Selected Papers of #AoIR2024:
The 25th Annual Conference of the
Association of Internet Researchers
Sheffield, UK / 30 Oct - 2 Nov 2024

THE POLITICAL ECONOMY OF AI AS PLATFORM: INFRASTRUCTURES, POWER, AND THE AI INDUSTRY

Fernando van der Vlist
University of Amsterdam

Anne Helmond
Utrecht University

Dieuwertje Maria Rebecca Luitse
University of Amsterdam

Bernhard Rieder
University of Amsterdam

Sam Hind
University of Manchester

Max Kanderske
University of Siegen

The artificial intelligence (AI) sector is experiencing rapid growth, with a projected market size of $1.3 trillion by 2032 according to industry reports (Bloomberg, 2023). The landscape shifted significantly with the launch of ChatGPT in late 2022, prompting major players like Google, Amazon, Microsoft, and Meta, alongside popular apps such as TikTok and Snapchat, to make substantial investments in AI. There has been an influx of new AI products and updates, reshaping the industry's structure and scale. Additionally, there has been a surge in acquisitions and investments in AI startups, particularly by Big Tech firms (Alcantara et al., 2023). Furthermore, partnerships between AI and major tech companies have proliferated, solidifying their dominant positions (Kak and Myers West 2023; Jacobides et al., 2021; Van der Vlist et al., 2024). In fact, as Kak and Myers West (2023: 5) succinctly state, 'There is no AI without Big Tech', raising critical issues around industry concentration and the political economy of

Suggested Citation (APA): van der Vlist, F., Helmond, A., Luitse, D., Rieder, B., Hind, S., et al. (2024, October). *The Political Economy of AI as Platform: Infrastructures, Power, and the AI Industry*. Panel presented at AoIR2024: The 25th Annual Conference of the Association of Internet Researchers. Sheffield, UK: AoIR. Retrieved from http://spir.aoir.org.

AI (e.g., Crawford, 2021; Ferrari, 2023; Luitse and Denkena, 2021; Mackenzie, 2019; Narayan, 2022; Rieder, 2022; Widder et al., 2023).

This panel posits that the driving force behind these transformative shifts is the evolution of AI as a *platform*. This evolution effectively propels the platformisation of AI, facilitating the integration of AI across diverse industry sectors (cf. Helmond, 2015). The resultant 'industrialisation' of AI marks the expansion of AI systems across various sectors and industries, triggering investments in necessary computational resources and posing challenges for governing AI (Van der Vlist et al., 2024). In short, this underscores that AI is much more than just a standalone application or tool, such as ChatGPT; it is a foundational technical system that underpins a broad array of apps and services.

In this context, the panel recognises the recent 'infrastructural turn' in media and internet studies, deliberately steering away from speculative discussions about the future impacts of AI. Instead, the emphasis shifts towards a focus on the 'mundanity and ordinariness of existing systems' (Hesmondhalgh, 2022). This highlights the importance of studying the foundational infrastructure, tools, and frameworks that shape AI development. Furthermore, it requires an understanding of the associated supply chains, investments, acquisitions, forms of ownership and support, control mechanisms, and the broader political economy surrounding AI. Such perspectives have been developed, for example, to study AI's industry relations in healthcare (Luitse et al., 2023), the global digital marketing and advertising industries (Van der Vlist and Helmond, 2021), journalism (Rieder and Skop, 2021), or the automotive industry (Hind et al., 2022).

The panellists examine how AI may be viewed as a *platform*, presenting critical perspectives on the platformisation of AI and its implications for industry relations and the media landscape. Through four distinct studies, they highlight: (1) the influence of platforms on the emerging AI ecosystem and their consolidation of power through reliance on cloud infrastructure, (2) the evolution of cloud infrastructure in the political economy of AI, (3) the actualisation of AI as a platform with 'general-purpose' applications, and (4) how challenges in machine vision shape innovation in AI. Each contribution revolves around a central question: *How is AI, particularly within the AI sector, evolving under the influence of platform logic?* In doing so, the panellists offer valuable insights informed by platform theory and methodologies, exploring their relevance for a comprehensive examination of AI and the broader AI sector. Furthermore, their perspectives provide methodological insights into understanding the material conditions and critical political economy of AI as a platform.

Collectively, these studies seek to advance the critical discourse on AI and its political economy, with a specific emphasis on the AI industry. They shed light on the evolving landscape of AI industry relations and dependencies within the platform ecosystem, tracing how these relationships have transformed over time.

References

Alcantara C, Schaul K, De Vynck G and Albergotti R (2023) How Big Tech got so big: Hundreds of acquisitions. *The Washington Post*, 26 September. Available at: https://www.washingtonpost.com/technology/interactive/2021/amazon-apple-facebook-google-acquisitions/.

Bloomberg (2023) Generative AI to Become a $1.3 Trillion Market by 2032, Research Finds. *Bloomberg*, 1 June. Available at: https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/.

Crawford K (2021) *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Ferrari (2023) Neural production networks: AI's infrastructural geographies. *Environment and Planning F* 2(4): 459–476. DOI: 10.1177/26349825231193226.

Helmond A (2015) The platformization of the Web: Making Web data platform ready. *Social Media + Society* 1(2). https://doi.org/10.1177/2056305115603080.

Hesmondhalgh D (2022) The infrastructural turn in media and internet research. In: McDonald P (eds) *The Routledge Companion to Media Industries*. Palgrave Macmillan, pp. 132–142.

Hind S, Kanderske M and van der Vlist FN (2022) Making the Car 'Platform Ready': How Big Tech Is Driving the Platformisation of Automobility. *Social Media + Society* 8(2). DOI: 10.1177/20563051221098697.

Jacobides MG, Brusoni S and Candelon F (2021) The evolutionary dynamics of the artificial intelligence ecosystem. *Strategy Science* 6(4): 412–435. DOI: 10.1287/stsc.2021.0148.

