FABIAN FERRARI

The age of digital democracy: A progressive vision for generative AI

A spectre is haunting Europe – the spectre of generative AI. This transformative technology, exemplified by ChatGPT, permeates our societies. It brings not only the convenience of use and innovation but also challenges for progressive politics. We must approach this spectre with a nuanced perspective, acknowledging the benefits of technological progress, while also ensuring it serves the public good. However, warnings of AI-powered misinformation are not only overblown, but they also distract from addressing underlying structural problems in the AI industry, including the formation of digital monopolies. This chapter outlines a new vision for AI governance that expands the political horizon beyond a narrow focus on regulation towards a more ambitious project of producing AI infrastructure as a public utility.

The spectre of generative AI

The fear that generative AI systems could manipulate elections and distort the fabric of our democratic processes haunts the EU. For example, in October 2023, the European Union Agency for Cybersecurity issued an alarming warning that the upcoming EU elections are at risk due to AI.[1] As the agency put it,[2] "we need to be alerted to the potential misuse of AI-powered chatbots in phishing attempts, information manipulation and cybercrime". This warning exemplifies a widespread perception of generative AI within the public debates: its role in eroding trust in democratic elections by facilitating the generation of misinformation at scale. It has never been easier to produce compelling pictures and text using apps like OpenAI's DALL-E and ChatGPT. Although deep fakes – using face-swapping techniques to modify videos – have been around for a while, the manipulative power of today's generative AI systems is unprecedented. Thus, as proponents of the misinformation hypothesis argue, we need to regulate the use of generative AI systems to mitigate those risks.

1   Bertuzzi, L. (2023) "EU cybersecurity body sounds alarm bell over AI-driven disruptions of European elections". Euractiv, 19 October.
2   "EU elections at risk with rise of AI-enabled information manipulation". ENISA, 19 October 2023.

However, this surface-level spotlight on the publicly visible expressions of AI-powered misinformation distracts from addressing structural problems in the AI industry. This chapter argues that progressives need to move beyond an exclusive focus on the most visible symptoms of current generative AI developments (e.g. fake news and information manipulation) to consider the underlying infrastructural causes for those symptoms. The key problem of the AI industry is not that it equips 'bad actors' (a term that often remains undefined) with new tools to produce misinformation. Its key problem is rather that a handful of Big Tech companies dominate this industry, evading and undercutting democratic control and accountability. As long as this situation persists, regulatory efforts to merely tamper with publicly identifiable symptoms will fall short of ensuring long-term democratic governance in the digital world.

This chapter proposes a new progressive perspective on AI governance, broadening the scope of policy interventions beyond a limited focus on regulating misinformation. I advocate for a more ambitious endeavour: producing AI infrastructure that underpins consumer-facing applications as a public utility. In other words, democratising the means of AI production. The age of digital democracy does not mean that democratic values and procedures need to be aligned with AI. Rather, it means that those values and procedures need to be imposed upon AI systems and their providers. The future trajectories of AI need to be aligned with democratic values – not the other way around.

Generative AI and Big Tech companies

Why is there a need to expand the EU's policy horizon beyond a narrow focus on regulating the most visible symptoms of generative AI systems? To answer this question, it is worth taking a step back to consider the fact that generative AI systems are constituted by three key components: consumer-facing applications; underlying foundation models; and computational infrastructure. Generative AI systems are powered by machine-learning techniques that can detect statistical patterns in training datasets (e.g. words, pixels) to produce outputs with some variations based on those patterns. Regardless of the type of training data (be it Hemingway novels or Shakespeare plays), the same basic logic applies. Those machine-learning techniques are commonly referred to as 'foundation models'[3] because they can be applied to a range of tasks, from churning out AI-generated scripts for new plays in the style of Shakespeare to producing grocery shopping lists in the style of Hemingway. Foundation models underpin consumer-facing generative AI applications, such as ChatGPT. Importantly, ChatGPT is not a foundation model; it is an application built on top of a foundation model – in this case, OpenAI's proprietary GPT-4 model.

This distinction between consumer-facing applications and foundation models is at the core of understanding uneven power relationships in the AI industry. For example, in the case of OpenAI, the same company owns the consumer applications and foundation

---

3   Ferrari, F., J. van Dijck and A. van den Bosch (2023) "Foundation models and the privatization of public knowledge". *Nature Machine Intelligence*, 5: 818-820.

