AI & SOCIETY
https://doi.org/10.1007/s00146-025-02625-y

OPEN FORUM

The politics of artificial intelligence supply chains

James Muldoon[1] · Ana Valdivia[2] · Adam Badger[3]

Received: 27 May 2025 / Accepted: 10 September 2025
© The Author(s) 2025

Abstract
The rising demand for generative artificial intelligence (AI) is fueling the growth of extractive supply chains to build and power the infrastructures this technology demands. However, there is ambiguity within the scholarly literature about what constitutes an AI supply chain. By connecting discussions across disciplinary boundaries, this article proposes a novel theoretical framework to conceptualise the AI supply chain as consisting of four inter-connected spheres of 'AI infrastructure', 'AI preparation', 'AI deployment' and 'e-waste'. It adopts the case study of OpenAI's ChatGPT to map one such example of an AI supply chain. In so doing, it analyses emerging forms of political contestation and resistance, revealing how the development of these supply chains gives rise to political issues of supply chain opacity, an increasing concentration of actors and power and new forms of coalitional politics. The article contributes to our understanding of AI systems through the development of a more holistic approach that examines end-to-end AI production as an iterative process, providing a new perspective on the journey of material flows within these logistical networks.

Keywords Artificial intelligence · Supply chains · AI infrastructure · Data work · Data centres · e-waste

1 Introduction

Accompanying the rapid deployment of artificial intelligence (AI) across many major platforms and systems are widespread concerns of the harms they could produce due to algorithmic bias and questions of fairness and accountability (Bender et al 2021; Buolamwini and Gebru 2018). Less attention has focussed on how these systems are produced and the possible harms resulting from the supply chains that make AI possible (Crawford 2021; Valdivia 2024; Muldoon et al 2024b; Brodie 2020). The emergent literature on AI supply chains has tended to be divided into disciplinary silos that contain a degree of ambiguity between scholars across different disciplines as to what exactly constitutes such a supply chain and how regulators should respond. This means that research on AI supply chains is conceptualised as discrete activities without sustained analysis of the interrelated and often iterative nature of the production of AI systems. Currently, different elements of this process have their own specific academic literatures which are rarely in dialogue: this includes data annotation (Tubaro et al. 2020; Casilli and Posada 2019; Muldoon et al 2024a), AI infrastructure (Crawford 2021; Robbins and Wynsberghe 2022; Dauvergne 2022), data centres (Velkova 2019; Brodie 2023) mineral extraction (Taffel 2015), model development and datasets (Gebru et al 2021; Bender et al 2021), AI vendors, deployment and marketing (Newlands 2021) and e-waste (Gabrys 2018; Wang et al 2024).

To remedy these issues, this article develops a conceptual framework that shows how these different activities constitute global AI supply chains that span across national boundaries, incorporating multiple economic actors interacting within complex systems. We propose an holistic approach that calls for joined-up ways of thinking that allow us to interrogate the interconnectedness of AI supply chains. AI's production, development, operation and use are multifaceted and complex, which requires a multi-disciplinary mode of thinking that encompasses the breadth and diversity of these challenges and perspectives. Furthermore, this

✉ James Muldoon
james.muldoon@essex.ac.uk

Ana Valdivia
ana.valdivia@oii.ox.ac.uk

Adam Badger
Adam.Badger@newcastle.ac.uk

[1] University of Essex, Colchester, United Kingdom

[2] University of Oxford, Oxford, United Kingdom

[3] Newcastle University, Newcastle upon Tyne, United Kingdom

perspective also reveals the possibilities for connections between political struggles over how AI distributes benefits and harms unevenly across the globe. The development of such an approach that examines end-to-end AI production as an iterative process provides a new perspective on the connections between different parts of these networks.

We define AI supply chains as the complex logistical processes that enable networked organisations to transform raw materials into AI systems with the hardware being recycled or discarded as e-waste. This end-to-end approach encompasses mineral extraction, infrastructural resources, chip manufacturing, human labour, energy, data, model development and deployment, and cyclical costs related to the maintenance of infrastructure and hardware. The main contribution of the article is to map and conceptualise the AI supply chain into four interconnected spheres: ‘AI infrastructure’, ‘AI preparation’, ‘AI deployment, and ‘E-waste’, some of which contain their own sub-groupings. We analyse OpenAI’s ChatGPT as a concrete example of how one of these supply chains operates because of its size, prominence in the sector and its relative maturity in comparison to newer ‘start-up’ entries to the market. Its CEO, Sam Altman, is a high-profile figure and is a leading voice in public debates about AI. As a result, there is a larger body of material for investigation relating to ChatGPT than other AI systems. The choice of OpenAI as a case study leads to a certain US-centric nature of this analysis, with the headquarters of OpenAI in addition to other key partners such as NVIDIA, Microsoft and Amazon based in the US. However, when outsourced labour, mineral extraction and e-waste are taken into consideration, this supply chain stretches across the globe. We also note that if, for example, a Chinese lead tech firm were selected, the supply chain would appear very different, with data annotation taking place through practices of ‘inland sourcing’ and with a different geopolitical configuration of key suppliers (Wu et al 2025).

In adopting this framework, the article understands politics not simply as institutional regulation or policy response, but as the contestation and disruption that occurs when established patterns of production, labour, and value in AI supply chains are called into question or contested. Following an agonistic conception of politics, it foregrounds moments where hidden or naturalised practices are made visible, challenged, and resisted by workers, communities, regulators, and transnational coalitions (Honig 1993). These struggles demonstrate how AI supply chains are not neutral or purely technical arrangements, but sites of power where competing claims over resources, labour, and global hegemony unfold. In this sense, the politics of AI supply chains emerges in the tensions, frictions, and interruptions that unsettle the seeming inevitability of technological development, opening space for alternative arrangements to be imagined and demanded. This includes geopolitical issues, such as growing tensions between China, the US and Taiwan over chip manufacturing, environmental concerns over the impact of mineral extraction, water usage and e-waste, in addition to the development of these technologies with a global political economy that incentivises the pursuit of profit over the satisfaction of human need. As a result, the development of these supply chains poses several risks that—if not confronted effectively—could cause serious harms to workers, consumers, nations, and the environment across the planet, and most acutely in the Global South.

The three central political issues that we unpack throughout the paper and then return to in the conclusion are the following. First, global AI supply chains are predominantly opaque. Due to questions of state security and competitive market pressures, many of the connections between the nodes of these networks are difficult to trace with information in the public domain. This raises the urgent need for greater levels of transparency and regulation of these networks. Second, there is an increasing concentration of power within AI supply chains. At certain points in these supply chains, a small number of companies dominate the supply of particular goods and services, such as AI chips and compute capacity. However, these bottlenecks in the supply chain also provide new possibilities for regulation and resistance. Third, the expansion of these supply chains has led to new forms of corporate and state coalitions in support of further developing AI infrastructure. At the same time, we have witnessed the growth of new coalitions of oppositional forces that challenge these supply chains. In this sense, AI could lead to new forms of political contestation that link the struggles of indigenous communities for land and water rights, with those of data workers in the Global South, with writers and artists whose data is used to train models, and protest movements by workers at large tech firms. By analysing one generative AI ecosystem—OpenAI’s ChatGPT—this paper seeks to uncover the new forms of politics that have emerged through the (re)composition of AI supply chains.

2 Understanding global AI supply chains

The production of AI systems relies on an opaque infrastructural network of production facilities, electronics, data, energy, human labour and natural resources that are extracted, manufactured and shipped across the globe. These supply chains rely on hundreds of companies and thousands of separate processes to deliver goods to consumers. Just considering AI hardware alone, Haydn Belfield (2024) describes the supply chain as ‘a worldwide technical and industrial achievement, comparable to massive energy infrastructure, container ships and ports, or the Earth’s web of satellites’. These supply chains are of critical importance to global markets due to the ongoing wave of investment in the AI industry and how tech companies drive gains in the stock market (Valdivia 2024).