Kak A and Myers West S (2023) *2023 Landscape: Confronting Tech Power*. AI Now Institute.

Luitse D and Denkena W (2021) The great Transformer: Examining the role of large language models in the political economy of AI. *Big Data & Society* 8(2): 1–14. DOI: 10.1177/20539517211047734.

Luitse DMR, Blanke T and Poell T (2023) AI Competitions as infrastructures: Examining power relations on kaggle and grand challenge in AI-driven medical imaging. *2022 AoIR Selected Papers of Internet Research*. https://doi.org/10.5210/spir.v2022i0.13044.

Mackenzie A (2019) From API to AI: Platforms and their opacities. *Information, Communication & Society* 22(13): 1989–2006. DOI: 10.1080/1369118X.2018.1476569.

Narayan D (2022) Platform capitalism and cloud infrastructure: Theorizing a hyper-scalable computing regime. *Environment and Planning A: Economy and Space* 54(5): 911–929. DOI: 10.1177/0308518X221094028.

Rieder B (2022) Towards a political economy of technical systems: The case of Google. *Big Data & Society*. DOI: 10.1177/2053951722113516.

Rieder B and Skop Y (2021) The fabrics of machine moderation: Studying the technical, normative, and organizational structure of Perspective API. *Big Data & Society* 8(2): 1–16. 10.1177/20539517211046181.

van der Vlist FN and Helmond A (2021) How partners mediate platform power: Mapping business and data partnerships in the social media ecosystem. *Big Data & Society* 8(1): 1–16. DOI: 10.1177/20539517211025061.

van der Vlist FN, Helmond A and Ferrari FL (2024) Big AI: Cloud infrastructure and the industrialisation of artificial intelligence. *Big Data & Society* 11(1). DOI: 10.1177/20539517241232630.

Widder DG, West S and Whittaker M (2023) Open (for business): Big Tech, concentrated power, and the political economy of open AI. Available at SSRN: https://doi.org/10.2139/ssrn.4543807.

CONVENING AI: HOW PLATFORMS SHAPE THE EVOLUTION OF THE AI ECOSYSTEM

Fernando van der Vlist
University of Amsterdam

Anne Helmond
Utrecht University

Introduction

The competition among leading technology companies in cloud-based artificial intelligence (AI), known as the 'cloud AI wars' (Goldman, 2022), is gaining momentum, with industry leaders proclaiming a transformative era driven by AI technologies. This shift, dominated by Amazon, Microsoft, and Google, is closely linked to their control over cloud infrastructure, significantly influencing the development of AI across various industries. This merger of AI innovation with the infrastructure and investments of major tech giants marks the rise of 'Big AI,' shaping not only the development and deployment of AI but also its commercialisation (Van der Vlist et al., 2024).

This paper critically explores cloud AI as a platform, emphasising the critical roles of Amazon, Microsoft, and Google. It demonstrates how these companies not only provide essential infrastructure services but also '*convene'* enterprises, organisations, and developers in the development and commercialisation of AI. These convening dynamics encapsulate what we refer to as the platform ecosystem logic of AI.

Through a 'technographic' approach (Bucher, 2016; Mackenzie, 2019; Van der Vlist et al., 2022), we examine the infrastructure support, investments, partnerships, and product offerings of these companies' cloud platforms. This shows the emergence of new platform ecosystems around their AI products and services. We show how the transformation of cloud AI into a platform-based model (platformisation) manifests, how Big Tech 'convenes' third-party businesses and developers to cultivate an ecosystem around cloud AI infrastructure, and discuss the consequences of this emerging ecosystem logic of AI.

Platforms and the Political Economy of Cloud AI

This paper builds upon existing critical research concerning the political economy of AI, platformisation, and platform studies from an interdisciplinary perspective. The political economy involves examining how the development, deployment, and impact of AI technologies, as well as their integration into broader economic systems and structures, *shape* and are shaped *by* major technology corporations.

Research shows that the development of foundational models like GPT, which use a 'bigger-is-better approach' (Economist, 2023), requires substantial cloud computing resources, thus favouring larger companies (Ferrari, 2023; Jacobides et al., 2021; Kak and Myers West, 2023; Luitse and Denkena, 2021). The unparalleled scalability offered by cloud computing, along with the incorporation of customised hardware and software, emerges as a defining feature of AI's material political economy (Narayan, 2022; Rieder, 2022; Raley and Rhee, 2023), leading to new dependencies and corresponding investments in computational resources (Van der Vlist et al., 2024). The platformisation of AI is rooted in Big Tech companies' control over essential infrastructure, echoing the statement that 'There is no AI without Big Tech' (Kak and Myers West, 2023). This platformisation is further facilitated by developers and partners who build and integrate new AI apps and services atop Big Tech's infrastructure (Van der Vlist et al., 2024).

The platformisation of AI is also marked by its 'industrialisation', representing the transition of AI systems from research and development to practical applications across diverse industries, including governments and public sectors, digital sectors, and traditional industries worldwide (Van der Vlist et al., 2024). Leading cloud infrastructure service providers like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform, through their offerings and boundary resources, have emerged as the material bedrock of this industrialisation process.

Big Tech's ambitions to expand their offerings and assume a central, 'infrastructural role' within society are evidenced by their attempts to consolidate complementors through acquisitions, 'convene' third-party businesses and developers, and expand their ecosystems by enhancing platform programmability (Egliston and Carter 2022; Helmond et al., 2019; Van der Vlist and Helmond, 2021). Similarly, partner programmes play a crucial role for platforms to grow into new markets or industry sectors and expand beyond their current boundaries (Helmond et al., 2019), using partnerships to solidify their 'infrastructural and strategic power' in the larger platform business ecosystem (Van der Vlist and Helmond, 2021; Van der Vlist, 2022). These studies emphasise the critical dynamics of '*convening'* complementors (Egliston and Carter, 2022)—initially developers and businesses, and later extending to media publishers, creators, and others—to actively contribute to the development, capture, and commercialisation of AI.