models. However, in other cases, smaller companies pay a fee to use OpenAI's GPT-4 model when building their own services and applications on the basis of the foundation model. OpenAI can act as a gatekeeper in controlling the downstream use of their models by other companies. Exempting the providers of such foundation models from the EU AI Act,[4] as representatives from Germany and France recently suggested, would put a high burden for compliance on smaller companies that use the foundation models, while their owners could evade responsibility. Crucially, there are only a handful of high-profile foundation models that underpin a range of generative AI start-ups in the EU, including Google's PaLM, Anthropic's Claude, OpenAI's GPT-4 and Meta's LLaMA models. Those foundation model providers either have exclusive partnerships with Big Tech companies (OpenAI/Microsoft, Anthropic/Amazon) or they represent direct subsidiaries of Big Tech companies (Google DeepMind, for example). Although their providers often tout the 'democratising' potential of AI, the political-economic reality is that a handful of platform monopolies[5] dominate this industry. As infrastructure providers like Microsoft and Amazon benefit from the widespread adoption of generative AI systems, they have no intrinsic economic interest in preventing the misuse of those applications by 'bad actors'. Therefore, any regulatory attempt to tamper with symptoms like fake news while neglecting to address uneven power relationships in the industry may only offer a superficial solution.

Beyond consumer-facing AI applications and foundation models, the third component of generative AI systems is computational infrastructure. This component refers to data centres, specialised chips to train large-scale machine-learning systems and the provision of processing power to ensure the day-to-day operations of systems such as ChatGPT. It requires a staggering amount of computing power to develop state-of-the-art generative AI systems in the first place. But their everyday operations also require infrastructural processes. Whenever we generate a text output using ChatGPT, a computing process is triggered in Microsoft's data centres worldwide. Although it is difficult to quantify the environmental impacts[6] of training generative AI and using them on an everyday basis, estimates suggest that a normal ChatGPT conversation of 20-50 question-and-answer pairs consumes approximately 500 ml clean freshwater to cool Microsoft's data centres. In other words, while the outputs of generative AI systems may feel artificial, the actual computing processes that underlie them are far from artificial, necessitating the extraction of material resources that are limited by planetary boundaries.[7]

This third component, computational infrastructure, is crucial for democratising the means of AI production. A progressive policy perspective on generative AI requires moving beyond the publicly visible and commonly discussed ramifications of this technology (misinformation, for example) to dig deeper and ask more fundamental and

---

4   Bertuzzi, L. (2023) "EU AI Act 'cannot turn away from foundation models', Spain's state secretary says". Euractiv, 17 November.
5   Srnicek (2020) "Data, compute, labour". Ada Lovelace Institute, 30 June.
6   Mann, T. (2023) "To quench AI's thirst, the way we build, operate datacenters needs to change". The Register, 15 May.
7   Agar, N., D. Blaustein-Rejto, M. Gomera et al. (2023) «Is AI a climate game-changer?" Project Syndicate, 11 September.

structural questions about the infrastructural underpinnings of this technology. Should computational resources be in the hands of a few powerful companies based in the US, allowing them to define what 'AI' should mean for the rest of us? Can there ever be a democratic governance of generative AI applications without democratising the infrastructure that makes them possible in the first place? And, perhaps most importantly, what can be done to stop a further concentration of economic and cultural power in the AI industry?

Public utility thinking in the age of generative AI

A promising answer to all those questions lies in the idea of transforming AI infrastructure into a 'public utility'. As Viktor Pickard explains, "public utilities are institutions that provide essential services and goods to the public. Different varieties are possible: they may be publicly or privately owned, cooperatively governed, locally controlled at the municipal level or maintained as a state monopoly".[8] For example, electricity, water and sewage systems, transportation and telecommunications are considered public utilities, as they provide essential services that are crucial to the well-being and functioning of society, and their accessibility and reliability are critical for public welfare and economic activities. While we cannot apply those sectors and examples one-to-one to AI infrastructure as a public utility, three important aspects of public utility governance can serve as reference points for progressive thinking and democratic debates about this pivotal topic.

Firstly, it is crucial to consider the dimension of high fixed costs and network effects. In simple terms, this means that it costs a staggering amount of money to assemble a worldwide network of data centres and computational resources. Amazon, Google and Microsoft – three Big Tech companies that are dominant infrastructure providers for generative AI systems – all benefit from a first-mover advantage: because they were very early in recognising the potential of renting out access to computational resources (a business model that is called infrastructure-as-a-service), they were able to optimise their services over the years. Additionally, they can use network effects: the more people use a service, the more data they generate that can be used to improve Big Tech's offerings. Because of a combination of high fixed costs and network effects, it is enormously difficult for EU-based companies to compete with Big Tech firms as infrastructure providers. As ambitious projects such as Gaia-X (an EU alternative to Big Tech's services) illustrate, Big Tech's dominance in the cloud computing industry is deeply entrenched.[9] In short, one key reason for public utility regulation is the impossibility of a level playing field for fair competition in digital markets. The only way to retain a level of European digital sovereignty is to gain sovereignty over the means of AI production.[10]

8   Pickard, V. (2022) "Democratizing the platforms: Promises and perils of public utility regulation". WACC, 19 August, p. 2.
9   Ambasna-Jones, M. (2023) "Is Gaia-X on course to challenge the big tech platforms?". Raconteur, 20 March.
10 Larsen, B. C. (2022) "The geopolitics of AI and the rise of digital sovereignty". Brookings, 8 December.