Emerging scholarly literature on 'AI supply chains' and 'AI value chains' examines questions of harms and accountability mechanisms related to these chains (Brown 2023; Cobbe, Veale and Singh 2023; Engler and Renda 2022; Küspert, Moës and Dunlop 2023; Widder & Wong 2023; Brodie 2020; Lehuedé 2024). However, these analyses are based on differing conceptions on what is involved in an AI supply chain and the policy implications of these differences. As an indicative sample, Nathan and Widder (2022) refer to 'AI supply chains' primarily as software production; Brown (2023) discusses dataset creation, AI models and AI vendors; Cobbe and colleagues (2023) analyse the development of cloud computer systems including 'AI-as-a-service'; Muldoon and colleagues (2024a) concentrate on data annotation; and Valdivia (2024) adopts a broader view to examine critical minerals, data centres, model development, and e-waste. This illustrates the ambiguity across disciplinary boundaries about what constitutes an AI supply chain and the consequences for the AI industry, consumers and regulators.

This article demonstrates what is distinctive about AI supply chains, by drawing on broader supply chain literatures and highlighting factors specific to AI. It interprets these supply chains as political in nature and key sources of political struggle and power in a modern capitalist society (Cowen 2014; MacDonald 2014; Tsing 2009). Macdonald (2014: 23) refers to global supply chains as 'the whole, spatially dispersed organisational system of functionally interconnected inputs and processes through which production and distribution are coordinated.' AI supply chains are typical of this broad definition in being geographically dispersed, distributed across multiple international actors and connected through heterogeneous networks of production and exchange. This literature also highlights that these networks are neither neutral nor apolitical. Tsing (2009; 2015) has argued that supply chains can become the structures for exploitative dynamics and lead to 'ecological ruinations'. Tsing's work illustrates how the ability to 'generate high profits depend[s] on firms that break not just national laws but also every conceivable humanitarian and environmental standard' (2009: 172). Early studies have found that AI supply chains are not immune to issues of sweatshop labour, exploitation of workers, unequal international divisions of digital labour, environmental degradation, excessive carbon emissions and weak regulatory frameworks (Crawford 2021; Bender et al 2021; Dauvergne 2022).

One characteristic of AI supply chains is that despite their enormous cost and complexity, parts of these supply chains are remarkably concentrated in just a few companies (and at points, within a single company—Sastry et al 2024). Designing AI chips, for example, is a highly skilled and knowledge-intensive process that creates a high barrier to entry for challenger firms without significant capital—thus allowing NVIDIA to dominate the chip market, with an estimated 80–95% market share of the top-end AI chips (IOT Analytics 2025).

NVIDIA's strategic partner in the manufacturing process, TSMC (Taiwan Semiconductor Manufacturing Company) has approximately 90% of the market share of AI chips and relies on just one company to provide it with manufacturing equipment (the Dutch company ASML) (Sastry et al 2024). In addition, two thirds of the world's cloud computing capacity required to train advanced AI models is held by just three leading tech companies: Amazon, Microsoft and Google (Richter 2024). Due to the need for access to this computational infrastructure, enormous amounts of capital and a handful of leading AI specialists, it is only the largest tech companies that also dominate the training of AI models and model deployment to consumers and other businesses. As a result, AI supply chains have high levels of concentration and bottlenecks.

Next, AI supply chain opacity makes them difficult to oversee and regulate due to competitive market pressures and security concerns (Sastry et al 2024). Many organisations within this supply chain consider the identity of their suppliers and the nature of the services procured to be a matter of strategic secrecy, exacerbated by AI's perceived geopolitical importance (Lee and Hawkins 2024). There is also the matter of the complexity of global networks that make it difficult to track every part of a supply chain given any single company can have hundreds of suppliers. Newlands (2021) identifies how AI vendors often perform a certain strategic secrecy in how they present AI products to clients, particularly in attempts to obscure the amount of human labour that is actually required at different stages of the process. An example of this secrecy is Sama, a data annotation and content moderation company, were ultimately traced back to their contract with Meta, causing negative publicity and reputational damage (Perrigo 2023). All of these factors result in a lack of public information about how AI supply changes are managed which makes it difficult for regulators to properly identify potential harms and develop policy approaches to minimise them (Morgan et al 2023).

The rapid growth of AI systems and the development of these supply chains have also given rise to new struggles and the possibility of new forms of coalitional politics. On the one hand, the growth of AI has led to new strategic partnerships of actors seeking to benefit from the development of this technology. The US state has provided generous subsidies to AI chip manufacturers to set up new facilities in the country and has passed laws prohibiting them from selling products to its adversaries (Lee and Hawkins 2024). Other national and regional governments have been eager to demonstrate their willingness to create attractive investment conditions for AI companies and other firms in the supply chain to do business in their jurisdiction. This risks creating negative externalities for workers, the environment, data subjects and nations in the Global South, raising the possibility of new strategic alliances between different protest and oppositional groups and the identification of forms of resistance between disparate actors connected by their participation in these supply chains (Tech Workers Coalitions 2018; Tarnoff 2020; Muldoon et al 2024a, b). Indeed, we follow Bull and Banik (2025: 195) in understanding the 'Global South' not as a single homogenous grouping of countries considered solely as the victims of global capitalism, but rather as 'characterized by internal diversity, evolving roles, and shifting alliances, while remaining anchored in the broader struggles for global justice and systemic transformation that give the concept its political meaning'. In the following, we examine each of these issues throughout AI supply chains and trace the emerging forms of politics that flow from their development.

2.1 AI infrastructure

We use the term AI infrastructure to refer to the layers of the technological stack of AI including the facilities, hardware (specialised GPUs and CPUs) and deployment tools (software) required for the hardware to function. They form the bedrock of AI systems and are a fundamental element of the AI supply chain that facilitate the work that occurs in the other three spheres of AI preparation and deployment, before being repurposed or disposed of as e-waste. This infrastructure is deeply material in nature, a facet that is often neglected in representations of technology that appear to be somehow 'virtual' (see Dourish 2017; Kinsley 2014). The relative sustainability of these supply chains is often sidelined by company reporting but organisational approaches to digital supply chains are growing in academic attention (see Wang and Zhang 2024). This infrastructure requires capital expenditure from tech companies and must be considered from a life cycle perspective in which it is subject to periodic reinvestment and replacement (Crawford 2021). AI infrastructure is specifically designed to perform computationally-intensive tasks with specialised hardware that can handle the complex functions AI workloads demand. Developing AI infrastructure is critical to OpenAI's mission. Their CEO Sam Altman (2024) has tweeted that the world 'needs more AI infrastructure, fab capacity, energy, data centres, etc. than people are currently planning to build' and that '[b]uilding massive scale AI infrastructure, and a resilient supply chain, is crucial to economic competitiveness'. In short, it is clear that AI companies are doing joined-up thinking with regards to their own supply chain, but governments and academics often remain siloed by discipline and/or geography. We outline four important aspects of AI infrastructure and unpack their interdependencies: i) mineral extraction ii) AI chips, iii) software and iv) data centres.

2.1.1 Mineral extraction

Minerals and metals are needed to build AI hardware including tantalum, copper, silver, bismuth, silicon, antimony, tin, gold, nickel, palladium and boron, among others (Euromines 2020). Mineral extraction involves the use and deterioration of land and other natural resources such as water. In many cases, this activity takes place on rural and Indigenous lands. Owen and colleagues (2022) estimate that 70% of deposits containing nickel, zinc, cobalt, or tungsten—key minerals for electronics and AI infrastructure—are located in Indigenous and peasant territories, most of them situated in Latin America, Africa, the Middle East, or Asia–Pacific. Once mined and processed, these materials are sent to fabrication plants where AI chips are produced. This process will eventually lead to the creation of e-waste as detailed later in the article, with AI infrastructure dumped in landfill when infrastructure is turned over.

OpenAI uses state of the art AI chips from NVIDIA. The NVIDIA Conflict Minerals report indicates that the company obtained tantalum, tin and gold from 259 companies across the globe for their AI chips (NVIDIA 2023b). These mining companies are mainly headquartered in the US and China, with the mining and processing of critical minerals stretching across the globe, but mainly taking place in proximity to the mineral extraction sites in South East Asia, Latin America and China (IEA 2023). There have been local instances of resistance by communities affected by these operations, for example activists' struggles in Santiago and the Atacama Desert, Chile against new data centre construction and ongoing water-intensive lithium extraction. Lehuedé (2024) theorises how their practices of resistance could be understood as an 'elemental ethics' in which water figures 'not just as a resource they [local communities] owned but instead as an agent present in their everyday life and as an enabler of their way of being'. This is indicative of a growing awareness among local communities of the material needs of AI systems and the drain they place on natural resources. Many of these resources are required to make advanced AI chips, which are central to the development of this technology and will be examined next.

