Selected Papers of #AoIR2023:
The 24th Annual Conference of the
Association of Internet Researchers
Philadelphia, PA, USA / 18-21 Oct 2023

BIG AI: THE CLOUD AS MARKETPLACE AND INFRASTRUCTURE

Fernando van der Vlist
University of Amsterdam

Anne Helmond
Utrecht University

Fabian Ferrari
Utrecht University

Cloud infrastructure platforms underpin most of today's internet and web-based services. In fact, they are seen as "where the internet lives" (Holt and Vonderau, 2015). Amazon's Web Services (AWS) in particular manifests as the dominant operating system of the internet with a third (34%) of all cloud services running on it, followed by Microsoft Azure (21%), Google Cloud (11%), and Alibaba (5%) (Statista, 2022). Moreover, these cloud service and product offerings also represent an important source of revenue for Big Tech companies.

While most of the current hype around ("generative") AI is focused on specific products and firms like OpenAI (ChatGPT, DALL·E 2) and Stability AI (Stable Diffusion), they would not have been possible without the significant infrastructural support and investments from Big Tech companies. In the past years, Big Tech has engaged in various multi-year partnerships and billion-dollar investments in AI initiatives and companies, especially Microsoft (OpenAI), Amazon (Stability AI, Hugging Face), and Google (CohereAI, DeepMind, Anthropic). In these arrangements, Big Tech typically functions as an exclusive or "preferred" cloud infrastructure provider or "cloud partner" (Financial Times, 2023). These strategic alliances between AI companies and cloud partners are further cementing Big Tech's central role as key infrastructure provider, which also raises potential concerns around (fair) competition in the booming AI economy.

The Political Economy of AI

This paper critically examines what we call *Big AI*, or those types and deployments of AI that simply would *not* be feasible or even possible without the infrastructural support,

Suggested Citation (APA): van der Vlist, F., Helmond, A., & Ferrari, F. (2023, October). *Big AI: The Cloud as Marketplace and Infrastructure*. Paper presented at AoIR2023: The 24th Annual Conference of the Association of Internet Researchers. Philadelphia, PA, USA: AoIR. Retrieved from http://spir.aoir.org.

partnerships, or investments provided by Big Tech companies. To account for this, we propose a broad understanding of "AI" to articulate the key components and how they are connected. Specifically, we explain that AI involves the distinct convergence of models, applications (tools, products, and services), and the various infrastructural layers of software and hardware components that are in place for "AI" to work in the first place. Ultimately, the phenomenon of "AI" concerns *how* these different key components—models, applications, and infrastructure—are brought together to enable AI solutions in specific settings.

This research provides the basis for a better understanding of the critical political economy of (Big) AI. Our contribution is situated in relation to the existing literature on the political economy of AI, machine learning, and large technical systems (Ferrari and McKelvey, 2022; Luitse and Denkena, 2021; Rieder, 2022). Here, the deep integrations between hardware and software components of these systems, as well as the dependence of external software developers on Big Tech's hardware have been understood to further cement Big Tech's role as an infrastructure provider in the digital economy (e.g. Narayan, 2023; Rieder, 2022). The integration or embedding of AI in other (existing) systems and products similarly reflects the infrastructuralisation of this technology (cf. Burkhardt, 2020; Dyer-Whiteford et al., 2019; Mackenzie, 2023; Plantin et al., 2016), and invites us to consider Big Tech's role in facilitating the deep integration of AI in various cultural and economic sectors of society (cf. Van Dijck et al., 2018).

"Following" Big AI

The paper studies Big AI as an emerging phenomenon that is still, in many ways, *under development*. By focusing on products and service offerings, third-party applications, and models, we gain a more comprehensive understanding of what Big AI *is*, or looks like today, and what it may become in the years to come—for which the infrastructure is being made right now (cf. Burkhardt, 2020; Egliston and Carter, 2022; Helmond et al., 2019). This enables a better understanding of the ongoing social processes of the infrastructuralisation and commodification of "AI" as a distinct technology "stack", as a type of infrastructure, as a category of novel tools, products, services, and as a development culture.

Because the phenomenon of "AI" is so difficult to define and demarcate *a priori*, we instead propose "following" the actors (Big Tech) themselves, as lead firms of this emerging technology, to learn how *they* understand AI. In doing so, we make a distinction between the cloud platform products and service offerings from Big Tech (i.e. the cloud as an *infrastructure* for AI) and Big Tech as the host or provider of marketplaces for diverse (AI-based) products and services from third-party businesses and developers (i.e. the cloud as a *marketplace* for AI). In both cases, we argue that Big Tech plays a vital role in "convening" businesses and developers to build, capture, and ultimately "sell" AI (Egliston and Carter, 2022: 13–14). "Following" these actors and their various offerings for businesses and developers, we learn how AI is connected to cloud platforms and infrastructure services, together constituting the Big AI cloud ecosystem, and how this is in turn connected to application domains like health care. This type of approach draws from the STS literature on "following" actors and technical objects, and by methodologies from economic geography on following the journeys of commodities (Cook et al., 2006), even if we do not focus on tangible commodities per se.

