and-academic-integrity-cdb3a14baaa0

Lab Employee Statement on Extreme AI Risks. (2024). calltolead.org. https://calltolead.org/?utm_source=Euractiv&utm_campaign=f73617c35b-EMAIL_CAMPAIGN_2024_02_23_09_54_COPY_01&utm_medium=email&utm_term=0_-a1a15a7236-%5BLIST_EMAIL_ID%5D

LaFrance, A. (2017, June 20). What an AI's Non-Human Language Actually Looks Like. *The Atlantic*. https://www.theatlantic.com/technology/archive/2017/06/what-an-ais-non-human-language-actually-looks-like/530934/

Laranjeira de Pereira, J. R. (2024, April 8). *The EU AI Act and environmental protection: the case for a missed opportunity*. Heinrich-Böll-Stiftung European Union. https://eu.boell.org/en/2024/04/08/eu-ai-act-missed-opportunity

Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016, May 23). *How We Analyzed the COMPAS Recidivism Algorithm*. Propublica. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Law Commission of England and Wales, & Scottish Law Commission. (2022, 01 25). *Automated Vehicles: joint report*. https://lawcom.gov.uk/project/automated-vehicles/

Lawler, R. (2025, January 21). *The Stargate Project is a $500 million AI data center plan for OpenAI*. The Verge. Retrieved April 4, 2025, from https://www.theverge.com/2025/1/21/24348816/openai-softbank-ai-data-center-stargate-project

Lecher, C., & Castro, A. (2019, April 25). *How Amazon automatically tracks and fires warehouse workers for 'productivity'*. The Verge. Retrieved December 15, 2024, from https://www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centers-productivity-firing-terminations

Lehuedé, S. (2023, April 6). *With Google as My Neighbor, Will There Still Be Water? - AlgorithmWatch*. Algorithm Watch. Retrieved November 10, 2024, from https://algorithmwatch.org/en/protests-against-data-centers/

Leloup, D. (2024, 03 21). Il y a cinquante ans, un article du « Monde » déclenchait la création de la CNIL. *Le Monde*. https://www.lemonde.fr/pixels/article/2024/03/21/il-y-a-cinquante-ans-un-article-du-monde-declenchait-la-creation-de-la-cnil_6223203_4408996.html

Le Moli, G. (2022). *AI vs Human Dignity: When Human Underperformance is Legally Required*. Groupe d'études géopolitiques. Retrieved December 16, 2024, from https://geopolitique.eu/en/articles/ai-vs-human-dignity-when-human-underperformance-is-legally-required/

Lewis, M., Dauphin, Y. N., Parikh, D., Batra, D., & Yarats, D. (2017, June 16). *Deal or No Deal? End-to-End Learning for Negotiation Dialogues*. arXiv. https://arxiv.org/pdf/1706.05125

Li, P., Yang, J., Ren, S., & Islam, M. A. (2023, April 6). *[2304.03271] Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models*. arXiv. Retrieved November 6, 2024, from https://arxiv.org/abs/2304.03271

Liminga, A., & Lindgren, S. (2024, September 24). *Mapping the discursive landscape of data activism: Articulations and actors in an emerging movement*. Sage Journals. https://journals.sagepub.com/doi/10.1177/20539517241266416

Livingstone, G. (2024, August 1). *Anger mounts over environmental cost of Google datacentre in Uruguay*. The Guardian. Retrieved November 10, 2024, from https://www.theguardian.com/global-development/article/2024/aug/01/uruguay-anger-environmental-cost-google-datacentre-carbon-emissions-toxic-waste-water

Luccioni, A. S., Viguier, S., & Ligozat, A.-L. (2022). *ESTIMATING THE CARBON FOOTPRINT OF BLOOM, A 176B PARAMETER LANGUAGE MODEL*. arxiv. https://arxiv.org/pdf/2211.02001

Madison, N., & Klang, M. (2019). *Recognizing Everyday Activism: Understanding Resistance to Facial Recognition*. Journal of Resistance Studies. https://resistance-journal.org/jrs_articles/recognizing-everyday-activism-understanding-resistance-to-facial-recognition/?utm

Mann, S. (2024, June 10). *Havering data centre: Plans for east London scheme trigger fury*. BBC. Retrieved November 10, 2024, from https://www.bbc.com/news/articles/c4nn52nvvljo