2.1.2 AI chips

Most companies training AI models rely on advanced AI chips designed and manufactured in a highly concentrated supply chain. The majority of these chips are specialised Graphics Processing Units (GPUs) and very few are produced in fully integrated production facilities.

The growth of AI has dramatically increased the value of dozens of companies involved in chip manufacture, including NVIDIA, which is valued at over $3 trillion. The production of most chips relies on a process of fabless manufacturing whereby the design and marketing of chips are undertaken by a fabless company (eg. NVIDIA), while the actual manufacture of the chips is outsourced to a fabrication company providing specialised manufacturing services (Yeung et al 2023; Sastry et al 2024). For context, AI chips for training AI models at data centres constitute a relatively small but growing portion of the chip market, less than one percent of all high-end chips (Sastry et al 2024).

ChatGPT was initially trained on Microsoft Azure supercomputing infrastructure, which was purpose built for OpenAI and powered by NVIDIA GPUs (Roth 2023). In 2023, research firm Trendforce estimated that ChatGPT required 30,000 NVIDIA GPUs to power its services (Liu 2023). NVIDIA employs other companies for all manufacturing processes, including wafer fabrication, assembly, testing and packaging (NVIDIA 2023a). Its prominent position at this stage of the AI supply chain allows it to outsource the manufacturing of its top-end AI chips to its strategic partner TSMC, which relies on highly specialised chip manufacturing equipment called extreme ultraviolet lithography (EUV) machines that it purchases from the only company in the world that produces them, ASML (Advanced Semiconductor Materials Lithography). These machines imprint patterns onto thin silicon wafers that are combined to create chips. The most advanced chips were for some time only produced in Taiwan, but TSMC has constructed facilities in Arizona with an investment of about $40 billion, which included the US government spending a proposed $6.6 billion in direct funding, up to US$5 billion in loans, and tax credits for up to 25% of the companies' total capital expenditure (TSMC 2024).

Sastry and colleagues (2024) have argued that this bottleneck in the supply chain might constitute an effective point of regulation for downstream AI models. They point to the fact that relative to other components in the AI supply chain, AI chips are more detectable, concentrated and quantifiable, suggesting that governance capacity could be built at this specific point, producing greater regulatory visibility and preventing malicious development and use of AI systems. This is an important intervention and an argument that could be extended to other strategic points in the AI supply chain, as we examine below. This bottleneck has also increased tensions between the US and China as Taiwan has become a central node in the AI supply chain, leading to heightened geopolitical risks of global conflict. Chip manufacturers are also directly connected in the production of software to operate these chips, with NVIDIA producing a whole suite of software required for their chips.

2.1.3 Software

The infrastructural layer of machine learning also consists of a complex ecosystem of dozens of software components necessary for training AI models. This includes software to run GPUs, data storage and management, frameworks and libraries, orchestration, tuning and optimisation and deployment tools, among others. OpenAI has not publicly disclosed the precise software it uses to train and operate ChatGPT, but standard industry practice—alongside analysis of their samples of open-source code—can be assumed for many of its operations. OpenAI (2020) has explicitly acknowledged that it uses the deep learning framework PyTorch to make it easier for its team to share optimised implementations of its models. The majority of OpenAI's research code is written in Python, although some is also in C + +, JavaScript, Ruby and Go code (OpenAI 2016, 2024a). Certain types of software would also be necessary based on the hardware the company relies on. This is a key interdependence between hardware and software: chip manufacturers can lock clients into their own proprietary software packages for operating their chips. According to NVIDIA chief executive Jensen Huang, 'What NVIDIA does for a living is not [just] build the chips, we build an entire supercomputer, from the chip to the system to the interconnects… but very importantly the software' (Bradshaw 2024). NVIDIA has developed CUDA (Compute Unified Device Architecture), an application programming interface that provides clients with direct access to GPUs to run their own programs and execute commands. In response, OpenAI developed Triton, an open-source programming language that seeks to provide an alternative to NVIDIA's monopoly, which has received backing from Meta, Microsoft and Google, all of which rely on NVIDIA chips and have interests in disrupting its monopoly (OpenAI 2021; Bradshaw 2024). This demonstrates the shifting suite of alliances between otherwise competitive tech firms as they move to weaken the monopolistic ties of NVIDIA and support OpenAI in developing tools to weaken these strong dependencies within the stack. Similarly, OpenAI Head of Infrastructure, Christopher Berner has also noted that the company uses Kubernetes 'as a batch scheduling system' for automating deployment and management of containerised applications (Kubernetes 2024). These, and dozens of other software packages, have been primarily developed in the United States and licensed to OpenAI by third-party companies that feed into this AI supply chain.

2.1.4 Data centres

AI models require significant computational resources and storage capacity to operate, which is often provided through cloud computing infrastructure as a distributed service to clients in which computational capacity is rented out as a flexible resource (Millard 2021). Data centres are constructed and maintained using their own complex supply chains that rely on a host of subsidiary companies supplying servers, cooling equipment, batteries, backup generators, and other equipment (Whitehead et al 2015; McKinsey & Company 2023; Valdivia 2024). Julia Velkova (2019) argues we should understand data centres from a perspective of 'impermanence' to highlight the wasteful materialities of their operation and the need for vast resources to build, repair and operate them, linking again to the creation of e-waste. There are now over 1000 hyperscale data centres with the number of data centres operated by large providers doubling over the past four years (Synergy Research Group 2024). Generative AI such as OpenAI's ChatGPT is a key driver for this large expansion of data centres, particularly by Big Tech firms (AWS, Microsoft, Alphabet, Meta). There are also several other key players in this industry such as Digital Realty, Equinix, NTT Global Data Centres, Oracle, and CloudHQ that provide services to the Big Tech companies (Valdivia 2024). This centralised control over AI infrastructure is one of the primary reasons that nearly all of the largest AI startups have formed a strategic partnership with a Big Tech firm that provides investment and access to computational resources (Muldoon et al 2024b). This is another point in the AI supply chain where regulation by governments and resistance by grassroots movements could provide a strategic point of leverage against the oligarchic ownership structure of current AI systems. Where these relationships must be formed by companies, states are provided a window of opportunity to intervene.

The rapid increase in new data centres relies on construction material for buildings such as concrete, steel and aluminium, *and* specialised computing and network equipment that require raw materials such as copper, silicon and lithium for back-up batteries (Swinhoe 2021). Tower and Townsend's (2023) survey of data centre operators revealed that 88% reported that demand for data centre capacity for AI is increasing rapidly. The same survey revealed 94% of respondents reported a shortage of experienced data centre construction teams, highlighting a skills shortage in the sector and a global demand for new centres.

Once constructed, these centres require large amounts of electricity and water to function and skilled technicians to maintain them. A hyperescale data centre can use between 11 and 19 million litres of water each day, with researchers estimating that ChatGPT consumes up to 500 millilitres of water each time it responds to between 5 and 50 prompts (Singh 2023). Similarly, Google's Sustainability report (2023) has revealed the company used 5.6 billion gallons of water in 2022, an increase of 20% over the previous year. Data centres also place significant demands on local electricity grids, with demand continuing to increase. For example, data centres in Ireland are estimated to require approximately 27% of the country's available electricity supply by 2028 (EirGrid 2020). Data centres can place additional demands on existing energy infrastructure, such as in Virginia, for example, when increased demand for electricity justified renewed investment in fossil fuel infrastructure (Atkins 2021).