To begin with, we explored the relation between Big Tech's cloud platforms and AI based on the various products and services offerings from Amazon, Microsoft, and Google using their respective developer-facing product information pages. Based on this information, we then analysed how these were related to business-facing AI tools, products, and services (how/where is AI situated or used within those offerings?). Next, we explored the contents of several cloud platform marketplaces (Amazon's AWS Marketplace, Google's Cloud Platform Marketplace, and Microsoft's Azure Marketplace), which list thousands of integrations and applications from third parties, including on "AI". This provides us with a better understanding of the marketisation and commodification of AI, and which actors are involved.

The "Trickle-Down" Power of Big AI

Overall, we find that Big Tech's cloud platforms simultaneously act as *infrastructure* for AI and as *marketplaces* for AI. In this regard, our findings echo Weigel's (2023) analysis of "Amazon's trickle-down monopoly", whereby the digital behemoth consolidates its unprecedented market power by exerting control over small business owners that act as third-party sellers on its marketplace. Although specific AI tools offered on Big Tech marketplace—including pre-trained models and data labelling offerings—are in high demand to streamline the production of AI, they remain hidden from public scrutiny. Unlike Weigel's study, however, there is also another side to this "trickle-down" power: Big Tech provides computing resources and infrastructure without which the AI solutions and integrations distributed on such proprietary marketplaces could not be deployed. Big AI is not just "big" because of this infrastructural dependence, but also because it provides marketplaces for the commodification of AI solutions. This gives Big Tech a unique form of corporate power that is "all-encompassing, shaping current and future markets" (Gerbrandy and Phoa, 2022: 176; Van der Vlist, 2022).

References

Burkhardt, M. (2020). Mapping the democratization of AI on GitHub: A first approach. In *Mapping the Democratization of AI on GitHub: A First Approach* (pp. 209–222). transcript Verlag. DOI: 10.1515/9783839447192-013.

Cook, I., et al. (2006). Geographies of food: Following. *Progress in Human Geography, 30*(5), 655–666. DOI: 10.1177/0309132506070183.

Dyer-Witheford, N., Kjøsen, A. M., & Steinhoff, J., (2019). *Inhuman Power: Artificial Intelligence and The Future of Capitalism*. Pluto Press.

Egliston, B., & Carter, M. (2022). 'The metaverse and how we'll build it': The political economy of Meta's Reality Labs. New Media & Society*, 0*(0). DOI: 10.1177/14614448221119785

Ferrari, F., & McKelvey, F. (2022). Hyperproduction: A social theory of deep generative models. *Distinktion: Journal of Social Theory, 0*(0). DOI: 10.1080/1600910X.2022.2137546.

Financial Times (2023). Big Tech companies use cloud computing arms to pursue alliances with AI groups. *Ars Technica*, February 6. Available at: https://arstechnica.com/tech-policy/2023/02/big-tech-companies-use-cloud-computing-arms-to-pursue-alliances-with-ai-groups/.

Gerbrandy, A., & Phoa, P. (2022). The power of big tech corporations as modern bigness and a vocabulary for shaping competition law as counter-power. In *Wealth and Power* (pp. 166–185). Routledge.

Helmond, A., Nieborg, D. B., & van der Vlist, F. N. (2019). Facebook's evolution: Development of a platform-as-infrastructure. *Internet Histories, 3*(2), 123–146. DOI: 10.1080/24701475.2019.1593667.

Holt, J., & Vonderau, P. (2015). "Where the Internet Lives": Data Centers as Cloud Infrastructure. In L. Parks & N. Starosielski (Eds.), *Signal Traffic: Critical Studies of Media Infrastructures* (p. 71–93). University of Illinois Press.

Luitse, D., & Denkena, W. (2021). The great transformer: Examining the role of large language models in the political economy of AI. *Big Data & Society, 8*(2). DOI: 10.1177/20539517211047734.

Mackenzie, A. (2023). 'This is a house': Large image collections and their platform embeddings. *Convergence, 29*(1), 169–182. DOI: 10.1177/13548565211052041

Narayan, D. (2023). Monopolization and competition under platform capitalism: Analyzing transformations in the computing industry. *New Media & Society, 25*(2), 287–306. DOI: 10.1177/14614448221149939.

Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2016). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society, 20*(1), 293–310. DOI: 10.1177/1461444816661553.

Rieder, B. (2022). Towards a political economy of technical systems: The case of Google. *Big Data & Society, 9*(2). DOI: 10.1177/20539517221135162

Statista. (2022). Infographic: Amazon, Microsoft & Google dominate cloud market. *Statista Infographics*, December 23. Available at: https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers

van der Vlist, F. N. (2022). *The Platform as Ecosystem: Configurations and Dynamics of Governance and Power* [Dissertation]. Utrecht University. DOI: 10.33540/1284.

van Dijck, J., Poell, T., & de Waal, M. (2018). *The Platform Society: Public Values in a Connective World*. Oxford University Press. DOI: 10.1093/oso/9780190889760.001.0001.

Weigel, M. (2023). Amazon's trickle-down monopoly. *Data & Society*, January 25. Available at: https://datasociety.net/library/amazons-trickle-down-monopoly/.