McCallum, S. (2023, April 1). *ChatGPT banned in Italy over privacy concerns*. BBC. Retrieved December 15, 2024, from https://www.bbc.co.uk/news/technology-65139406

McCarthy, J. (2007, November 12). *What is AI?* Stanford.edu. Retrieved October 20, 2024, from http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955, August 31). *A Proposal for the Dartmouth Summer Research Project on Artificial intelligence*. Stanford.edu. http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

McDonald, H. (2020, August 4). *Home Office to scrap 'racist algorithm' for UK visa applicants*. The Guardian. Retrieved December 31, 2024, from https://www.theguardian.com/uk-news/2020/aug/04/home-office-to-scrap-racist-algorithm-for-uk-visa-applicants

McVeigh, T. (2013, February 23). *Killer robots must be stopped, say campaigners | Robots*. The Guardian. Retrieved November 26, 2024, from https://www.theguardian.com/technology/2013/feb/23/stop-killer-robots

Mello, M. M., & Guha, N. (2024, February). Understanding Liability Risk from Healthcare AI. In *Stanford HAI Policy & Society Policy Brief*. https://hai.stanford.edu/sites/default/files/2024-02/Liability-Risk-Healthcare-AI.pdf

Merchant, B. (2024, July 23). *AI Is Already Taking Jobs in the Video Game Industry*. WIRED. Retrieved April 6, 2025, from https://www.wired.com/story/ai-is-already-taking-jobs-in-the-video-game-industry/

Microsoft Source. (2024, September 17). *BlackRock, Global Infrastructure Partners, Microsoft and MGX launch new AI partnership to invest in data centers and supporting power infrastructure*. Microsoft.com. https://news.microsoft.com/2024/09/17/blackrock-global-infrastructure-partners-microsoft-and-mgx-launch-new-ai-partnership-to-invest-in-data-centers-and-supporting-power-infrastructure/

Miller, G. (2023, December 8). *US Senate AI 'Insight Forum' Tracker*. Tech Policy Press

Milmo, D. (2023, February 3). *ChatGPT reaches 100 million users two months after launch*. The Guardian. Retrieved November 27, 2024, from https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app

Mittelstadt, B. D., Taddeo, M., Allo, P., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716679679

Molnar, P. (2020). *Technological Testing Grounds: Migration Management Experiments and Reflections from the Ground Up*. Sarah Chander, EDRi; Chris Jones, Statewatch; Antonella Napolitano, Privacy International; Kostantinos Kakavoulis, Homo Digitalis. https://edri.org/wp-content/uploads/2020/11/Technological-Testing-Grounds.pdf

Molnar, P. (2024). *The Walls Have Eyes: Surviving Migration in the Age of Artificial Intelligence*. New Press.

Moor, J. (2006). The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. *AI Magazine*, *27*(4), 87-91.

Murray, A. (2024, 11 04). Automated Public Decision Making and the Need for Regulation. *LSE Public Policy Review*, *3*(3). https://ppr.lse.ac.uk/articles/10.31389/lseppr.110

Mytton, D. (2021, February 15). *Data centre water consumption*. Nature.com. https://www.nature.com/articles/s41545-021-00101-w

Nadibaidze, A., Bode, I., & Zhang, Q. (2024, November). *AI in Military Decision Support Systems: A Review of Developments and Debates*. Odense: Center for War Studies. https://usercontent.one/wp/www.autonorms.eu/wp-content/uploads/2024/11/AI-DSS-report-WEB.pdf?media=1629963761

Narayanan, A., & Kapoor, S. (2024). *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*. Princeton University Press.

National Academy of Sciences & National Research Council. (1966). *LANGUAGE AND MACHINES: COMPUTERS IN TRANSLATION AND LINGUISTICS*. https://mt-archive.net/50/ALPAC-1966.pdf

NBC Boston. (2020, June 24). NBC Boston. https://www.nbcboston.com/news/local/boston-approves-ban-on-facial-recognition-technology/2148450/

Newlove-Eriksson, L., Giacomello, G., & Eriksson, J. (2018). The invisible hand? Critical information infrastructures, commercialisation and national security. *The International Spectator*, *53*(2), 124–140.