While OpenAI initially kept the question of where ChatGPT was trained a secret, a Microsoft executive announced that it was trained on an Azure supercomputer Microsoft had built in Iowa (O'Brien and Fingerhut 2023). This training facility was instrumental in developing OpenAI's GPT-4 model (Bach 2023). When the two companies formed a partnership in 2019, the physical data centre infrastructure was already in place and work began to develop a custom computing system to train OpenAI's models. For the ongoing inference required to operate ChatGPT, OpenAI draws on computational resources from across the world. Microsoft has deployed GPUs across multiple regions to facilitate millions of customers interacting with the chatbot and requiring access to small amounts of computer power to fulfil requests in seconds. Crucially, this global deployment is also a vital tool in maintaining operations throughout conflict, natural disasters or other events that threaten to take systems offline.

Big Tech companies have also been found to employ temporary and short-term contract workers in these data centres, creating a two-tier system between full-time employees with greater employment benefits and what is known at Google as TVCs (temps, vendors and contractors). A report by Data Center Dynamics in 2021 found that Google employed an estimated 130,000–150,000 TVCs, which was more than the total number of full-time Google employees (Moss 2021). These workers reported being treated as second class citizens with vastly different rights, expectations and responsibilities than Google's employees. This highlights the potential for transnational solidarity between tech workers involved in various positions of the supply chains including data annotators, data centre workers and machine learning engineers.

The above analysis demonstrates that data centres bring together a broad coalition of forces that could cohere into a broad allegiance for mounting critical resistance strategies. Indigenous activists have opposed the use of their land, local communities are concerned about the use of natural resources, and there are open questions about the extent to which data centres provide decent jobs for workers (Lehuedé 2024). While the construction of data centres can bring together a political coalition in favour of providing tech companies with favourable conditions for investment (Burrel 2020), they can also lead to local governments mobilising alongside citizen groups against the expansion of the data centre industry and for increased democratisation of the decision making procedures for digital infrastructure (Hogan 2015; Rone 2023).

2.2 AI preparation

2.2.1 Dataset production

Generative AI models are trained on large datasets of text, images and video that enable the model to understand statistical patterns in the data and generate outputs in response to prompts. This data and compute power is housed on, and processed through AI infrastructure (sphere 1) and forms a major element of AI development. Once complete, it is deployed for use (sphere 3). When infrastructure reaches end of life, data resulting from AI production needs to be either ported to newly built infrastructure, stored, or deleted. Large increases in the size of datasets is one of the reasons why models such as ChatGPT had a significant increase in their abilities. ChatGPT3 was trained on a series of different datasets including a filtered version of Common Crawl, WebText2, Books1, Books2 and Wikipedia articles, consisting of approximately 570 GB of data (Brown et al. 2020). The origins of Books1 and Books 2 is questionable and this opacity has brought a range of creatives and writers together to challenge OpenAI to become more transparent. The other datasets are more transparent, and contain large amounts of text scraped from the Internet: some from published books and Wikipedia articles, but others from Reddit forums and other online posts (Gebru et al 2021). Biases in training data can cause significant issues for AI models as they can reproduce inaccuracies and harmful stereotypes that reinforce social hierarchies (Bender et al 2021; Noble 2018). We do not know what changes were made to ChatGPT's training data after the release of ChatGPT-4 because OpenAI declined to publish information about how the model was trained for commercial reasons. As such, what limited transparency there was regarding OpenAI's training data has now reduced to nothing.

OpenAI *is* open about sourcing much of its training data from information it describes as 'publicly available on the internet,' but this does not necessarily mean it is free to use or out of copyright. OpenAI has faced dozens of lawsuits for its use of copyrighted material, most prominently from the New York Times and Sarah Silverman (Muldoon 2024). One exception to copyright infringement is what is known as the 'fair use' doctrine, which includes such purposes as criticism, teaching, research or news reporting, among others. OpenAI (2019) has declared that 'we believe that courts would and should rule that training AI systems on copyrighted works constitutes fair use,' a position it—unsurprisingly—shares with nearly every generative AI company. OpenAI is also pursuing other avenues such as finding private collections of information not available to be scraped and even developing a transcription model to transcribe over a million hours of YouTube videos to train GPT-4. In response to legal issues over its training data, OpenAI has been seeking out licencing deals with publishing partners to obtain copies of 'high quality' text data that it can use to train its future models, including deals with Axel Springer, Le Monde, Financial Times, and the Associated Press (Muldoon 2024). Similarly, Microsoft brokered a $10 million deal with Informa (the parent company of Taylor & Francis) to grant them access to—among other things—their academic publication library (Battersby 2024; Jack 2024). The rationale behind these deals is for the company to avoid expensive lawsuits whilst gaining access to new sources of training data. This aspect of generative AI has proven highly contentious and has led to a significant backlash against tech companies seeking to develop generative AI tools. It also gives rise to potential alliances between creative workers whose data is used to train models, workers whose precarious labour is exploited for data annotation, and community groups whose resources are used for data centres (Muldoon et al 2024a, b).

2.2.2 Model design and development

Generative AI models such as ChatGPT are developed through a training process led by machine learning engineers in AI startups and large tech companies. Training foundation models is incredibly expensive, with Sam Altman putting the cost of training ChatGPT at 'more than' $100 million and Anthropic CEO Dario Amodei asserting that models costing more than $1 billion will appear soon (Knight 2023; Wang 2023). Recent computational advances in hardware, the development of transformers and the use of large datasets have led to a rapid increase in the capacity of these general-purpose models. ChatGPT was developed in two phases involving a pre-training phase of unsupervised learning on a large text corpus described above, and a second phase of fine-tuning on specific tasks such as question-answering and dialogue (Radford et al 2019). The initial stages of model development and training are undertaken by highly-specialised machine learning engineers. There are only a handful of people in the world with industry-leading knowledge and expertise in machine learning who can perform this work and competition between the top firms is fierce. Big Tech companies offer generous compensation packages of upwards of $1 million dollars with CEOs personally reaching out to individuals to poach top AI talent (Hays and Thomas 2024). Ownership of cutting-edge hardware exercises a gravitational pull over AI talent since individuals want to work at leading organisations producing the most advanced models (Muldoon et al 2024b). As a result, top graduates tend to work for large US-based tech companies with the resources and incentive packages to retain them (Maslej et al 2024). At OpenAI, their recent growth has resulted in a particularly secretive organisational structure that has emerged partly in response to the scrutiny it has faced following its meteoric rise (French-Owen 2025; Wang and Zhang 2025a). Bort (2025: n.p.) reports that this organisational culture is a young one and that 'OpenAI doesn't seem to know yet that it's a giant company, right down to running entirely on Slack. It feels very much like move-fast-and-break-things Meta in its early Facebook years'.

Attracting and retaining AI talent is an important bottleneck in the AI supply chain. The popularity of OpenAI as a workplace and of the organisational culture that exists there was demonstrated by over seven-hundred staff writing a letter to threaten en-masse resignation if the company's board did not resign and reinstate Altman as CEO after an attempt to oust him (Knight and Levy 2023). The limited number of top AI scientists available is such that it provides these tech workers with a degree of leverage against—even their own—companies that they have used to contest the types of contracts tech companies accept. These have become apparent in moments of geopolitical crisis and subsequent controversies at firms. For instance, Google and Amazon workers part of a collective action group called 'No Tech for Apartheid' have protested these companies' contracts with the Israeli state and their participation in this country's war efforts, including the use of an AI system called 'Lavender' which selects targets for bombing (Davies et al 2023). The potential for action at OpenAI if workers unite to call for change in areas of the supply chain is yet unknown, but the board's reaction to their call to re-instate their CEO is clearly demonstrative of their power in the highest levels of decision making.

The fine-tuning process of model development also takes advantage of a process of supervised learning known as reinforcement learning from human feedback (RLHF), which is thought to have made a decisive improvement to ChatGPT's capabilities (Heikkilä 2023). This process is undertaken with human trainers who rank the responses of the models and create a reward structure that fine-tunes the model through multiple iterations. Some of this work is undertaken by machine learning engineers, but much of it is outsourced to precarious and low-paid data annotators in various locations in the Global South, which we examine below (Muldoon et al 2024b).

