A transition orchestrated from Big Tech clouds?

Authors: Cecilia Rikap and Silvia Weko

Long abstract

Amazon, Microsoft and Google are the big winners of the AI age. This paper's contribution is to show how, on this basis, they are positioning themselves to steer and even control an ecological transition fit for their purpose.

These three giants control the whole AI value chain from the required datasets and research networks up to the start-ups producing AI applications. It all runs on their clouds, depends on their models, relies on their funding or is indirectly influences by their control of the field. Evidence of their research field dominance is that Microsoft and Google occupy the most central role of the global academic network of co-authorship of papers presented at the most prestigious AI academic conventions and that, together with Amazon and sometimes other large tech companies they sit in the boards organizing those conferences (Rikap, 2024b).

Mushrooming AI startups, sooner or later, integrate these three giants' spheres of control. Most of these start-ups rely on their corporate venture capital investments. Amazon, Microsoft and Google poured more money into the AI start-up universe than any venture capitalist in 2023, investing two-thirds of a total of USD 27 billion raised.[1] Looking at the top 5 investors of every AI company receiving funding since December 2022 until February 2024, Google ranks third in number of companies funded and Microsoft tenth (Rikap, 2024a). Wherever we look at, there is Big Tech.

And this is only the beginning because AI start-ups coding large language models (LLMs), as in the case of OpenAI, need processing power and a whole set of computing services to train and run their models. This has meant getting locked-in to a Big Tech cloud. Amazon, Microsoft and Google, in that order, concentrate 66% of the cloud computing market globally and this share is even higher in some countries.[2] The majority of the AI-startups are not coding new LLMs but using ChatGPT or Meta's LLM "Llama" as computing services offered on at those giants' clouds to code apps or other cloud computing services. Their business model is dependent on these cloud hegemons.

Even in the open-source space the largest projects in number of contributors are those shared by Microsoft, Google and Meta. These giants open source to turn portions of their solutions into industry standards and then develop platforms or other complementary products around them. Meta's Llama is in open source and available on GitHub, the most used developers' platform acquired by Microsoft. However, running an LLM requires a lot of processing power, which is unaffordable for most organizations. Llama can thus be accessed on one of the other Big Tech's clouds. Every time someone queries the model from those clouds it will be paying for the processing power and probably for many other computing services.

---

[1] https://www.ft.com/content/c6b47d24-b435-4f41-b197-2d826cce9532

[2] OfCom estimated in its cloud investigation that the three companies' market share could be near 80% in the UK.

This extended control has wide impacts on industry and research. The literature speaks widely of machine learning as a general-purpose technology and even an emerging general (or general-purpose) method of invention (Cockburn et al., 2018; Goldfarb et al., 2023). Machine learning algorithms significantly automate discoveries and expand the types of problems that can be addressed by analysing big data. As AI uptake expands as a method of invention, those controlling its key inputs -data, digital infrastructure and talent- can eventually control the production of many types of knowledge. This is only one reason why Big Tech dominance of the AI value chain is dangerous beyond economic concentration.

Some scholars are worried about another risk of the rapid expansion of AI and the cloud for science and industry: their ecological footprint (see for instance Crawford, 2021). Datacentres are powered and cooled with large amounts of energy and water. A new UNCTAD (2024) report showed that electricity consumption by the major datacentres' operators has increased around 150% between 2018 and 2022. Amazon's increase was 176%. The carbon footprint to train BERT, an AI LLM model of 110 millions of parameters, was roughly equivalent to a trans-American flight (Strubell et al., 2019). All else equal, a model like the recently released GPT-4, which is supposed to have a trillion parameters, would have polluted as much as 9,090 trans-American flights. Training is only the beginning because LLMs consume much more energy than other AI models when they are used. Google's BARD, then renamed Gemini, consumers 15 times more energy than a normal search on Google. Cooling servers down with water can also have environmental impacts, especially as data centres tend to be built in more water-scarce areas. Water demand is also rising. In 2022, Microsoft's water use grew 34% and Google's 22%.[3]

Big Tech tries to reduce their data centre emissions by using renewable electricity, and major cloud providers have rapidly installed renewables and negotiated power purchasing agreements with energy companies. However, even if their data centres run on 100% renewables, their massive energy demand still poses ecological problems for two reasons. First, using renewable electricity for AI means that there will be less available for other uses. The energy transition requires a massive amount of new renewable electricity to decarbonize sectors like transport and heat. If Big Tech eats up this electricity – as they can also pay more and have a better negotiating position than other actors – this could mean slower decarbonization overall. The other problem is that there are still environmental impacts associated with data centres, even if they are run purely on renewable electricity including electronic waste and water stress. Using AI when it is not needed therefore adds to the ecological breakdown.

But this is not the only effect of Big Tech dominance of AI and the cloud on the prospects of ecological transitions. Among AI's countless applications, Weko (2024) notes that the energy transition offers a unique opportunity to apply AI for managing energy flows, which results in a potentially profitable business for those that control key technologies and digital infrastructures, a.k.a. the cloud hegemons. This seminal work opens an area of research that examines how cloud hegemons benefit from ecological transitions.