Three Ways of Seeing Cloud AI as a Platform

Drawing from our technographic analysis, this paper views the political economy of AI through the critical lens of platform ecosystems, characterised by the combination of a central technical platform in the 'cloud' (accessible via application programming interfaces and other platform boundary resources) and a network of complements (applications and services) and complementors (the businesses and developers behind these applications and services) (Van der Vlist, 2022). This framework illustrates how application and development tools, partner programmes, and investments form a comprehensive collection of boundary resources. These resources 'convene' third-party businesses and developers to build an ecosystem of new products and services around cloud AI infrastructure platform providers.

The paper examines cloud AI as a platform from three distinct perspectives. First, it explores infrastructure support relationships, such as partnerships and investments, which often evolve into dependencies on platforms over time. Second, it analyses the existing cloud AI platform products and services, showing how they target the diverse AI needs of various organisations and industry sectors. Third, it examines the ecosystems of applications and solutions developed by third parties on top of Microsoft, Google, and Amazon's cloud platforms, accessible via their cloud app stores and marketplaces. These viewpoints show the material conditions of AI as a development platform and highlight the importance of analysing its political economy from a critical 'ecosystem' perspective (Van der Vlist, 2022).

Shaping the Evolution of the AI Ecosystem

The study demonstrates how AI is not merely an emerging discursive phenomenon but also comprises extensive suites of tools, products, and services, along with respective documentation for external stakeholders, intended to facilitate an ecosystem of complementors atop Microsoft, Google, and Amazon's platforms.
Crucially, these platform 'ecosystems', as proposed by Van der Vlist (2022) and illustrated by Egliston and Carter (2022), are not static or passive entities but are actively *cultivated* through strategic and infrastructural initiatives by leading platform companies. Central to this cultivation is the platform's ability to *convene* external stakeholders, such as partners and developers, and orchestrate their contributions to the ecosystem's value. Essentially, platform owners seek to wield what the paper conceptualises as 'convening power', aligning participants' interests and business and development activities to reinforce their vision. Their visions range from Meta's 'future of connection in the metaverse' (Meta), Microsoft's vision for AI in the enterprise (Microsoft), Amazon's ambition of 'AI becoming the new electricity in our homes' (Zeghari, 2023), to Google's portrayal of AI as 'a foundational and transformational technology that will provide compelling and helpful benefits to people and society through its capacity to assist, complement, empower, and inspire people in almost every field of human endeavor' (Google AI).

In conclusion, this research offers insights into the political economy of AI, viewed through the critical lens of platform ecosystems, highlighting the pivotal role played by cloud infrastructure service providers in shaping the industrialisation of AI. In this context, 'Big AI' is reshaping traditional industries ranging from healthcare and the public sector to manufacturing, automotive, retail, and energy. This process is not only influencing market dynamics but also raising critical questions about governance and control within the AI platform ecosystem.

References

Economist (2023) The bigger-is-better approach to AI is running out of road. Available at: https://www.economist.com/science-and-technology/2023/06/21/the-bigger-is-better-approach-to-ai-is-running-out-of-road.

Egliston B and Carter M (2022) 'The metaverse and how we'll build it': The political economy of Meta's Reality Labs. *New Media & Society* 0(0): 1–25. DOI: https://doi.org/10.1177/14614448221119785.

Ferrari, F (2023) Neural production networks: AI's infrastructural geographies. *Environment and Planning F* 2(4), 459–476. DOI: 10.1177/26349825231193226.

Goldman S (2022) Big Tech cloud AI wars heat up. In: *VentureBeat*, 14 October. Available at: https://venturebeat.com/ai/the-ai-beat-big-tech-cloud-ai-wars-heat-up/.

Google AI (n.d.) Why we focus on AI (and to what end). Available at: https://ai.google/why-ai/.

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

Jacobides MG, Brusoni S and Candelon F (2021) The evolutionary dynamics of the artificial intelligence ecosystem. *Strategy Science* 6(4): 412–435. DOI: 10.1287/stsc.2021.0148.

Kak A and Myers West S (2023) *2023 Landscape: Confronting Tech Power*. AI Now Institute. Available at: https://ainowinstitute.org/2023-landscape.

Luitse D and Denkena W (2021) The great Transformer: Examining the role of large language models in the political economy of AI. *Big Data & Society* 8(2): 1–14. DOI: 10.1177/20539517211047734.

Meta (n.d.) The metaverse is the future of digital connection. Available at: https://about.meta.com/metaverse/.

Microsoft (n.d.) Microsoft's vision for AI in the enterprise. Available at: https://info.microsoft.com/rs/157-GQE-382/images/Microsoft%20Enterprise%20AI%20white%20paper.pdf.

Narayan D (2022) Platform capitalism and cloud infrastructure: Theorizing a hyper-scalable computing regime. *Environment and Planning A: Economy and Space* 54(5): 911–929. DOI: 10.1177/0308518X221094028.

Raley R and Rhee J (2023) Critical AI: A field in formation. American Literature 95(2): 185–204.

Rieder B (2022) Towards a political economy of technical systems: The case of Google. *Big Data & Society*. DOI: 10.1177/2053951722113516.

van der Vlist FN (2022) *The Platform as Ecosystem: Configurations and Dynamics of Governance and Power*. PhD Thesis, Utrecht University. DOI: 10.33540/1284.

van der Vlist FN and Helmond A (2021) How partners mediate platform power: Mapping business and data partnerships in the social media ecosystem. *Big Data & Society* 8(1): 1–16. DOI: 10.1177/20539517211025061.

van der Vlist FN, Helmond A and Ferrari FL (2024) Big AI: Cloud infrastructure and the industrialisation of artificial intelligence. *Big Data & Society* 11(1). DOI: 10.1177/20539517241232630.