2.2.3 AI data work

The AI supply chain contains many 'AI data workers' who perform the behind-the-scenes work of preparing and annotating datasets, providing human reinforcement for models, and verifying the results of AI training programs (Miceli and Posada 2022). Muldoon and colleagues (2024a) define AI data work as 'the human labour required to support machine learning algorithms through the preparation and evaluation of datasets and model outputs that is often outsourced to low-paid and marginalised workers.' Studies have estimated that as much as 80% of project time on AI models is this type of work (Cognilytica Research 2019). According to digital labour researchers Tubaro and Casilli (2019) there are no signs of this labour being automated away as it forms a structural component that supports machine learning algorithms. This type of data work can be performed both by geographically-dispersed independent contractors on digital platforms such as Amazon Mechanical Turk and by employees in Business Process Outsourcing (BPO) centres, which are frequently located in various countries in the Global South on account of lower wages and weaker labour laws (Miceli and Posada 2022). Time magazine reported how OpenAI outsourced this operation through the data annotation firm, Sama. Sama distributed these tasks to Kenyan workers earning less than $2 an hour to label harmful content and help train its models to become less 'toxic' (Perrigo 2023). This takes a psychological toll on these workers and could even cause anxiety for other workers in the supply chain such as engineers at firms like OpenAI, which could be seen as another potential catalyst for resistance (see Zhang and Wang 2025b).

In response, AI data workers have been organising within their companies, forming the African Content Moderators Union and pursuing legal action against data annotation companies (Muldoon et al 2024a, b). However, because many of these workers are on short-term contracts and the data annotation companies do not have the same reputational concerns as large tech companies like Meta, these workers have faced many obstacles in achieving their demands. One additional strategy suggested by the Tech Workers Coalition (2018) is for the further development of solidarity across the entire tech industry, including between what could be considered white and blue collar tech work and including 'all occupations and stratas: everybody from cafeteria workers, to customer service reps, to data scientists'. This is another of the new coalitional possibilities that is created by the ongoing spread of AI and the expansion of the industry and its need for human labour.

2.3 AI deployment

Once the model has been developed and is ready to be deployed we shift from the upstream to the downstream part of the AI supply chain. So far, this article has focused on questions of how AI systems are produced, but for a sense of the end-to-end process, we also provide an overview of how AI is deployed in real-world applications. This element of the supply chain is reliant-upon the spheres that occur before it. Deployment often occurs alongside continued development and refinement (meaning it is a distinct, if not an entirely discrete sphere of the overall supply chain). Without spheres 1 and 2, deployment would be impossible. OpenAI provides access to its model to individuals and businesses through the ChatGPT chatbot, allowing customers to use its capabilities through a monthly subscription model or pay-per-use through an API (OpenAI 2024b). In addition, other developers can fine-tune the model and develop features on top that offer additional value to customers and for which they can charge a premium. For example, Jasper.ai is a writing tool that generates text content, Woebot is a therapy bot, Ada performs customer service tasks, AI Dungeon is a text-based adventure game, and Wealthfront offers AI-powered financial advice. All of these applications use ChatGPT as the basis for their product. On top of this, large tech companies sell 'AI-as-a-service' (AIaaS) via their cloud computing platforms (Tubaro et al. 2020; Newlands 2021). A number of small and mid-sized startups also have AIaaS incorporated into their business model (Metelskaia et al 2018). In this model, third parties can access machine learning tools via the company's cloud computing system to develop their own tools. Further downstream there are vendors that sell AI-powered products that they have bought from other developers but use in their products to sell to their clients (Newlands 2021).

2.4 E-waste

One additional important aspect of the AI supply chain that is often missed is what happens to all of the old hardware and infrastructure when it is no longer functional (Valdivia 2024). This is frequently invisibilised by firms that seek to promote the scale of their infrastructure (sphere 1), the power of their development and training models (sphere 2) and the volume of users on the system (sphere 3). Whilst the previous three spheres of the supply chain can be actively mobilised to investors, governments and users, the fourth sphere of e-waste is an integral, but less marketable element. The computing-intensive nature of generative AI results in greater use of hardware and more need for investment by large tech companies. Microsoft reported a 30% rise in their emissions between 2020 and 2023 largely due to the construction of data centres that power their AI and cloud computing systems (Hodgson 2024). Similarly, Google's emissions climbed nearly 50% in five years due to increased energy demand from AI (Milmo 2024). In 2024, Microsoft, Google, Meta and Amazon increased their profits by almost $10 billion by extending the working life of servers in their data centres from four or five to six years as part of increased spending on cloud computing and their push into generative AI (Kinder et al 2024). All of the physical components of data centres once they reach their end of service date end up in landfill or recycling (Whitehead et al 2015). According to Crownheart (2024) this could total 5 million tonnes of e-waste per year. It is important to note the environmental costs of this final stage of the AI supply chain when accounting for the overall impact of AI systems.

3 Conclusion: AI supply chain politics

From the perspective of this end-to-end analysis of AI development, we conclude with a reflection on the political consequences and possibilities for our three issues of supply chain opacity, a concentration of power and new forms of coalitional politics. First, AI supply chain opacity raises the urgent need for stricter regulatory measures to increase transparency in supply chains and combat harms that occur downstream. Very few countries have adequate legislation that would place duties on lead firms to take responsibility for conditions occurring along their supply chain. One exception to this is the 2023 German Supply Chain Law, which states that German companies with a certain number of employees (now 1,000) must ensure that minimum conditions are met by their suppliers. These include issues such as forced labour, slavery, human rights violations and prohibitions on freedom of association. However, for context, AI firms that outsource various elements of their supply chain do not have large employment figures; whether through the use of Temps, Vendors & Contractors (TVCs) or through running small agile teams. For example, OpenAI, arguably regarded as a global leader in this field despite being a non-profit organisation, only has 1,200 employees (Davalos 2024). Smaller firms would still escape this supply chain transparency threshold. While there are other countries with some laws related to supply chains, these tend to focus exclusively on modern slavery and child labour, without addressing broader concerns about working conditions and environmental standards throughout the supply chain. For example, the Canadian Bill S-211, the UK and Australian Modern Slavery Acts, the French Corporate Duty of Vigilance Law and the Norwegian Transparency Act (Muldoon et al 2024a, b). The German law represents a step forward in placing greater moral and legal responsibility on lead firms in ensuring that potential harms cannot be externalised onto vulnerable subjects at other points in the supply chain. A similar European law, the Directive on Corporate Sustainability Due Diligence, came into force across the EU in July 2024. This law ensures companies identify and seek to remedy any adverse environmental and human rights impacts of their activities or in their supply chains. It remains too early to judge the effects of these laws as they have just come into effect, but they offer a framework and language for greater pressure to be leveraged against AI companies to provide greater transparency to their supply chains and take more responsibility for conditions downstream. As AI systems continue to develop this is likely an area in which further legislative measures will be needed.

A second key political issue with AI supply chains is the concentration of power and market share that create asymmetrical power relations in the supply chain (Valdivia 2024). This adds greater stress and vulnerability to global supply chains as the entire market is reliant on a small number of actors for certain aspects of the service, but it also provides possibilities for governance mechanisms. At the key bottlenecks in AI supply chains: AI chip manufacture, computer provision and model development—there is the possibility of creating a regulatory framework around such ideas as an AI chip registry, required reporting and caps on computational provision, safety requirements for new model development and so on (Sastry et al 2024). At certain points in the AI Supply Chain the concentration of actors makes it easier to establish visibility over transactions and to create enforcement mechanisms to ensure compliance with legislation. In addition to the possibility for more strict regulation, these bottlenecks also provide the potential for bottom-up forms of resistance. If companies are reliant on one particular chip manufacturer or workers at a handful of companies, these workers occupy a strategic position in the production process, which they can use to their advantage in threatening to withhold their labour and draw attention to conditions at other points in the supply chain (Muldoon et al 2024a, b). This is compounded by the highly specific, and limited supply of production machinery that —for now— makes even the largest of transnational firms significantly less mobile. The concentration of power both enhances the bargaining position of the firms in these networks, but it also makes them vulnerable to being targeted by workers movements and strike activity.