The space for Big Tech to offer digital solutions for ecological issues is growing as governments and businesses invest in research and development for finding solutions to the ecological crisis and integrate ecological risks and needed action to their economic calculations. As we

[3] Cristina Criddle and Kenza Bryan, 'AI boom sparks concern over Big Tech's water consumption', Financial Times, 25 February 2024.

mentioned above, AI is a method of invention increasingly applied to every science and technology domain, managing and researching on ecological matters is not an exception. In fact, there is a political push in Europe and the US that prompts what has been dubbed the "twin transition", where digital technologies should help accelerate sustainability transitions. However, this initiative and the research that studies whether there is a twin transition neglects that digital technologies are increasingly monopolized by a handful of giant companies (see also Rikap & Lundvall, 2021).

In this contribution, we move a step further and argue that cloud hegemons are not simply creating profitable businesses given the need for a transition, as noted by Weko (2024). As we show here, these giants have hatched and honed a comprehensive strategy for an ecological transition that reinforces their economic and political power. Their control of crucial intangibles for the energy transition (Weko, 2024) is one of many interrelated pieces of a comprehensive strategy to plan an ecological transition that reinforces their intellectual monopolies, expands their profits and ultimately entrenches their ruling power.

To provide evidence of this strategy, we offer unique comparison of a virtually untapped set of sustainability reports by Amazon and Google. Our findings point to a 3-tier strategy.

First, cloud giants promote a siloed -business friendly- approach to the transition. They offer discrete solutions based on applications of the AI supply chain by finding data, processing them with AI and using resulting insights. They thus offer ecologically friendlier applications and insist on how their offerings enable a more efficient use of energy for consumers without acknowledging the ecological footprint of the production of digital technologies. From the perspective of the ecological transition, this hinders the systemic view that is needed for guiding us towards a sustainable future.

Second, theirs is an energy transition that pushes for an unrestricted use of AI and data solutions for ecological problems. Since the ecological footprint is outsourced on the cloud, it contributes to their clients' outsourcing and thus (strategic) ignorance of the environmental impact of their actions. Cloud hegemons are aiming to use exclusively renewable energy in their datacentres, but this comes with its own problems. As above, using clean electricity for AI rather than for decarbonizing challenging sectors like transport and heating may slow down the transition to renewables. In addition, using renewable electricity does not address the other environmental impacts of data centres, including waste and water consumption. By pushing AI and data solutions for ecological problems, Google and Amazon ensure the ever-expanding consumption of cloud services, thus energy, water and materials. As they foster a transition that poses no limits to the consumption of data, AI and their associated use of nature, other transition pathways that could be more rapid and inclusive lose grounds.

Third, the dominant discourse imposed by cloud hegemons' transition not only endorses techno-solutionism, which is certainly problematic because it neglects the political economy and geopolitics of the ecological crisis, but also fosters the use of their AI-powered solutions, thus their cloud, as the method for developing the technologies that are expected to solve the ecological breakdown. This means promoting science and technology that lays under their control. As more, often public, resources are dedicated to conduct R&D for the transition, which is already a priority in most of the world's science and technology bodies, a larger share will eventually rely on AI and other digital technologies.

The whole environmental sciences field could become subordinated to Big Tech priorities as the discipline massively adopts AI, and thus the cloud, for their R&D. On top of getting richer, cloud hegemons position themselves in a place to steer the R&D for the transition. This would strengthen Big Tech intellectual monopolization by further positioning their intangible assets and their associated infrastructure -seen as a means of appropriation of those intangibles (Bensussan et al., 2023)- as the linchpin of the transition.

All in all, as cloud hegemons curate an ecological science and industry based in AI and the cloud, they are doing much more than simply expanding their business. They are encroaching what should be states' prerogatives by advancing (implicit) rules on how the transition should take place and positioning their solutions as unavoidable passages in environmental supply chains.

Far from being just and respectful of planetary boundaries, a transition orchestrated from the(ir) clouds is not a just transition but a business of the few that is paid by the many. Their transition relies on green accumulation as it is anchored on economic growth and accumulation goals. Our article thus finishes by arguing that a transition planned by cloud hegemons is a dead end. The only possible transition is one that replaces the economic efficiency and accumulation mantra by being centred around human and planetary needs. This transition requires the termination of intellectual monopolies putting every form of knowledge at the service of a common production of alternatives.

References

Bensussan, H., Durand, C., & Rikap, C. (2023). 100 years of Corporate Planning. *University of Geneva Working Paper*.

Cockburn, I. M., Henderson, R., & Stern, S. (2018). *The impact of artificial intelligence on innovation*. National bureau of economic research.

Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Goldfarb, A., Taska, B., & Teodoridis, F. (2023). Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings. *Research Policy*, *52*(1), 104653.

Rikap, C. (2024a). *Dynamics of Corporate Governance Beyond Ownership in AI*. Common Wealth. https://www.common-wealth.org/publications/dynamics-of-corporate-governance-beyond-ownership-in-ai

Rikap, C. (2024b). Varieties of corporate innovation systems and their interplay with global and national systems: Amazon, Facebook, Google and Microsoft's strategies to produce and appropriate artificial intelligence. *Review of International Political Economy*, 1–29. https://doi.org/10.1080/09692290.2024.2365757

Rikap, C., & Lundvall, B.-Å. (2021). *The Digital Innovation Race: Conceptualizing the Emerging New World Order*. Palgrave Macmillan.

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *arXiv Preprint*.

UNCTAD. (2024). *Digital Economy Report 2024. Shaping an environmentally sustainable and inclusive digital future*. United Nations.

Weko, S. (2024). New sites of accumulation? Why intangible assets matter for energy transitions. *Review of Political Economy*.