Newman, M. (2024, January 12). *Thiel's Palantir, Israel Agree Strategic Partnership for Battle Tech* Bloomberg.com. https://www.palantir.com/assets/xrfr7uokpv1b/3MuEeA8MLbLDAyxixTsiIe/9e4a11a7fb058554a8a1e3cd83e31c09/C134184_finaleprint.pdf

Nichols, M. (2023, June 12). *UN chief backs idea of global AI watchdog like nuclear agency*. Reuters. Retrieved December 15, 2024, from https://www.reuters.com/technology/un-chief-backs-idea-global-ai-watchdog-like-nuclear-agency-2023-06-12/

Nolan, B. (2023, March 16). *The latest version of ChatGPT told a TaskRabbit worker it was visually impaired to get help solving a CAPTCHA, OpenAI test shows*.

Business Insider. https://www.businessinsider.com/gpt4-openai-chatgpt-taskrabbit-tricked-solve-captcha-test-2023-3?op=1

Nolan, B. (2024, December 11). *Lawsuit Claims Character.AI Chatbot Suggested a Teen Kill His Parents*. Business Insider. Retrieved December 16, 2024, from https://www.businessinsider.com/characterai-google-lawsuit-chatbot-teen-kill-parents-2024-12?utm

Not Here Not Anywhere. (2023, September). *Data Centres – Not Here Not Anywhere*. Not Here Not Anywhere. Retrieved November 10, 2024, from https://notherenotanywhere.com/campaigns/data-centres/

O’Brien, I. (2024, September 15). *Data center emissions probably 662% higher than big tech claims. Can it keep up the ruse?* The Guardian. https://www.theguardian.com/technology/2024/sep/15/data-center-gas-emissions-tech

O’Brien, M. (2020, June 11). *Microsoft joins Amazon, IBM in pausing face scans for police*. AP. https://apnews.com/article/e5dfcb8c0b003c1134137d33add4c301

O'Donovan, C. (2024, October 5). *These small towns are resisting the spread of energy-hungry data centers*. The Washington Post. Retrieved November 10, 2024, from https://www.washingtonpost.com/technology/2024/10/05/data-center-protest-community-resistance/

OECD. (2024, March 15). *Using AI in the workplace: Opportunities, risks and policy responses*. OECD. https://www.oecd.org/en/publications/using-ai-in-the-workplace_73d417f9-en.html

OECD. (2025). *OECD AI Principles overview*. OECD. https://oecd.ai/en/ai-principles

Office of the President of the Republic. (2025, February 10). *MAKE FRANCE AN AI POWERHOUSE*. elysee.fr. Retrieved April 4, 2025, from https://www.elysee.fr/admin/upload/default/0001/17/d9c1462e7337d353f918aac7d654b896b77c5349.pdf

Olazaran, M. (1996, Aug.). A Sociological Study of the Official History of the Perceptrons Controversy. *Social Studies of Science,*, *Vol. 26,*(No. 3), pp. 611-659.

OpenAI. (2022, November 30). *Introducing ChatGPT*. OpenAI. Retrieved November 27, 2024, from https://openai.com/index/chatgpt/

OpenAI. (2023, March 27). *arXiv:submit/4812508 [cs.CL] 27 Mar 2023*. OpenAI. Retrieved December 15, 2024, from https://cdn.openai.com/papers/gpt-4.pdf

OpenAI. (2023, April 5). *Our approach to AI safety*. OpenAI. Retrieved December 11, 2024, from https://openai.com/index/our-approach-to-ai-safety/

Padden, M. (2023, August 8). *The transformation of surveillance in the digitalisation discourse of the OECD: a brief genealogy*. Internet Policy Review. Retrieved November 16, 2024, from https://policyreview.info/articles/analysis/transformation-of-surveillance-in-digitalisation-discourse

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

Pasquinelli, M. (2023). *The Eye of the Master: A Social History of Artificial Intelligence*. Verso Books.

PauseAI. (2025). *Next International Protest: Paris AI Summit — February 7–11*. https://pauseai.info/2025-february

Pequeño IV, A. (2024, Jul 16). *JD Vance And Peter Thiel: What To Know About The Relationship Between Trump's VP Pick And The Billionaire*. Forbes. https://www.forbes.com/sites/antoniopequenoiv/2024/07/16/jd-vance-and-peter-thiel-what-to-know-about-the-relationship-between-trumps-vp-pick-and-the-billionaire/

Perez, C. (2002). *Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages*. Edward Elgar.

