AI interregnum

Nick Srnicek, *Silicon Empires: The Fight for the Future of AI* (Cambridge: Polity Press, 2026). 162pp., £14.99 pb., 978 1 50955 049 4

In May 2024, Microsoft's CEO Satya Nadella announced the integration of Artificial Intelligence (AI) and Personal Computing in the form of Copilot+ PCs. Portraying AI's future as the future of distributed interactivity, Nadella stated that its integration with PCs would turn the 'world itself' into a prompt. The statement points to a wide-ranging transformation of personal computing, driven by the hype around AI since the public launch of OpenAI's ChatGPT model in late 2022.

Turning the world itself into a prompt for AI might sound like a dystopian vision of the future. I use 'the future' in the singular and with a definite article here because computational capitalism can see only one future, which is the future of endless growth premised on AI. This future promises connectivity and interactivity, adding on to the network effects of the Internet, amplifying and indeed changing them in many ways. For Microsoft, as for Big Tech actors such as GAMA (Google, Amazon, Meta and Apple), the promise of a totally networked life is emancipatory for modern productivity and creativity. For some, this might seem like just another hyped narrative that is being sold by tech bros and enthusiasts, one which can consequently be ignored. But there is more at stake here than mere promises. Microsoft's announcement, worrying as it is, indicates the next stage in the endless growth of the demand for computing which intertwines geopolitics, trade and technology.

Nick Srnicek's new book, *Silicon Empires*, responds to the global struggle for computing power. The demand for computing power is driving new economies of scale – hyperscaling – that are reconfiguring relations between the transnational capitalist class and the state, in addition to upending the well-established neo-liberal principles of global trade. Srnicek's book is a serious and systematic attempt at unravelling the complex 'geopolitical economy of contemporary AI'. But what exactly is the geopolitical economy of contemporary AI? To elaborate, Srnicek first turns to setting up his thinking around AI's value.

We have come to associate AI primarily with chatbots, such as those developed by OpenAI (ChatGPT), Anthropic (Claude) and Google (Gemini). But for Srnicek the story only begins there. The geopolitical economy of AI moves substantially beyond inference computing. Srnicek is interested in how proposed AI strategies think about innovation, international cooperation and the diffusion of automation. His approach here parallels recent literature on social histories of AI, such as Matteo Pasquinelli's *The Eye of the Master* and Shoshanna Zuboff's *Surveillance Capitalism*, but is distinguished from the former by its focus on geopolitical competition and Big Tech and from the latter by its timely intervention into debates on changes from neoliberal modes of trade to protectionism.

Srnicek offers an analytical structure for the political economy of AI: the vertical stack. The vertical stack consists of four layers: hardware, infrastructure, models and applications. At the hardware level, Nvidia, an American company, has the monopoly, holding up to 92% of the market for graphics processing units (GPUs). GPUs are essential for performing AI calculations. Though Nvidia's advantage is slowly melting as big players, which include AMD and GAMAM (GAMA + Microsoft), enter competition for hardware, their sheer power over this element in the vertical stack makes them a key factor in the political economy of AI. Closely connected is the infrastructure layer, which is an oligopoly shared between Microsoft Azure, Amazon Web Services and Google Cloud. They make up around 64% of the global cloud market. This layer 'racks' chips together in energy and water-intensive data centres to form networked systems that underpin cloud computing.

In contrast to the other two layers, the model layer seems to be highly competitive. It is where the development of Large-Language Models (LLMs) is located. The model layer seems to benefit from wide distribution in that many organisations – as many as 96 in 2023 – release their own models. The most famous and competitive layer is made of applications. This is where we find ChatGPTs and other apps. Besides being the most competitive, the apps layer is also the most unstable, with many companies taking their approaches to apps development

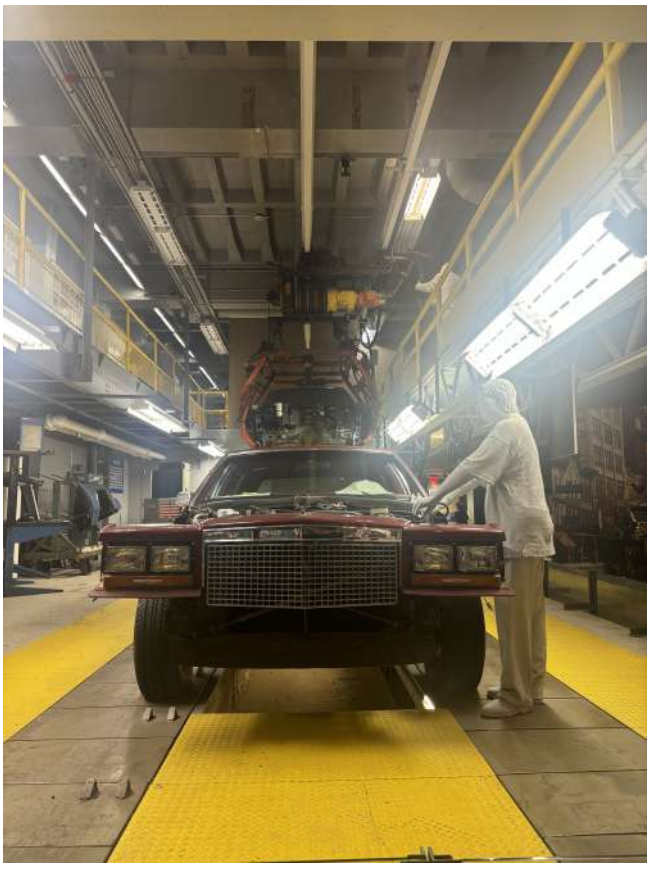
based on preferred models and available infrastructures.