Zeghari A (2023) Amazon's vision: AI Becoming the new Electricity in our homes. Medium, 23 June. Available at: https://medium.com/@ayoub.zeghari/amazons-vision-ai-becoming-the-new-electricity-in-our-homes-b120bb0ed1.

PLATFORM POWER IN AI: THE EVOLUTION OF CLOUD INFRASTRUCTURES IN THE POLITICAL ECONOMY OF ARTIFICIAL INTELLIGENCE

Dieuwertje Maria Rebecca Luitse
University of Amsterdam

Introduction

Powered by the promise that Google CEO Sundar Pichai (2017) declared as the shift from a 'mobile first to an AI first world,' Amazon, Microsoft, and Google have become the three dominant developers of artificial intelligence (AI) infrastructures and services (Srnicek, 2022). Beyond leveraging their vast amounts of data, these corporations have equipped themselves with technical knowledge through the attraction of AI expertise, the acquisitions of AI startups (e.g., Google/Deepmind) and extensive business partnerships (e.g., Brockman, 2019). Additionally, the last years have seen significant investments in the infrastructural expansion of their cloud computing platforms: Amazon Web Services (AWS), Microsoft Azure, and Google Cloud. These platforms now provide full-stack tools and services for the production and deployment of machine-learning systems and applications thriving on their proprietary infrastructures for computing power at scale (Luitse and Denkena, 2021; Van der Vlist et al., 2024). The most recent developments include the provision of foundation models (Bommasani et al., 2022) such as OpenAI's DALL·E (Ramesh et al., 2021) or Google's PaLM (Chowdhery et al., 2022).

Driven by AI hype, corporate cloud AI infrastructures and services are increasingly implemented across industries. Following research on AI's political economy (Luitse and Denkena, 2021; Widder et al., 2023), this has allowed AWS, Microsoft Azure and Google Cloud to leverage them as core 'commercial computing assets' (Narayan, 2022) resulting in a rapid concentration of economic and political power (monopolisation) toward these corporations (Kak and Myers West, 2023; Srnicek, 2022; Whittaker, 2021). Building on this body of work, this paper provides a case study on how these three dominant cloud platforms have strategically been operationalising their power in AI production and deployment through the AI infrastructures and services they operate.

The critical analyses on big tech's monopoly power in AI have put forward valuable macro-level insights. However, relatively little attention has been paid to specific micro-material ways power in AI is operationalised by AWS, Microsoft Azure and Google Cloud through the evolution of their cloud AI infrastructures and services. Such research is important as 'cloud computing arrangements […] are foundational to platform expansion' (Narayan, 2022: 915). As evolving assemblages of infrastructural hardware and software services they set the conditions for AI production and deployment (Jacobides et al., 2021; Rieder, 2022), shaping the present in terms of 'what we do (and do not) know about [AI]' (Whittaker, 2021). These cloud platforms therefore warrant a deeper examination into the cloud AI infrastructures and services they have developed over time. As such, we can better understand the ways in which these companies have strategically manifested themselves in the field and attempt to exercise power in AI through their cloud ecosystems.

This paper provides such an inquiry through a critical empirical investigation into the evolution of AWS, Microsoft Azure and Google Cloud in the context of AI in the run-up to the current shift to foundation models. That is, primarily since Sundar Pichai's 2017 defining proclamation up to April 2021. Following a political economy and evolutionary platform studies research approach, I investigate the development of AWS, Microsoft Azure and Google cloud's AI infrastructures and services over this period of time and show how these cloud platforms operationalise specific forms of infrastructural power to further solidify their position.

Evolutionary Technographic Analysis of Infrastructures for Cloud AI

To conduct this study, I developed a methodological approach that I call evolutionary platform technography. Adapted from Bucher's (2018) notion of technography, and Helmond and Van der Vlist's (2021) work on historical platform studies, this method draws from a wide variety of sources including (1) archived product pages and documentation from each platform available through the Internet Archive Wayback Machine; (2) AWS, Azure and Google Cloud's corporate blog posts and press releases; (3) industry reports.

The evolutionary technographic analysis entailed two complementary lines of inquiry. First, I examined the collection of archived product pages to reconstruct the development of AWS, Microsoft Azure and Google Cloud's cloud AI infrastructures and services over time. This mapping of the evolution of these cloud platforms provides empirical insight into how the AI infrastructures and services have been developed according to the service layers of the cloud stack—Infrastructure-as-a-service (IaaS); Platform-as-a-service (PaaS) or AI-as-a-service (AIaaS)—as well as their application domain.

As technography requires the critical observation, description, and interpretation of technical systems on their own material-discursive terms (Bucher, 2012), the second step involved a document analysis and close reading of the set of collected archived product documentations, blog posts, press releases and industry reports. Here, I focused on distinguishing the strategic positions of AWS, Microsoft Azure and Google Cloud in mobilising and operating their growing collection of AI infrastructures and services. By outlining these corporate dynamics, this technographic exploration into the evolution of their platform-specific operations allows me to gain empirical insight into the specific ways AWS, Microsoft Azure and Google Cloud seek to exert infrastructural power in the political economy of AI.

Infrastructural Power in Cloud AI: Vertical integration, Complementary innovation, Abstraction

The evolution of AWS, Microsoft Azure and Google Cloud's infrastructures for cloud AI shows that these platforms strategically operationalise infrastructural power in three significant ways: through 1) vertical integration; 2) complementary innovation; and 3) the power of abstraction. First, the widespread vertical integration of resources across the cloud stack strengthens these companies' attempts to privatise and standardise entire machine-learning workflows. This creates path-dependencies which could lead to user and vendor lock-in (cf. Van Dijck, 2020).