Perez, C. (2024, March 11). *What Is AI's Place in History? by Carlota Perez*. Project Syndicate. Retrieved October 12, 2024, from https://www.project-syndicate.org/magazine/ai-is-part-of-larger-technological-revolution-by-carlota-perez-1-2024-03

Perrigo, B. (2023, January 18). *Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic*. Time. https://time.com/6247678/openai-chatgpt-kenya-workers/

Pierson, B. (2024, 10 24). *Mother sues AI chatbot company Character.AI, Google over son's suicideMother sues AI chatbot company Character.AI, Google over son's suicide*. Reuters. https://www.reuters.com/legal/mother-sues-ai-chatbot-company-characterai-google-sued-over-sons-suicide-2024-10-23/

Pollina, E., & Coulter, M. (2023, February 3). *Italy bans U.S.-based AI chatbot Replika from using personal data*. Reuters. Retrieved May 14, 2025, from https://www.reuters.com/technology/italy-bans-us-based-ai-chatbot-replika-using-personal-data-2023-02-03/?utm

Pomfret, J., & Pang, J. (2024, November 1). *Exclusive: Chinese researchers develop AI model for military use on back of Meta's Llama*. Reuters. Retrieved November 26, 2024, from https://www.reuters.com/technology/artificial-intelligence/chinese-researchers-develop-ai-model-military-use-back-metas-llama-2024-11-01/

Protect Not Surveil. (2024). EU AI | Protect Not Surveil. Retrieved December 31, 2024, from https://protectnotsurveil.eu/#resources

Public Law Project. (2023, February 17). *Legal action launched over sham marriage screening algorithm*. Public Law Project. Retrieved December 31, 2024, from https://publiclawproject.org.uk/latest/legal-action-launched-over-sham-marriage-screening-algorithm/

Quinn, B., & Milmo, D. (2024, November 26). *How the far right is weaponising AI-generated content in Europe*. The Guardian. Retrieved December 10, 2024, from https://www.theguardian.com/technology/2024/nov/26/far-right-weaponising-ai-generated-content-europe

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018, June 11). *Improving Language Understanding by Generative Pre-Training*. OpenAI. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

Rahman, I., & Fabbri, T. (2024, January 31). *ChatGPT: Italy says OpenAI's chatbot breaches data protection rules*. BBC. Retrieved December 15, 2024, from https://www.bbc.co.uk/news/technology-68128396

Reismann, H. (2023, 07 13). *What Is Deepfake Porn and Why Is It Thriving in the Age of AI?* University of Pennsylvania Annenberg School for Communication. https://www.asc.upenn.edu/news-events/news/what-deepfake-porn-and-why-it-thriving-age-ai

Ritchie, H., & Roser, M. (2024). *CO2 emissions*. Our World in Data. https://ourworldindata.org/co2-emissions

Robinson, D. (2024, Sep 6). *Datacenters to emit 3x more carbon dioxide because of generative AI*. The Register. https://www.theregister.com/2024/09/06/datacenters_set_to_emit_3x/

Romine, T. (2022, December 7). *Amid public outcry, San Francisco rejects police use of robots to kill*. CNN. Retrieved December 1, 2024, from https://edition.cnn.com/2022/12/07/us/san-francisco-rejects-police-controlled-robots/index.html

Rone, J. (2024,). The shape of the cloud: Contesting data centre construction in North Holland. *New media & society*, *26*(10), 5999–6018. https://doi.org/10.1177/14614448221145928

Russell, S., Perset, K., & Grobelnik, M. (2023, November 29). *Updates to the OECD's definition of an AI system explained*. OECD AI Policy Observatory. Retrieved December 15, 2024, from https://oecd.ai/en/wonk/ai-system-definition-update

Ryan-Mosley, T. (2023, 06 12). It’s time to talk about the real AI risks. *MIT Technology Review*. https://www.technologyreview.com/2023/06/12/1074449/real-ai-risks/

Saenz, A. D., Harned, Z., Banerjee, O., Abramoff, M. D., & Rajpurkar, P. (2023, 10 06). Autonomous AI systems in the face of liability, regulations and costs. *npj Digital Medicine*, *6*. https://www.nature.com/articles/s41746-023-00929-1