The vertical stack helps to explain the growing AI bubble. Srnicek argues that AI development has so far been built on promises backed by major investments but without much profit. This is complicated by the enormous costs – both financial and energy – involved in developing AI. For example, Google's Gemini 1.0 (2023) cost $200 million. Price competition also drives issues across the stack, with companies offering models at increasingly lower prices to out-price competitors. Such constraints to profits drive different approaches to AI.

Gaining a competitive edge in the AI economy is a no holds barred game. Big Tech is making smaller developers dependent on their funding for growth. In 2023, two-thirds of USD27 billion investment in AI came just from Amazon, Google and Microsoft. Acquisition too has become a strategy of what Srnicek calls 'control without ownership'. It involves poaching the workforce of start-ups without buying the companies themselves. Such strategy avoids regulatory oversight while transferring epistemic and social capital to the big players.

Different approaches to AI also reflect the structural needs of companies. For example, approaches taken by OpenAI and Microsoft are different because OpenAI is focusing on developing the base AI model that can then be tweaked by each user – individual and company alike – for their own purposes. Known as the frontier strategy, it reduces costs and increases vertical adoption, which brings profits. By contrast, Amazon and Microsoft pursue an infrastructure strategy. They do not focus as much on AI development – Google outperforms them in AI research – but try instead to leverage their cloud infrastructures to capture the market. Microsoft's deal with Paige – a healthcare AI provider – sees the former investing in the latter's solutions while bringing them on to its Azure cloud. Though most companies deploy more than a single strategy, they serve as 'analytical lenses' to make sense of Big Tech activities.

For Srnicek, such strategies are projected internationally and seep into both global and daily politics. The silicon empires of the US are aiming to expand computation as much as possible, which brings them into conflict with their Chinese counterparts – Alibaba, Tencent and Huawei – reflecting on the conduct of world politics and shaping culture. Reflecting this, Srnicek switches from the analysis of silicon empires to geopolitics halfway through the book. He says that 'if the book had been written a decade ago', the vertical stack is where the story would have stopped. But capitalist competition has been reconfigured since then.



The move from AI markets to world politics and back, which Srnicek makes with ease, is also what makes the book so powerful. It connects the vertical stack, and the different strategies deployed to making AI solutions widely available, to both geopolitical manoeuvring at the highest levels and to daily political cultures. Long gone are the days of the Silicon Valley model of neoliberal free trade. We now experience global technological hegemony, led by approximately ten companies and a similar number of individuals. The main features of this hegemony are increasing tech protectionism – via trade bans and tariffs – and reactionary politics. They drive a new kind of post-Cold War geopolitics whose focus is 'the pursuit of control over strategic hubs in global networks as opposed to containment policy.' Furthermore, hegemony has brought a transnational technological class into view, amplifying the fault lines in democratic societies. In short, an interregnum.

The US, China and their tech elites are central to

the current interregnum. If the US strategy is advancing an innovation model in general-purpose technologies (GPT), the Chinese strategy is more open, focusing on diffusion. The two models argue respectively that innovation needs to be maximised by promoting specific AI models and that diffusion across various industries is more favourable economically. These are not the only strategies out there – Europe, stuck between the two, might still offer something new – but the two great powers have largely profiled themselves along the described lines. The two approaches cause complications because to advance an innovation model the US needs to engage in trade protectionism, while China needs to find ways to navigate constraints in attaining cutting-edge computing.

Srnicek utilises Jeffrey Ding's analysis of GPT diffusion, articulated in *Technology and the Rise of Great Powers* (2024), to argue that the Chinese strategy might prevail over the US's in the long-term. Ding has argued that innovation strategies (also known as Leading Sector) emphasise innovation over distribution. Srnicek follows him here to argue that what matters with tech development is not so much innovation (the US strategy) but diffusion (the Chinese strategy). However, China does have its work cut out for itself. The US has distinct advantages in terms of the vertical stack, including access to funding and cutting-edge computation. The way China navigates these constraints while advancing its AI strategy will determine if it can win the AI race.

The emergence of Big Tech geopolitics and promises of AI have become intertwined, coupling big money interests with protectionist policies at an unprecedented scale. Srnicek is careful to distinguish between Big Tech and smaller but no less important tech companies, which are increasingly involved in political life due to their strategic importance, *inter alia*, in defence and arms development. 'The new tech right' is contrasted to the 'more orthodox and unified neoliberals of Big Tech' because of its diversity, ranging from 'libertarian network state utopias' to 'openly eugenicist views', and its connection to the rise of the far right across the world. Srnicek takes care not to lump these diverse actors under one ideological banner, but points to 'anti-woke' discourse as a unifying element.

The new tech right benefits more directly than Big Tech from state contracts because it is made up of smaller companies and is therefore more protectionist and less interested in free trade. Palantir and Anduril are examples, with the latter edging Microsoft out of a USD22 billion 'augmented reality programme for the military'. Srnicek's distinction provides ample analytical opportunities by focusing away from Big Tech and onto smaller tech actors that, disparate as they are, play an important role in the geopolitics of AI dominance.

If we live in an interregnum, with several different possibilities for the future of AI before us, then we need a map to understand struggles that will define those futures. Srnicek's *Silicon Empires*, written in a style that is accessible to both generalist and specialist audiences, and that soberly elucidates the transformation that we are witnessing, can serve as one such map.

Toni Čerkez