Second, because AI systems can be designed for many different application domains (Aradau and Blanke 2022; Rieder, 2022), these technologies enable AWS, Microsoft Azure and Google Cloud to facilitate 'complementary innovation' (Gawer, 2014) in two directions. On the one hand, this facilitates the strategic development of complementary services to infrastructurally expand themselves into different application domains, such as healthcare, manufacturing or retail and set the conditions of possibility for AI development in these respective areas moving forward. On the other hand, the analysis shows that new machine-learning capabilities operate in complementary ways that seamlessly align with other parts of the larger platform ecosystems of Amazon, Microsoft and Google such as Amazon Mechanical Turk.

Third and lastly, I locate that these cloud platforms strategically operationalise 'the power of abstraction' to strengthen their position in AI's political economy. While abstraction is considered to play a central role in computer science practices (Selbst et al., 2018; Rieder, 2020), the analysis shows how these platforms mobilise the ability to hide the complex operations of their cloud infrastructures and services across the layers of the stack. While this simplifies and speeds up machine-learning development pipelines, it also creates significant advantages for cloud platforms to structure these practices in ways that further reinforce the creation of AI systems and applications that rely on the data, computing infrastructures and machine-learning expertise they own and operate. Additionally, the operationalisation of abstraction thwarts the critical scrutiny and evaluation of AI systems even though there have been ever more calls for critical oversight (e.g., Buolamwini and Gebru, 2018; Bender et al., 2021; Kak and Myers West, 2023). As such, abstraction provides AWS, Microsoft Azure and Google Cloud with the infrastructural power to shift the focus away from alternative resources and the development of new approaches that contribute to different understandings about AI outside of the confining ecosystems of the cloud.

References

Aradau C and Blanke T (2022) *Algorithmic Reason: The New Government of Self and Other*. New York: Oxford University Press.

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*, New York, NY, USA, 3 March 2021, pp. 610–623. FAccT '21. Association for Computing Machinery. DOI: 10.1145/3442188.3445922.

Bommasani R, Hudson DA, Adeli E, et al. (2022) On the Opportunities and Risks of Foundation Models. arXiv:2108.07258. arXiv. http://arxiv.org/abs/2108.07258.

Brockman G (2019, 22 July) Microsoft invests in and partners with OpenAI to support us building beneficial AGI. Available at: https://openai.com/blog/microsoft-invests-in-and-partners-with-openai.

Bucher T (2012) *Programmed sociality: A software studies perspective on social networking sites* (Doctoral dissertation). University of Oslo, Oslo, Norway.

Bucher T (2018). *If…Then: Algorithmic Power and Politics*. New York: Oxford University Press.

Buolamwini J and Gebru T (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research 812018 Conference on Fairness, Accountability, and Transparency*, (1–15). https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Chowdhery A, Narang S, Devlin J, et al. (2022) PaLM: Scaling Language Modeling with Pathways. arXiv:2204.02311. arXiv. http://arxiv.org/abs/2204.02311.

Gawer A (2014) Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, *43*(7): 1239–1249. DOI: 10.1016/j.respol.2014.03.006 .

Gray Widder D, West S and Whittaker M (2023) Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.4543807.

Helmond A and van der Vlist FN (2021) Platform and app histories: Assessing source availability in web archives and app repositories. In D Gomez, E Demidova, J Winters and T Risse (Eds), *The Past Web: Exploring Web Archives*. Cham, Switzerland: Springer, pp. 203–214. DOI: 10.1007/978-3-030-63291-5_16.

Jacobides MG, Brusoni S and Candelon F (2021) The evolutionary dynamics of the artificial intelligence ecosystem. *Strategy Science* 6(4): 412–435. DOI: 10.1287/stsc.2021.0148.

Kak A and Myers West S (2023) *2023 Landscape: Confronting Tech Power*. AI Now Institute. Available at: https://ainowinstitute.org/2023-landscape.

Luitse D and Denkena W (2021) The great Transformer: Examining the role of large language models in the political economy of AI. *Big Data & Society* 8(2): 1–14. DOI: 10.1177/20539517211047734.

Narayan D (2022) Platform capitalism and cloud infrastructure: Theorizing a hyper-scalable computing regime. *Environment and Planning A: Economy and Space* 54(5): 911–929. DOI: 10.1177/0308518X221094028.

Pichai S (2017) *Google I/O'17 Keynote*. Google. https://www.youtube.com/watch?v=Y2VF8tmLFHw.

Ramesh A, Pavlov M, Goh G, et al. (2021) Zero-Shot Text-to-Image Generation. arXiv:2102.12092. arXiv. http://arxiv.org/abs/2102.12092.

Rieder B (2020) *Engines of Order: A Mechanology of Algorithmic Techniques*. Amsterdam: Amsterdam University Press.

Rieder B (2022) Towards a political economy of technical systems: The case of Google. *Big Data & Society*. DOI: 10.1177/2053951722113516.

Selbst AD, Boyd D, Friedler SA, et al. (2019) Fairness and Abstraction in Sociotechnical Systems. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*, New York, NY, USA, 29 January 2019, pp. 59–68. FAT* ’19. Association for Computing Machinery. DOI: 10.1145/3287560.3287598.

Srnicek N (2022) Data, compute, labor. In: Graham M, Ferrari F (eds) *Digital Work in the Planetary Market*. Cambridge MA: MIT Press, pp. 241–261.

van der Vlist FN, Helmond A and Ferrari FL (2024) Big AI: Cloud infrastructure and the industrialisation of artificial intelligence. *Big Data & Society* 11(1). DOI: 10.1177/20539517241232630.

van Dijck J (2020) Seeing the forest for the trees: Visualizing platformization and its governance. *New Media & Society*, *23*(9): 2801-2819. DOI: 10.1177/1461444820940293.

Whittaker M (2021) The steep cost of capture. *Interactions* 28(6): 50–55. DOI: 10.1145/3488666.