Sallai, D., Cardoso Silva, J., Barreto, M., Panero, F., Berrada, G., & Luxmoore, S. (2024, 11 04). Approach Generative AI Tools Proactively or Risk Bypassing the Learning Process in Higher Education. In *LSE Public Policy Review*. https://ppr.lse.ac.uk/articles/10.31389/lseppr.108

Samhan, B. (2018, January). Revisiting Technology Resistance: Current Insights and Future Directions. *Australasian Journal of Information Systems*, *22*. DOI:10.3127/ajis.v22i0.1655

Samuelson, A. (2024, June 13). *How grassroots climate activists are taking on Big Tech*. HEATED | Emily Atkin. Retrieved November 10, 2024, from https://heated.world/p/how-grassroots-climate-activists

Sánchez-Monedero, J., & Dencik, L. (2022). The politics of deceptive borders: 'biomarkers of deceit' and the case of iBorderCtrl. *Information, Communication & Society*, *25*(3), 413-430. https://doi.org/10.1080/1369118X.2020.1792530

Schaake, M. (2024). *The Tech Coup: How to Save Democracy from Silicon Valley*. Princeton University Press.

Seuferling, P., & Pfeifer, M. (2024, 02 01). *Smart borders and their critiques are too focused on the tech: Why we need a historical approach to envision different futures*. Media@LSE. https://blogs.lse.ac.uk/medialse/2024/02/01/smart-borders-and-their-critiques-are-too-focused-on-the-tech-why-we-need-a-historical-approach-to-envision-different-futures/

Sharkey, N. (2007, August 8). *Robot wars are a reality*. Guardian. https://www.theguardian.com/commentisfree/2007/aug/18/comment.military

Silburn, B. (2001, 11 11). Can computers be creative? *BBC Click Online*.

Şimşek, C. (2017, February). Uzaktan Kumandalı ve Otonom Silah Sistemlerinin Uluslararası İnsancıl Hukuka Etkisi. *Hukuk Kuramı*, *4*(1), 1-25. https://www.hukukkurami.net/media/file/18_19_01_simsek.pdf

Şimşek, C. (2021, May). *Algorithmic Transparency in the EU*. Sciences Po. https://www.sciencespo.fr/public/sites/sciencespo.fr.public/files/SIMSEK%20Can.pdf

Singer, B., Bingham, D. R., Corbett, B., Davenport, C., & Gandolfi, A. (2024, April 28). *GS SUSTAIN Generational Growth AIdata centers' global power surge and the Sustainability impact*. Goldman Sachs. Retrieved November 6, 2024, from https://www.goldmansachs.com/images/migrated/insights/pages/gs-research/gs-sustain-generational-growth-ai-data-centers-global-power-surge-and-the-sustainability-impact/sustain-data-center-redaction.pdf

Smuha, N. A., & Yeung, K. (2025). The European Union’s AI Act: beyond motherhood and apple pie? In *The Cambridge Handbook on the Law, Ethics and Policy of Artificial Intelligence* (Cambridge University Press ed.). Nathalie A. Smuha.

Social Media Victims Law Center. (2025). *Character.AI Lawsuits*. Character.AI Lawsuits. https://socialmediavictims.org/character-ai-lawsuits/

Standage, T. (2002). *The Turk: The Life and Times of the Famous Eighteenth-Century Chess-Playing Machine*. Walker.

Stanford University Human-Centered Artificial Intelligence. (2024). *Artificial Intelligence Index Report 2024*.

Stokel-Walker, C. (2022, 12 09). AI bot ChatGPT writes smart essays — should professors worry? *Nature*. https://www.nature.com/articles/d41586-022-04397-7

Thomas, U. (1971). *Computerised data banks in public administration : trends and policies issues*. OECD.