Secondly, an infrastructural view of the utility providers that underpin the expansion of other services that are built on top of their offerings can have sweeping cultural implications. As legal scholar K. Sabeel Rahman argues, applying public utility governance to Big Tech companies may "alter the very business model and market dynamics of the firms in question to head off potential downstream conflicts, power disparities, and likelihood of exploitation".[11] This reasoning assumes the root cause of all concerns surrounding the role of generative AI in undermining democratic values "lies in the way these platforms operate as modern economic and social infrastructure". The fact that a tiny handful of Big Tech companies control the provision of computational infrastructure for generative AI systems derives from the substantial fixed costs associated with constructing such infrastructural arrangements.

As a result of their business models, their dominance creates distinct power disparities, including the exploitation of low-paid, outsourced workers that annotate training datasets for AI,[12] as well as the disregard for the accumulated cognitive work of copyright holders.[13] By contrast, turning AI infrastructure into a public utility could go hand in hand with requirements for downstream users to ensure fundamental human rights and fair labour standards.

Thirdly, contrary to the perception that state intervention is the enemy of innovation, the production of AI infrastructure as a public utility may ensure the development of technological innovations in line with the public interest of EU citizens – rather than the private interests of American Big Tech firms. As the economist Mariana Mazzucato argues in her influential book, *The Entrepreneurial State*, corporately claimed innovations are often the result of state-funded investments in research and education. From early public investments in internet infrastructure and fundamental research and open datasets, the state is an enabler of innovation. However, a typical pattern in the AI industry is that the fruits of publicly funded work get turned into closed and commercial systems like ChatGPT. State investments made ChatGPT possible, be it in the sense of training datasets or processing power, but states do not benefit from this privatisation, especially in the EU. In the context of AI governance, Mazzucato and her colleagues therefore argue that a focus on governing the symptoms of AI "without improving the kind of institutional and infrastructural environments which avoid lock-in and path dependencies can lead to under-performing innovation systems".[14] The unfettered power of Big Tech companies also poses more structural concerns for democracies. The Open Markets Institute argues there is a "major threat to economic and societal resilience posed by the reliance of our governments and key industries on a handful of geographically concentrated cloud providers".[15] This

---

11 Rahman, K. S. (2018) "Regulating informational infrastructure: Internet platforms as the new public utilities". *Georgetown Law Technology Review*, 2(2): 234-251.
12 Perrigo, B. (2023) "Exclusive: OpenAI used Kenyan workers on less than $2 per hour to make ChatGPT less toxic". *Time*, 18 January.
13 Appel, G., J. Neelbauer and D. A. Schweidel (2023) "Generative AI has an intellectual property problem". *Harvard Business Review*, 7 April.
14 Mazzucato, M., M. Schaake, S. Krier et al. (2022) "Governing artificial intelligence in the public interest". Working paper, WP 2022/12, p. 16. UCL Institute for Innovation and Public Purpose.
15 Lynn, B., M. von Thun and K. Montoya (2023) "AI in the public interest: Confronting the monopoly threat". Open Markets Institute, November.

means that, if Big Tech's services are disrupted, so will public and private sector data flows that rely on those infrastructural offerings for their day-to-day operations.

These three drivers of public utility thinking in the context of AI infrastructure – stifled competition, downstream effects and state-funded innovation – require further elaboration and democratic debate. All three aspects also need to be systematically juxtaposed with the EU's regulatory action in shaping the digital single market in the form of the Digital Markets Act, the Digital Services Act and the AI Act.

Democratising the means of AI production

Fake news, misinformation, deep fakes – those buzzwords distract our attention from the more pressing political project of democratising the means of AI production. It is not enough to look at the symptoms of the AI industry's structural problem. Instead, progressive thinking needs to tackle their underlying driving force: a concentration of infrastructural power in the hands of a few companies. Historically, such tendencies of industry concentration have reliably triggered social, economic and regulatory transformations. The rise of the robber barons during the late 19th century in the US, characterised by the consolidation of power by a handful of industrial magnates, prompted comprehensive reforms in antitrust laws and sweeping regulations to curb monopolistic practices. Similarly, the formation of oil and gas monopolies led to widespread concerns about the social and environmental implications of market power abuse.

What will future historians say about the dominant providers of AI infrastructure? The answer to this critical question will hinge on the dialectical interplay between the control over the means of AI production and the socio-economic structures it perpetuates. Just as industrial monopolies spurred a re-evaluation of capitalist structures in the past, the current state of AI is a historical opportunity. The choices we make now will determine whether the age of digital democracy will amplify concentrated power or empower us collectively.