BECOMING PLATFORM: ON THE HETEROGENEOUS ACTUALISATION OF AI'S 'GENERAL-PURPOSE' POTENTIAL

Bernhard Rieder
University of Amsterdam

As 'artificial intelligence' (AI) has moved into various application spaces over the last two decades, investigations into the political economy of AI have proliferated, and the recent emergence of 'foundation models' (Bommasani et al., 2021)—large language models (LLMs), large vision-language models (VLMs), and other large pre-trained models—has added fuel to the fire. Economists (e.g., Trajtenberg 2018) have described AI as a 'general purpose technology' that is adopted and has impacts across a large number of economic sectors. More recently, Eloundou et al. (2023) have argued that even single LLMs like OpenAI's GPT-4 have 'general-purpose potential' as they are applicable to many different tasks in many different application settings.

Embedded within the broader rise to dominance of cloud-based provisioning of computing services (e.g., Narayan, 2022), the emergence of AI-as-a-Service (AIaaS) has increasingly led commentators to think of AI more generally and foundation models more specifically as *platforms* in both the computational sense as 'a computing system of any sort upon which further computing development can be done' (Bogost and Montfort, 2009, 2) and a broader commercial sense that includes—but is not limited to—the common 'multi-sided market' (Rochet and Tirole, 2006) setting. Here, attention is paid to the various ways hardware (e.g., storage and compute), software (e.g., libraries and frameworks), and higher-level services (e.g., Amazon's Business Metrics Analysis, OpenAI's GPT-4) are becoming key components in the development of virtually any AI application, both enabling and orienting AI production and deployment. From a political economy perspective, this sets up already dominant technology companies like Amazon, Google, and Microsoft to also control AI (e.g., Jacobides et al., 2021, Burkhardt and Rieder, 2024), as they can build upon their vast compute resources, abundant data, access to users, and synergies with existing products to develop crucial infrastructure for themselves and third parties.

While this overall narrative indeed covers important parts of what is currently playing out, many conceptual and practical questions remain when we talk about AI 'as a platform'. This paper seeks to provide an overview and critical investigation into the different ways 'platformisation' (Helmond, 2015) proceeds in and around AI, paying specific attention to the 'general purpose potential' that undergirds most claims about economic—and by extension social and political—impact. Instead of taking this potential for granted, as simply inherent to a technology or set of technologies, I approach it as a technical and not-just-technical construct or achievement, as something that must be produced, requiring great effort and investment. Paying attention to how AI is composed and orchestrated also leads us away from the idea that is *a* technology, *a* platform. Neither AI nor even foundation models are singular, monolithic things that come with stable, clearly defined properties, even if Google's transformer architecture (Vaswani et al., 2017) currently captures attention and investment.

Although companies like Amazon and Meta increasingly use the moniker ‘artificial general intelligence’ (AGI) to describe their research and industrialisation efforts around foundation models, the question of whether their systems can imitate human intelligence is less relevant from a political economy perspective than their clear ambition to create widely deployable means of production that justify the truly colossal investments being made. To achieve this, AI providers—the companies that build AI platforms for others to use—seek not only to boost the performance of their systems in specific benchmarks, to improve resource efficiency, and so forth, but they also pursue breadth in terms of application potential and ease of specification when it comes to adapting to actual production tasks. ‘Traditional’ machine learning has already allowed for the use of the same hardware and software across many different types of domains: chips suited to the computational requirements of ML, whether they are produced by equipment manufacturers such as Nvidia or in-house, are largely task-agnostic and the same goes for software frameworks such as PyTorch and TensorFlow. The considerable breadth achievable based on these two layers explains why investments and the pursuit of economies of scale were already significant before the latest peak in AI hype. Here, platformisation manifests primarily as the emergence of *compute platforms* that hide technical complexities behind easy-to-use abstraction layers but still require much work when it comes to specification, as producing task-specific capabilities involves training models on often copious amounts of relevant examples that may be very costly to acquire or create.

To address customers who cannot afford or prefer not to pursue their own model training, the big cloud providers have progressively added *service platforms* on top of their compute stacks. These services provide models trained for tasks deemed sufficiently common (e.g., document classification, content moderation, machine vision, etc.) and/or sufficiently valuable (e.g., medical diagnosis, fraud detection, etc.) to justify the investment in domain-specific data and expertise. (Luitse, forthcoming) Compute platforms come with a high level of general-purpose potential and thus built-in economies of scale but remain at the lower end of AI value chains. Service platforms along the lines of what Amazon and Microsoft, in particular, are offering are much more complicated to construct, as each domain-specific system comes with its specific requirements, caveats, problems, and so forth, but reach up further in terms of value creation.

The promise of foundation models, at least to AI providers, is the ability to create what could be called *universal service platforms*. If Turing’s universal machine is capable of simulating any other symbol-manipulating device by virtue of being programmable, foundation models show the potential to be ‘adaptable’ to many downstream tasks through techniques like fine-tuning and prompting. While this metaphorical comparison only goes so far, it sheds light on why Meta ordered 350k Nvidia H100 GPUs in early 2024: foundation models promise both ‘universal’ breadth and forms of specification that are more lightweight and accessible than traditional data-heavy training. While the initial investments are colossal, a single model could be used to create a very large number of higher-level services, allowing providers to move up the value chain without having to train a separate model for every task domain they want to tackle. OpenAI, with the help of Microsoft’s money and compute park, is currently furthest along in this ‘third wave’ of platformisation in and of AI, providing several third-party interfaces (Chatbot, APIs, ‘custom’ GPTs, etc.) to access its underlying models.