Third, the growth of AI creates new coalitions and alliances between disaffected groups at different points in the supply chain. While there are currently only a limited number of examples of such coalitions forming, and of those they tend to be one-sided acts of support rather than joint organising, there is the emergence of new political fault lines as a result of the expansion of the AI industry. One prominent example of transnational worker organising was performed by the Amazon Employees for Climate Justice group, who organised a global walkout of over 3,000 workers in solidarity with the youth-led climate movement (Associated Press 2023). Another example is the coalitions that are forming in support of data annotators based in the Global South, with other workers in the AI supply chains providing donations and statements of solidarity from across the globe (Perrigo 2023). Workers in the Alphabet Workers Union include both highly paid software engineers and less well paid contractors performing administrative and janitorial work. This radical type of organising allows those workers who are in more powerful positions to lend their voices to the more marginalised. Workers at Amazon and Google have also protested these firms' contracts with the Israeli government in response to Israel's genocide in Gaza under the banner of 'No tech for apartheid!'. Sebastián Lehuéde's (2024) concept of an 'elemental ethics' also highlights Indigenous and local resistance to extractive AI infrastructures, specifically through grassroots activism against a Google data centre in Santiago and lithium extraction in Atacama, an action which demonstrates Indigenous communities mobilising against resource-hostile AI developments.

AI's global reach and the rapid expansion of its supply chains will provide ample opportunities in the future for further alliances to be formed at critical junctures in the supply chains. The sheer complexity of the supply chains that comprise the AI ecosystem are bringing organisations, states, and communities into new constellations. Within these supply chains, bottlenecks have the potential to transform into vital chokepoints in the struggle for a fairer, and more just future of AI. Equally though, these bottlenecks come to represent an intense concentration of power in one location. Future research work is needed to investigate these solidarities and bottlenecks further. These supply chains encompass people, places and systems, and have the potential to become a vital battleground in future ecological and geopolitical battles. At present, the envisioning of what the future holds for technology and society is led by corporate entities (Bock-Brown et al 2024). If we are to build a future of AI that is more just, transparent and democratically governed, then multi-stakeholder research and praxis must be part of the future agenda for work. By tracing the iterative and cyclical process of AI development this article has sought to draw attention to many of the costs of this technology and to reveal new forms of politics that are likely to grow in importance with the ongoing rise of AI.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by JM and AV. The first draft of the manuscript was written by JM and AV and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

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

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Altman S (2024) Sam Altman [Twitter] 8 February 2024. https://twitter.com/sama/status/1755294743565930726 (Accessed 23 Jul 2024).

IOT Analytics (2025) The leading generative AI companies. 4 March 2025. https://iot-analytics.com/leading-generative-ai-companies. (Accessed 23 Apr 2025).

Atkins E (2021) Tracing the 'cloud': Emergent political geographies of global data centres. Polit Geogr. https://doi.org/10.1016/j.polgeo.2020.102306

Bach D (2023) How a small city in Iowa became an epicenter for advancing AI. Microsoft News. 8 September 2023. https://news.microsoft.com/source/features/ai/west-des-moines-iowa-ai-supercomputer (Accessed 23 Jul 2024).

Battersby M (2024) Taylor & Francis set to make £58m from AI in 2024 as it reveals second partnership. https://www.thebookseller.com/news/taylor-francis-set-to-make-58m-from-ai-in-2024-as-it-reveals-second-partnership (Accessed 2 Aug 2024).

Belfield H (2024) What role can 'compute' play in AI governance? Leverhulme Centre for the Future of Intelligence. http://lcfi.ac.uk/news-and-events/news/2024/feb/14/compute-ai-governance/ (Accessed 23 Jul 2024).

Bender EM, Gebru T, McMillan-Major A et al. (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada, 3 March 2021, pp. 610–623. ACM. https://doi.org/10.1145/3442188.3445922

Bock-Brown O, Badger A, Adey P (2024) Winging it: visions, automation, and narrating alternative mobility futures. Mobilities 20(2):255–270. https://doi.org/10.1080/17450101.2024.2414785

Bort J (2025) A former OpenAI engineer describes what it's really like to work there. TechCrunch. https://techcrunch.com/2025/07/15/a-former-openai-engineer-describes-what-its-really-like-to-work-there/ [Accessed 4 Sept 2025].

Bradshaw T (2024) Nvidia's rivals take aim at its software dominance. Financial Times. 22 May 2024. https://www.ft.com/content/320f35de-9a6c-4dbf-b42f-9cdaf35e45bb (Accessed 23 Jul 2024).

Brodie P (2020) Climate extraction and supply chains of data. Media Cult Soc 42(7–8):1095–1114

Brodie P (2023) Data infrastructure studies on an unequal planet. Big Data Soc 10(1):20539517231182402

Brown T et al. (2020) Language Models are Few-Shot Learners. Arxiv. https://arxiv.org/abs/2005.14165

Brown I (2023) Expert explainer: Allocating accountability in AI supply chains. Ada Lovelace Institute. 29 June 2023. https://www.adalovelaceinstitute.org/resource/ai-supply-chains/ (Accessed 23 Jul 2024).

Bull B, Banik D (2025) The rebirth of the Global South: geopolitics, imageries and developmental realities. Forum Dev Stud 52(2):195–214. https://doi.org/10.1080/08039410.2025.2490696

Burrel J (2020) On half-built assemblages: waiting for a data center in Prineville, Oregon. Engag Sci Technol Soc 6:283–305

Cobbe J, Veale M and Singh J (2023) Understanding accountability in algorithmic supply chains. FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 1186–1197. https://doi.org/10.1145/3593013.3594073

Cognilytica Research (2019) Data Engineering, Preparation, and Labeling for AI 2019. Available at: https://www.cloudfactory.com/reports/data-engineering-preparation-labeling-for-ai

Cowen D (2014) The Deadly Life of Logistics: Mapping violence in global trade. University of Minnesota Press, Minneapolis

Crawford K (2021) Atlas of AI: power, politics and the planetary costs of artificial intelligence. Yale University Press, Yale and London

Crownhart C (2024) AI will add to the e-waste problem. Here's what we can do about it. MIT Technology Review. https://www.technologyreview.com/2024/10/28/1106316/ai-e-waste/ (Accessed 4 Sept 2025).

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

Davalos J (2024) OpenAI sees tremendous growth in corporate version of ChatGPT. https://www.bloomberg.com/news/articles/2024-04-04/openai-sees-tremendous-growth-in-corporate-version-of-chatgpt (Accessed 2 Aug 2024)

David W (2024) OpenAI transcribed over a million hours of YouTube videos to train GPT-4. The Verge. 6 April 2024. https://www.theverge.com/2024/4/6/24122915/openai-youtube-transcripts-gpt-4-training-data-google (Accessed 23 Jul 2024).

Davies H, McKernan B and Sabbagh D (2023) '"The Gospel": How Israel uses AI to select bombing targets in Gaza. The Guardian, 1 December 2023. https://www.theguardian.com/world/2023/dec/01/the-gospel-how-israel-uses-ai-to-select-bombing-targets (Accessed 23 Jul 2024).

Dourish P (2017) The Stuff of Bits: An Essay on the Materialities of Information. MIT Press, Cambridge, MA

EirGrid (2020) 'All-Island Generation Capacity Statement 2020–2029'. https://www.eirgridgroup.com/site-files/library/EirGrid/All-Island-Generation-Capacity-Statement-2020-2029.pdf (Accessed 23 Jul 2024)

Engler A C and Renda A (2022). Reconciling the AI value chain with the EU's Artificial Intelligence Act. Centre for European Policy Studies. https://www.ceps.eu/cepspublications/reconciling-the-ai-value-chain-with-the-eus-artificialintelligence-act/ (Accessed 23 Jul 2024).

Euromines (2020) The electronics value chain and its raw materials. Retrieved August 4, 2023, from https://euromines.org/files/key_value_chain_electronics_euromines_final.pdf (open in a new window)

Euromines (2022) The Electronics Value Chain and Its Raw Materials. https://euromines.org/files/key_value_chain_electronics_euromines_final.pdf

French-Owen C (2025) *Reflections on OpenAI*. https://calv.info/openai-reflections [Accessed 4 September 2025].