Thornton, R., & Miron, M. (2020, May 28). Towards the ‘Third Revolution in Military Affairs’: The Russian Military’s Use of AI-Enabled Cyber Warfare. The RUSI Journal,. *165*(3), 12–21. https://www-tandfonline-com.acces-distant.sciencespo.fr/doi/full/10.1080/03071847.2020.1765514

Tong, A. (2023, March 21). *What happens when your AI chatbot stops loving you back?* Reuters. https://www.reuters.com/technology/what-happens-when-your-ai-chatbot-stops-loving-you-back-2023-03-18/?utm

Turing, A. (1950). COMPUTING MACHINERY AND INTELLIGENCE. *M I N D*, *LIX*(236), 433-460. https://academic.oup.com/mind/article/LIX/236/433/986238

Tusseau, G. (2023, June). *Taking Chaos Seriously: From Analogue to Digital Constitutionalism(s)*. SciencesPo. https://www.sciencespo.fr/public/chaire-numerique/wp-content/uploads/2023/11/chaire-digitale-g-tusseau-consitutionalism.pdf

UK CMA. (2024, April 29). *AI strategic update*. Competition and Markets Authority. https://www.gov.uk/government/publications/cma-ai-strategic-update/cma-ai-strategic-update#fnref:7

UNESCO. (2025). *Artificial Intelligence and the Evolution of AI (Model) Capabilities A Conceptual Primer*. UNESCO. Artificial Intelligence and the Evolution of AI (Model) Capabilities A Conceptual Primer

UN Secretary-General & President of the ICRC. (2023, October 5). *Note to Correspondents: Joint call by the United Nations Secretary-General and the President of the International Committee of the Red Cross for States to establish new prohibitions and restrictions on Autonomous Weapon Systems | United Nations ...* the United Nations. Retrieved November 26, 2024, from https://www.un.org/sg/en/content/sg/note-correspondents/2023-10-05/note-correspondents-joint-call-the-united-nations-secretary-general-and-the-president-of-the-international-committee-of-the-red-cross-for-states-establish-new

Urquieta, C., Dib, D., Blanck, N., Kaphle, A., & Dosunmu, D. (2024, May 31). Data centers bring environmental concerns, like excess water use, to Chile. *Rest of World*. https://restofworld.org/2024/data-centers-environmental-issues/

U.S. Department of Defense. (2023, August 10). *DOD Announces Establishment of Generative AI Task Force*. Defense.gov. Retrieved December 16, 2024, from https://www.defense.gov/News/Releases/Release/Article/3489803/dod-announces-establishment-of-generative-ai-task-force/

Valdivia, A. (2024). The supply chain capitalism of AI: a call to (re)think algorithmic harms and resistance through environmental lens. *INFORMATION, COMMUNICATION & SOCIETY*. https://doi.org/10.1080/1369118X.2024.2420021

Van Den Meerssche, D. (2022, February). Virtual Borders: International Law and the Elusive Inequalities of Algorithmic Association. *EJIL*, *33*(1), Pages 171–204. https://academic.oup.com/ejil/article/33/1/171/6583470

Van Sant, S., & Gonzales, R. (2019, May 14). *San Francisco Approves Ban On Government's Use Of Facial Recognition Technology*. https://www.npr.org/2019/05/14/723193785/san-francisco-considers-ban-on-governments-use-of-facial-recognition-technology

Varoufakis, Y. (2023). *Technofeudalism: What Killed Capitalism*. Bodley Head.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., N. Gomez, A., Kaiser, Ł., & Polosukhin, I. (2017, August 2). *Attention Is All You Need*. arXiv. Retrieved November 27, 2024, from https://arxiv.org/pdf/1706.03762

Vergouw, B., Bondt, G., Custers, B., & Nagel, H. (2016). Drone Technology: Types, Payloads, Applications, Frequency Spectrum Issues and Future Developments. In B. Custers (Ed.), *The Future of Drone Use: Opportunities and Threats from Ethical and Legal Perspectives*. T.M.C. Asser Press.

Vincent, J. (2023, February 6). *Getty Images sues AI art generator Stable Diffusion in the US for copyright infringement*. The Verge. Retrieved December 17, 2024, from https://www.theverge.com/2023/2/6/23587393/ai-art-copyright-lawsuit-getty-images-stable-diffusion

Virginia Data Center Reform Coalition – The Piedmont Environmental Council. (2024, January 22). The Piedmont Environmental Council. Retrieved November 10, 2024, from https://www.pecva.org/work/energy-work/data-centers/virginia-data-center-reform-coalition/

Vöpel, H. (2024, 9 9). The AI Revolution: A New Paradigm of Economic Order. *The Economists' Voice*. https://www.degruyter.com/document/doi/10.1515/ev-2024-0054/html

Warren, T., & Peters, J. (2025, April 4). *Microsoft employee disrupts 50th anniversary and calls AI boss 'war profiteer'*. The Verge. https://www.theverge.com/news/643670/microsoft-employee-protest-50th-annivesary-ai