At the same time, success is not guaranteed. The transformer architecture and the ‘train on all the data, then adapt’ logic driving foundation models are giving rise to impressive technical objects, but we can already see that the road to becoming a platform in a more commercial sense may well be more complicated, requiring—at least in part—the ‘domain work’ we know from more traditional AI service platforms. This includes collecting, cleaning, and licensing data, various interventions to ensure security and factuality, optimising for specific interaction patterns, and other things, but also forms of semantic and cultural engineering that are only partially captured with the term ‘alignment’, as Google recently had to learn when users asked its Gemini model to create images of German soldiers in 1943.

Within these three layers or vectors of platformisation, there is considerable variation, and there are many synergies between them that go far in explaining why the main actors in the field pursue all three simultaneously. Foundation models may, in fact, not succeed or succeed only partially as user-facing universal platforms but still play a central role in the creation of task- or domain-specific models (similar to the role of BERT in NLP). It is highly likely, however, that the diffusion of AI will not neatly resemble previous general-purpose technologies, such as electricity or steam power, as it reaches much further into culturally contested terrain, making fragmentation and diversification more likely. The goal of this paper is thus to better understand how AI is not becoming a platform but giving rise to different (kinds of) platforms that produce and explore its general-purpose potential in different ways, with potentially different effects on power structures. Insights into these variations are necessary to calibrate our political responses to complex and contradictory developments that do not teleologically converge on a single outcome.

References

Bogost I and Montfort N (2009) Platform Studies Frequently Questioned Answers. Available at: https://escholarship.org/uc/item/01r0k9br.

Bommasani R, Hudson DA, Adeli E, et al. (2021) On the Opportunities and Risks of Foundation Models. arXiv:2108.07258 [cs].

Eloundou T, Manning S, Mishkin P, et al. (2023) GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. arXiv:2303.10130.

Helmond A (2015) The Platformization of the Web: Making Web Data Platform Ready. *Social Media + Society* 1(2): 1–11. DOI: 10.1177/2056305115603080.

Jacobides MG, Brusoni S and Candelon F (2021) The evolutionary dynamics of the artificial intelligence ecosystem. *Strategy Science* 6(4): 412–435. DOI: 10.1287/stsc.2021.0148.

Luitse D (forthcoming) Platform Power in AI: The Evolution of Cloud Infrastructures in the Political Economy of Artificial Intelligence.

Narayan D (2022) Platform capitalism and cloud infrastructure: Theorizing a hyper-scalable computing regime. *Environment and Planning A: Economy and Space* 54(5): 911–929. DOI: 10.1177/0308518X221094028.

Rieder B and Burkhardt S (2023) Foundation models are platform models: Prompting and the political economy of AI. *Big Data & Society* 11(2): 1–15. DOI: 10.1177/20539517241247839.

Rochet J-C and Tirole J (2003) Platform Competition in Two-Sided Markets. *Journal of the European Economic Association* 1(4): 990–1029.

Trajtenberg M (2018) AI as the next GPT: A Political-Economy Perspective. NBER Working Paper 24245. Available at: http://www.nber.org/papers/w24245.

Vaswani A, Shazeer N, Parmar N, et al. (2017) Attention Is All You Need. arXiv:1706.03762.

HOW MACHINE VISION CHALLENGES STRUCTURE AI INNOVATION

Sam Hind
University of Manchester

Fernando van der Vlist
University of Amsterdam

Max Kanderske
University of Siegen

Introduction

Challenges, competitions and prizes have long played a role in driving technological innovation. In the late 1980s, the concept of 'Grand Challenges' emerged as a framework for realising research in science and technology. More precisely, Raj Reddy's 1988 Presidential Address to the Association for the Advancement of Artificial Intelligence (AAAI) aimed to propel AI research towards tangible, concrete applications. Chess-playing machines and autonomous vehicles were presented, amongst others, as 'bold national initiatives' intended to 'capture the imagination of the public' (Reddy, 1988: 18). Despite 25 years of financial support from the likes of the Defense Advanced Research Projects Agency (DARPA), National Science Foundation (NSF), and NASA, Reddy contended that AI research—in the US, at least—now needed to enter an 'era of accountability' (Reddy, 1988: 9).

In recent years, talk of the 'Grand Challenges' of AI has receded. In its place—as the era of accountability turns into an era of accumulation—a series of altogether less grand challenges. These might, instead, be understood as 'incremental challenges': a host of competitions organised by start-ups, research centres, and platform firms to facilitate cutting-edge innovation in AI, machine learning (ML) and machine vision more specifically (Hind et al., 2024). The argument here is that incremental challenges—different in scale, form, and purpose from Grand Challenges—serve as a critical organizing principle for the development of new ML and machine vision techniques (Ribes et al., 2019). Methodologically, challenges represent a fascinating setting for studying the everyday work of computer scientists working on ML model design, testing, and application.

Challenges As: Internal Competition, Collaborative Setting, Platformised Environment

Incremental challenges are an important mechanism through which (a) training data (b) computing power ('compute') and (c) expert forms of labour come together in the contemporary AI economy (Srnicek, 2022). However, competitions and challenges have long been used by capitalist firms to stimulate activity in different settings. Three trajectories are worth mentioning.

Firstly, capitalist firms often invite different kinds of internal competition, whether between rival departments, teams, or projects. Phillips and Rozworski (2019) discuss how US retailer Sears implemented an internal market to drive competition (rather than 'competitions', per se) between divisions. Similarly, messaging platform Telegram previously ran an internal competition to develop a simplified web version of the service (Telegram, 2021a, Contest, 2024), resulting in the public launch of two platforms, Version K and Version Z (Telegram, 2021b). Secondly, ML and machine vision challenges function as an opportunity to draw together expertise to tackle shared computational problems. The PASCAL VOC Challenge (2005–2012), for example, offered researchers the opportunity to design visual object-recognition methods in a competitive setting (Everingham et al., 2010). In such cases, the stated aim was to offer a 'standard evaluation methodology' for comparing different methods, and to 'measure the state of the art' in visual object-recognition (Everingham et al., 2010: 303). Thirdly, and most recently, the challenge format has itself become 'platformised' over the course of the last 10–15 years, with the likes of Kaggle (owned by Google/Alphabet since 2017) offering access to thousands of ML and data science challenges, across numerous domains such as medical imaging and plant recognition (Kaggle, 2024). In such cases, the competition format—already a way of standardising and comparing methods—is itself subject to the same conditions: with would-be competitors able to compare and evaluate different competitions across different domains, utilising different skill sets, with varying prize funds on offer.