Gabrys J (2018) Digital rubbish: A natural history of electronics. The University of Michigan Press, Ann Arbor, USA

Gebru T, Morgenstern J, Vecchione B et al. (2021) Datasheets for Datasets. arXiv:1803.09010. arXiv. https://doi.org/10.48550/arXiv.1803.09010

Goffman E (1959) The presentation of self in everyday life. Doubleday, New York

Harvey D (2003) The New Imperialism: Accumulation by Dispossession. Oxford University Press, Oxford (UK)

Hays K and Thomas E (2024) Personal CEO calls and $1 million pay packages: How tech companies recruit top AI talent amid a 'gold rush'. *Business Insider*. 26 April 2024. https://www.businessinsider.com/ai-talent-wars-big-tech-ceos-recruiting-paying-huge-packages-2024-4 (Accessed 23 Jul 2024).

Heikkilä M (2023) How OpenAI is trying to make ChatGPT safer and less biased. MIT Technology Review. 21 February 2023. https://www.technologyreview.com/2023/02/21/1068893/how-openai-is-trying-to-make-chatgpt-safer-and-less-biased/ (Accessed 23 Jul 2024).

Helmore E (2024) Nvidia reports stratospheric growth as AI boom shows no sign of stopping. The Guardian. 22 May 2024. https://www.theguardian.com/technology/article/2024/may/22/nvidia-quarterly-earnings (Accessed 5 Jun 2024).

Hodgson C (2024) Microsoft's emissions jump almost 30% as it races to meet AI demand. Financial Times. 15 May 2024. https://

www.ft.com/content/61bd45d9-2c0f-479a-8b24-605d5e72f1ab (Accessed 23 Jul 2024).
Hogan M (2015) Data flows and water woes: the Utah Data Center. Big Data Soc. https://doi.org/10.1177/2053951715592429
Honig B (1993) Political theory and the displacement of politics. Cornell University Press, Ithaca, NY
International Energy Association (2023) Critical Minerals Market Review 2023. https://www.iea.org/reports/critical-minerals-market-review-2023/implications (Accessed 23 Jul 2024).
Jack P (2024) Academic backlash as publisher lets Microsoft train AI on papers. Times Higher Education. https://www.timeshighereducation.com/news/academic-backlash-publisher-lets-microsoft-train-ai-papers (Accessed 2 Aug 2024).
Kinder T, Hodgson C and Criddle C (2024) Big Tech boosts profits by $10bn with accounting change to server life estimate. *Financial Times*. 5 February 2024. https://www.ft.com/content/ad2f407c-633a-431e-874b-44df53acc68a (Accessed 23 Jul 2024).
Kinsley S (2014) 'The matter of 'virtual' geographies'. Prog Hum Geogr 38(3):364–384
Knight W and Levy S (2023) OpenAI staff threaten to quit unless board resigns. WIRED. https://www.wired.com/story/openai-staff-walk-protest-sam-altman/ (Accessed 4 Sept 2025).
Knight W (2023) OpenAI's CEO Says the Age of Giant AI Models Is Already Over. Wired. 17 April 2023. https://www.wired.com/story/openai-ceo-sam-altman-the-age-of-giant-ai-models-is-already-over/ (Accessed 23 Jul 2024).
Kubernetes (2024) Case Study: OpenAI. https://kubernetes.io/case-studies/openai/ (Accessed 23 Jul 2024).
Küspert S, Moës N and Dunlop C (2023) The value chain of general-purpose AI. Ada Lovelace Institute Blog. 10 February 2023. https://www.adalovelaceinstitute.org/blog/value-chaingeneral-purpose-ai/ (Accessed 23 Jul 2024).
Kwet M (2019) Digital colonialism: US empire and the new imperialism in the Global South. Race Class 60(4):3–26
Lee JL and Hawkins M (2024) AMD, like Nvidia, tried to make a weaker chip for the Chinese market—but Washington says it's still too powerful. Futurism. 5 March 2024. https://fortune.com/asia/2024/03/05/amd-tried-to-make-weaker-china-chip-biden-export-controls-china-semiconductors-blocked/ (Accessed 23 Jul 2024).
Lehuedé S (2024) An elemental ethics for artificial intelligence: water as resistance within AI's value chain. AI & Soc. https://doi.org/10.1007/s00146-024-01922-2
Liu Z (2023) ChatGPT Will Command More Than 30,000 Nvidia GPUs: Report. Tom's Hardware. 1 March 2023. https://www.tomshardware.com/news/chatgpt-nvidia-30000-gpus (Accessed 23 Jul 2024).
Macdonald K (2014) The politics of global supply chains: Power and governance beyond the State. Malden, MA: Polity.
Maslej N, Fattorini L, Perrault R, Parli V, Reuel A, Brynjolfsson E, Etchemendy J, Ligett K, Lyons T, Manyika J, Niebles JC, Shoham Y, Wald R and Clark J (2024) The AI Index 2024 Annual Report. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024. https://aiindex.stanford.edu/report/ (Accessed 23 Jul 2024).
McKinsey & Company (2023) Investing in the rising data center economy. McKinsey. 17 January 2023. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/investing-in-the-rising-data-center-economy.
Metelskaia I, Ignatyeva O, Denef S, et al (2018) A business model template for AI solutions. In: Proceedings of the international conference on intelligent science and technology (ICIST '18). New York, NY: Association for Computing Machinery, 35–41.
Miceli M, Posada J (2022) The data-production dispositif. In Proceed ACM Human-Comput Interaction 6(CSCW2):1–37
Millard C (2021) Cloud Computing Law. Oxford University Press
Milmo D (2024) Google's emissions climb nearly 50% in five years due to AI energy demand. The Guardian. 2 July 2024. https://www.theguardian.com/technology/article/2024/jul/02/google-ai-emissions (Accessed 23 Jul 2024).
Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, Gebru T et al (2019) Model cards for model reporting. FAT *19: Proceedings of the Conference on Fairness, Accountability, and Transparency. January 2019, 220–229. https://doi.org/10.1145/3287560.3287596 (Accessed 23 Jul 2024).
Morgan TR, Gabler CB, Manhart PS (2023) Supply chain transparency: theoretical perspectives for future research. Int J Logist Manag 34(5):1422–1445
Moss S (2021) 'Underpaid and overworked: Behind the scenes with Google's data center contractors.' Data Center Dynamics. 2 December 2021. https://www.datacenterdynamics.com/en/analysis/underpaid-and-overworked-behind-the-scenes-with-googles-data-center-contractors (Accessed 23 Jul 2024).
Muldoon J, Cant C, Wu B, Graham M (2024a) A typology of artificial intelligence data work. Big Data Soc. https://doi.org/10.1177/20539517241232632
Muldoon J, Graham M, Cant C (2024b) Feeding the Machine: the hidden human labour powering AI. Canongate, London
Muldoon J (2024) Generative AI and Copyright Law: A Primer. Substack. 14 April 2024. https://doesnotcomputeai.substack.com/p/generative-ai-and-copyright-law-a (Accessed 23 Jul 2024).
Newlands G (2021) Lifting the curtain: strategic visibility of human labour in AI-as-a-service. Big Data Soc. https://doi.org/10.1177/20539517211016026
Nvidia Newsroom (2024) NVIDIA Blackwell Platform Arrives to Power a New Era of Computing. Nvidia. 18 March 2024. https://nvidianews.nvidia.com/news/nvidia-blackwell-platform-arrives-to-power-a-new-era-of-computing (Accessed 23 Jul 2024).
Noble S (2018) Algorithms of oppression: how search engines reinforce racism. NYU Press, New York
Nvidia (2023a) Nvidia Corporation Annual Review. https://s201.q4cdn.com/141608511/files/doc_financials/2023/ar/2023-Annual-Report-1.pdf (Accessed 23 Jul 2024).
Nvidia (2023b) Conflict Minerals Report of NVIDIA Corporation. https://fintel.io/doc/sec-nvidia-corp-1045810-ex101-2023-april-26-19473-799.
O'Brien M and Fingerhut H (2023) Artificial intelligence technology behind ChatGPT was built in Iowa — with a lot of water. *Associated Press*. 9 September 2023. https://apnews.com/article/chatgpt-gpt4-iowa-ai-water-consumption-microsoft-f551fde98083d17a7e8d904f8be822c4 (Accessed 23 Jul 2024).
OpenAI (2016) Infrastructure for Deep Learning. OpenAI. 29 August 2016. https://openai.com/index/infrastructure-for-deep-learning/ (Accessed 23 Jul 2024).
OpenAI (2019) Comment Regarding Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation. https://www.uspto.gov/sites/default/files/documents/OpenAI_RFC-84-FR-58141.pdf (Accessed 23 Jul 2024).
OpenAI (2020) OpenAI standardizes on PyTorch. OpenAI. 20 March 2020. https://openai.com/index/openai-pytorch/ (Accessed 23 Jul 2024).
OpenAI (2021) Introducing Triton: Open-source GPU programming for neural networks. OpenAI. 28 July 2021. https://openai.com/index/triton/ (Accessed 23 Jul 2024).
OpenAI (2024a) OpenAI Repositories. *GitHub*. https://github.com/orgs/openai/repositories (accessed 23 July 2024).
OpenAI (2024b) Pricing. https://openai.com/api/pricing/ (Accessed 23 Jul 2024).
Perrigo B (2023) Exclusive: OpenAI USed Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. https://time.com/6247678/openai-chatgpt-kenya-workers/ (Accessed 23 Jul 2024).