Watercutter, A. (2023, December 25). *The Hollywood Strikes Stopped AI From Taking Your Job. But for How Long?* WIRED. Retrieved December 15, 2024, from https://www.wired.com/story/hollywood-saved-your-job-from-ai-2023-will-it-last/

Webster, R. A. (2025, April 10). *TIGER, the Algorithm Banning Louisiana Prisoners from Parole — ProPublica*. ProPublica. Retrieved May 14, 2025, from https://www.propublica.org/article/tiger-algorithm-louisiana-parole-calvin-alexander

The White House. (2025, February 21). *Defending American Companies and Innovators From Overseas Extortion and Unfair Fines and Penalties*. whitehouse.gov. https://www.whitehouse.gov/presidential-actions/2025/02/defending-american-companies-and-innovators-from-overseas-extortion-and-unfair-fines-and-penalties/

Wintour, P. (2024, June 14). *Pope calls on G7 leaders to ban use of autonomous weapons*. The Guardian. Retrieved November 28, 2024, from https://www.theguardian.com/world/article/2024/jun/14/pope-tells-g7-leaders-ai-can-be-a-both-terrifying-and-fascinating-tool

Woollacott, E. (2024, November 3). *The environmental campaigners fighting against data centres*. BBC. https://www.bbc.com/news/articles/cz0mlrx0jxno

World Economic Forum. (2020, October). *The Future of Jobs Report*. Weforum. https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf

Yasar, A. G., Chong, A., Dong, E., Gilbert, T., Hladikova, S., Mougan, C., Shen, X., Singh, S., Stoica, A.-A., & Thais, S. (2024, 04 23). *Integration of Generative AI in the Digital Markets Act: Contestability and Fairness from a Cross-Disciplinary Perspective*. LSE Legal Studies Working Paper No. 4/2024. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4769439

Young People’s Alliance, Encode, & Tech Justice Law Project. (2025, January 13). *Replika FTC Complaint v2*. Tech Justice Law Project. Retrieved May 14, 2025, from https://techjusticelaw.org/wp-content/uploads/2025/01/Complaint-and-Petition-for-Investigation-Re-Replika.pdf?

Yu, D., Rosenfeld, H., & Gupta, A. (2023, January 16). *The 'AI divide' between the Global North and Global South*. The World Economic Forum. Retrieved December 11, 2024, from https://www.weforum.org/stories/2023/01/davos23-ai-divide-global-north-global-south/

Zhou, E., & Lee, D. (2024, 3 5). Generative artificial intelligence, human creativity, and art. *PNAS Nexus*. https://academic.oup.com/pnasnexus/article/3/3/pgae052/7618478

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books.

About the authors

Can Şimşek LL.M. is a lawyer and policy researcher specialising in the governance of emerging technologies, with a particular focus on artificial intelligence. He is currently a Humboldt Institute for Internet and Society research fellow.

E-mail address: can.simsek@sciencespo.fr

Dr. Ayşe Gizem Yaşar is an assistant professor (education) at LSE Law School and a CREATe Fellow at the University of Glasgow. Her work focuses on innovation, history of technological change, and the regulation of new technologies.

E-mail address: ayse.yasar@sciencespo.fr

This report was commissioned by **Sciences Po's Digital, Governance and Sovereignty Chair**. It was submitted in December 2024 and updated in March 2025. The authors would like to thank the Digital, Governance and Sovereignty Chair and Professor Florence G'sell for their support.

About the Digital, Governance and Sovereignty Chair

Sciences Po's Digital, Governance and Sovereignty Chair's mission is to foster a unique forum bringing together technical companies, academia, policymakers, civil societies stakeholders, public policy incubators as well as digital regulation experts.

Hosted by the **School of Public Affairs,** the Chair adopts a multidisciplinary and holistic approach to research and analyze the economic, legal, social and institutional transformations brought by digital innovation. The Digital, Governance and Sovereignty Chair is chaired by **Florence G'sell,** Professor of Law at the Université de Lorraine, lecturer at the Sciences Po School of Public Affairs, and visiting professor at the Cyber Policy Center of Stanford University.

The Chair's activities are supported by:

sopra steria next

Caisse des Dépôts GROUPE