Waymo Open Dataset Challenges

In March 2020, Google/Alphabet's autonomous vehicle division, Waymo, launched the 'Open Dataset Virtual Challenge', a soon-to-be annual competition designed to leverage their previously released Waymo Open Dataset (Anguelov, 2020). Composed of camera/lidar data captured by Waymo vehicles in various locations (Phoenix, San Francisco), the dataset constituted the largest, and arguably most diverse, autonomous vehicle dataset ever publicly released. The Waymo Open Dataset Challenge (the 'Virtual' was later dropped) has now run across four iterations (2020, 2021, 2022, 2023), each time inviting teams of computer scientists to tackle evolving ML and machine vision problems—using Google's data (Waymo Open Dataset) and tools (Google Cloud Storage, Google Colab, TensorFlow). Given three months, teams could select one of multiple specific ML/machine vision challenges to tackle, relating either to detection tasks or prediction tasks. A public leaderboard ranks each team's submission, and winners are invited to present their work at an associated workshop on autonomous driving (WAD), held at the annual CVPR (Computer Vision and Pattern Recognition) conference.

Through collaborative work on these Waymo challenges (Hind et al. 2024) we have encountered a number of key themes. Firstly, challenges serve as *interfaces* between platform firms, external collaborators, and ML tools and services. Secondly, that challenges increasingly drive *hyper-incrementality* in the development of ML/machine vision methods. Thirdly, *metrics* and *benchmarks* routinely evolve to help evaluate competition success, and rank technical performance. Fourthly, challenges engender a particular operating AI/ML *vernacular*, smoothing interpretation and understanding in, and across, competing teams. Then lastly, that participation in challenges is driven by the allure of the *applied domain*—helped along by the presence, and prestige, of platform firms such as Google/Alphabet.

Incremental Gains, Incremental Losses

What, then, might we understand as the wider effects or outcomes of the platformisation of ML/machine vision challenges? We can understand, rather straightforwardly, that the (incremental) gains offered through competitive formats such as the ML challenge are designed to be captured by the platform firms organizing them. Whether this is *directly* (through ML/machine vision techniques being directly incorporated into relevant AI products and systems), or *indirectly* (through fostering AI development through specific proprietary tools and software) is an empirical question. In any case, the desire is to secure a *competitive advantage* in the development of ML/machine vision techniques *within specific domains*. In Waymo's case, to secure a competitive advantage in the deployment of autonomous vehicles.

Inversely, what might the incremental *losses* be—and for whom? Everingham et al.'s (2015) reflection on the running of the PASCAL VOC Challenge suggests a few possible outcomes. Firstly, that (hyper-)incrementalism breeds (hyper-)incrementalism: teams work purely (or mainly) on building models/techniques that deliver just enough incremental gains to place higher than previous winners. The consequence of this, as Everingham et al. (2015) contend, is a move away from developing new models from scratch. Secondly, inevitably, that increasing monopoly status over developments in ML/machine vision by those who develop specific tools and software will lead to increased *dependence* on particular tools and software, and therefore on specific platform operators. Everingham et al. (2015) considered the development of 'out of the box' software for everything from ML training to validation to be extremely useful, from a technical perspective. In a platformised world, it looks increasingly problematic as the extractive/accumulative logics of Google/Alphabet run counter to principles of open science and collaboration.

References

Anguelov D (2020) Announcing Waymo's Open Dataset virtual challenges. Available at: https://waymo.com/blog/2020/03/announcing-waymos-open-dataset-challenges/.

Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A (2010) The PASCAL Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision* 88(2): 303–338. DOI: 10.1007/s11263-009-0275-4

Everingham M, Ali Eslami SM, Van Gool L, Williams CKI, Winn J, Zisserman A (2015) The PASCAL Visual Object Classes Challenge: A retrospective. *International Journal of Computer Vision* 111: 98–136. DOI: 10.1007/s11263-014-0733-5.

Contest (2024) Developer challenges. Available at: https://contest.com/.

Hind S, van der Vlist FN and Kanderske M (2024) Challenges as catalysts: How Waymo's Open Dataset Challenges shape AI development. *AI & Society: Knowledge, Culture and Communication*. DOI: 10.1007/s00146-024-01927-x.

Kaggle (2024) Kaggle Competitions. Available at: https://www.kaggle.com/competitions.

Phillips L and Rozworski M (2019) *People's Republic of Walmart: How the World's Biggest Corporations are Laying the Foundation for Socialism*. London: Verso.

Reddy R (1988) Foundations and grand challenges of artificial intelligence. *AI Magazine* 9(4): 9–21. https://ojs.aaai.org/index.php/aimagazine/article/download/950/868.

Ribes D, Hoffman AS, Slota SC, Bowker GC (2019) The logic of domains. *Social Studies of Science* 49(3): 281–309. DOI: 10.1177/0306312719849709

Srnicek N (2022) Data, compute, labor. In: Graham M, Ferrari F (eds) *Digital Work in the Planetary Market*. Cambridge MA: MIT Press, pp. 241–261.

Telegram (2021a) JavaScript Contest, stage 2. Available at: https://web.archive.org/web/20220929202333/https://contest.com/javascript-web-2.

Telegram (2021b) New Telegram Web Apps. Available at: https://web.archive.org/web/20220927215401/https://bugs.telegram.org/c/4002.