Associated Press (2023) Amazon workers upset over layoffs and return-to-office mandate plan walkout. https://apnews.com/article/amazon-walkout-mandate-climate-change-seattle-dc921dce60c90b8b9d1c48ad8dc49c48 (Accessed 23 Jul 2024).

Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I (2019) Language Models Are Unsupervised Multitask Learners. OpenAI Blog. https://life-extension.github.io/2020/05/27/GPT%E6%8A%80%E6%9C%AF%E5%88%9D%E6%8E%A2/language-models.pdf (Accessed 23 Jul 2024).

Richter F (2024) Amazon Maintains Cloud Lead as Microsoft Edges Closer. Statista. 2 May 2024. https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/ (Accessed 23 Jul 2024).

Rone J (2023) The shape of the cloud: Contesting data centre construction in North Holland. New Media & Society.

Roth E (2023) Microsoft spent hundreds of millions of dollars on a ChatGPT supercomputer. The Verge 13 March 2023. https://www.theverge.com/2023/3/13/23637675/microsoft-chatgpt-bing-millions-dollars-supercomputer-openai (Accessed 23 Jul 2024).

Rowberry S (2025) The value of books in the age of generative AI training data. Convergence. https://doi.org/10.1177/13548565251358020

Sastry G et al (2024) Computing Power and the Governance of Artificial Intelligence. Arxiv. https://www.cser.ac.uk/media/uploads/files/Computing-Power-and-the-Governance-of-AI.pdf (Accessed 23 Jul 2024).

Singh P (2023) Every time you talk to ChatGPT it drinks 500ml of water; here's why. Business Today. 12 September 2023. https://www.businesstoday.in/technology/news/story/microsofts-water-usage-surges-by-thousands-of-gallons-after-the-launch-of-chatgpt-study-397951-2023-09-11 (Accessed 23 Jul 2024).

Google Sustainability (2023) '2023 Environmental Report'. https://sustainability.google/reports/google-2023-environmental-report/.

Swinhow D (2021) Sustainable data centers require sustainable construction. Data Center Dynamics. 21 May 2021. https://www.datacenterdynamics.com/en/analysis/sustainable-data-centers-require-sustainable-construction/ (Accessed 23 Jul 2024).

Synergy Research Group (2024) Hyperscale Data Centers Hit the Thousand Mark; Total Capacity is Doubling Every Four Years. Synergy Research Group. 17 April 2024. https://www.srgresearch.com/articles/hyperscale-data-centers-hit-the-thousand-mark-total-capacity-is-doubling-every-four-years (Accessed 23 Jul 2024).

Taffel S (2015) Towards an Ethical Electronics? Ecol Congol Conf Min Westminster Papers Commun Cult 10(1):18–33

Tarnoff B (2020) The making of the tech workers movement. Logic(s), New York

Tech Workers Coalition (2018) Tech Workers, Platform Workers, and Workers' Inquiry. *Notes From Below*. 30 March 2018, notesfrombelow.org/article/tech-workers-platform-workers-and-workers-inquiry (Accessed 23 Jul 2024).

Tong A, Max A, Cherney CB and Nellis S (2024) Exclusive: ChatGPT-owner OpenAI is exploring making its own AI chips. *Reuters*. 6 October 2023. https://www.reuters.com/technology/chatgpt-owner-openai-is-exploring-making-its-own-ai-chips-sources-2023-10-06/ (Accessed 23 Jul 2024).

Tower and Townsend (2023) Data centre cost index 2023. https://reports.turnerandtownsend.com/dcci-2023/ (Accessed 23 Jul 2024).

Tsing AL (2009) Supply chains and the human condition. Rethinking Marx 21(2):148–176

TSMC (2024) News Archives. https://pr.tsmc.com/english/news/3122 (Accessed 23 Jul 2024).

Tubaro P, Casilli AA (2019) Micro-work, artificial intelligence and the automotive industry. J Ind Bus Econ 46(3):333–345

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

Valdivia A (2024) The supply chain capitalism of AI: a call to (re)think algorithmic harms and resistance through an environmental lens. Information, Communication & Society (online first).

Velkova J (2019) Data Centres as Impermanent Infrastructures. Culture Machine 18. http://culturemachine.net/vol-18-the-nature-of-data-centers/data-centers-as-impermanent/ (Accessed 23 Jul 2024).

Wang P, Zhang L-Y, Tzachor A, Masanet E and Chen W (2024) E-waste Challenges of Generative Artificial Intelligence. Preprint. https://doi.org/10.21203/rs.3.rs-3978528/v1 (Accessed 23 Jul 2024).

Wang S, Zhang H (2024) Inter-organizational cooperation in digital green supply chains: a catalyst for eco-innovations and sustainable business practices. J Clean Prod. https://doi.org/10.1016/j.jclepro.2024.143383

Wang S, Zhang H (2025a) Digital transformation and innovation performance in small-and medium-sized enterprises: a systems perspective on the interplay of digital adoption, digital drive, and digital culture. Systems 13(1):43

Wang S, Zhang H (2025b) Artificial intelligence digital employees and sustainable innovation in online retail: the mediating role of ambidextrous green innovation and the moderating role of ethical anxiety. J Retail Consum Serv. https://doi.org/10.1016/j.jretconser.2025.104235

Wang S (2023) [Twitter]. 23 August 2023. https://twitter.com/sarahdingwang/status/1694415161837453534 (Accessed 23 Jul 2024).

Ward I and Lung N (2023) Big Tech's Year of Partnering Up With AI Startups. Bloomberg. 18 December 2023. https://www.bloomberg.com/news/articles/2023-12-18/big-tech-s-year-of-partnering-up-with-ai-startups (Accessed 23 Jul 2024).

Warren T (2023) Microsoft is finally making custom chips — and they're all about AI. The Verge. 15 November 2023. https://www.theverge.com/2023/11/15/23960345/microsoft-cpu-gpu-ai-chips-azure-maia-cobalt-specifications-cloud-infrastructure (Accessed 23 Jul 2024).

Whitehead B, Andrews D, Shah A (2015) The life cycle assessment of a UK data centre. Int J Life Cycle Assess 20:332–349

Widder DG and Wong R (2023) Thinking upstream: Ethics and policy opportunities in AI supply chains. https://arxiv.org/pdf/2303.07529.pdf.

Widder DG, Nafus D (2023) Dislocated accountabilities in the 'AI supply chain': Modularity and developers' notions of responsibility. Big Data Soc. https://doi.org/10.1177/20539517231177620

Wu T, Muldoon J, Xia B (2025) Global data empires: analysing artificial intelligence data annotation in China and the USA. Big Data Soc. https://doi.org/10.1177/20539517251340600

Yeung HW, Huang S and Xing Y (2023) From fabless to fabs everywhere? Semiconductor global value chains in transition. In: Global Value Chain Development Report 2023: Resilient and Sustainable GVCs in Turbulent Times (pp. 132–187).

Yeung HW (2022) Interconnected worlds: Global electronics and production networks in East Asia. Innovation and Technology in the World Economy Series, Stanford, US: Stanford University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.