Original Research Article

Big AI: Cloud infrastructure dependence and the industrialisation of artificial intelligence

Fernando van der Vlist[1] (iD)**, Anne Helmond**[2] (iD)
and Fabian Ferrari[2] (iD)

Abstract

Critical scholars contend that 'There is no AI without Big Tech'. This study delves into the substantial role played by major technology conglomerates, including Amazon, Microsoft, and Google (Alphabet), in the 'industrialisation of artificial intelligence'. This concept encapsulates the shift of AI technologies from the research and development stage to practical, real-world applications across diverse industry sectors, resulting in new dependencies and associated investments. We employ the term 'Big AI' to encapsulate the structural convergence of AI and Big Tech, characterised by the profound interdependence of AI with the infrastructure, resources, and investments of these major technology companies. Using a 'technographic' approach, our study scrutinises the infrastructural support and investments of Big Tech in the AI sector, focussing on corporate partnerships, acquisitions, and financial investments. Additionally, we conduct a detailed examination of the complete spectrum of cloud platform products and services offered by Amazon, Microsoft, and Google. We demonstrate that AI is not merely an abstract idea but an actual technology stack encompassing infrastructure, models, applications, and an ecosystem of applications and companies relying on this stack. Significantly, these tech giants have seamlessly integrated all three components of the stack into their cloud offerings. Furthermore, they have developed industry-focussed solutions and marketplaces aimed at attracting third-party developers and businesses, fostering the growth of a broader AI ecosystem. This analysis underscores the intricate interdependence between AI and cloud infrastructure, emphasising the industry-specific aspects of cloud AI.

Keywords

Cloud infrastructure, artificial intelligence, Big Tech, AI industrialisation, AI ecosystems, platform studies

Introduction: No AI without Big Tech

The ongoing competition among major technology companies in cloud-based artificial intelligence (AI), known as the 'cloud AI wars' (Goldman, 2022), is gaining momentum. Industry leaders proclaim that we are in a transformational era, where AI technologies are profoundly reshaping society. Influential figures like Bill Gates (2023) and Alphabet's Sundar Pichai emphasise the foundational and transformative nature of AI (Manyika, 2023). However, this alleged transformation is primarily driven by a select few technology giants, specifically Amazon, Microsoft, and Google. Their unparalleled ability to rapidly scale—known as 'hyper-scalability'—is attributed to their cloud computing arrangements, which have established the fundamental socio-technical infrastructure driving expansive growth and platform capitalism (Narayan, 2022).

This article delves into the substantial implications of these firms' dominance in what we term the 'industrialisation

of AI'. By using this term, we refer to the transition process of AI systems from the research and development phase to becoming practical, real-world commercial products and services in a variety of industry sectors. This transition results in new dependencies on cloud infrastructure and the corresponding investments in computational resources required for the industrial-scale development and deployment of AI applications and solutions. As we explain, the story of AI's

[1]Media Studies, University of Amsterdam, Amsterdam, the Netherlands
[2]Department of Media and Culture Studies, Utrecht University, the Netherlands

Corresponding author:
F. N. (Fernando) van der Vlist, Media Studies, University of Amsterdam, Turfdraagsterpad 9, 1012 XT, Amsterdam, the Netherlands.
Email: f.n.vandervlist@uva.nl

industrialisation is intimately tied to cloud infrastructure dominance. Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP), as the top-three cloud platforms, form the essential backbone of this continuing industrialisation process. These platforms, 'where the Internet lives' (Holt and Vonderau, 2015), have a profound impact on businesses across industries. When a major cloud platform like AWS experiences an outage, as it did on June 13, 2023, it has far-reaching repercussions for a wide array of clients, including the Associated Press (AP), McDonald's, and Reddit. These clients faced disruptions due to their reliance on AWS (Ropek, 2023). According to market research estimates, AWS holds a particularly dominant position, being the primary operating system of the Internet with approximately a third (32%) of all cloud services running on it, followed by Microsoft Azure (22%) and Google Cloud (11%) (Richter, 2023). The extensive suites of cloud products and services offered by these companies contribute significantly to their revenues (Klinge et al., 2023).

Discursively, the term 'AI' serves as a powerful marketing tool, attracting substantial investments and pushing AI startups to seek partnerships with Amazon, Microsoft, and Google (Alphabet), who in turn 'pour billions' into costly cloud computing (Hodgson, 2023). These major technology companies actively position themselves as essential infrastructure providers through their support and investments. With AI entering its 'industrial age' (*The Economist*, 2022), it is becoming increasingly important to understand the value chains of AI in the digital platform economy for strategic, political, and economic reasons.

Therefore, it is crucial to theorise and closely monitor the industrialisation of AI, a phenomenon characterised by its pervasive expansion and integration across industries and sectors, extending far beyond traditional computing. This process mirrors the 'platformisation' observed in various sectors and industries (Van Dijck et al., 2018). Currently, the global AI market is experiencing rapid growth, but only a handful of major technology companies exercise control over the dominant infrastructure platforms. They have been referred to as the 'Great Houses of AI' (Crawford, 2021: 20) and play a foundational role in the development and deployment of AI. As they lead the 'pursuit of AI dominance', concerns have risen about potential risks associated with 'industrial capture', as highlighted by journalists (*The Economist*, 2023a; Murgia, 2023). The dominance of these companies is intrinsically linked to their control over infrastructure, which is crucial for AI development and deployment, raising concerns regarding access and competition (Khan, 2023). Their structural dominance is attributed to three key advantages: access to vast amounts of data, substantial computational resources, and a geopolitical edge (Kak and Myers West, 2023; Whittaker, 2021). As Kak and Myers West assert, 'There is no AI without Big Tech' (2023: 5).

This research aims to enhance our understanding of the advantages and ramifications of Big Tech's cloud infrastructure dominance and smaller players' dependency on cloud infrastructure. It delves into the convergence of AI and Big Tech, emphasising the large-scale industrialisation of AI, facilitated by Big Tech's hyper-scalable cloud computing arrangements (cf. Narayan, 2022). This raises the questions: *How can we theorise and investigate the role of Big Tech in the industrialisation of artificial intelligence? What are the multifaceted implications associated with (large-scale) reliance on cloud infrastructure?*

To capture the scale and scope of this structural convergence, we introduce the concept of 'Big AI', characterised by the interdependence between AI and the infrastructure, resources, and investments of major technology conglomerates such as Amazon, Microsoft, and Google (Alphabet). This structural dependency underlies the ongoing industrialisation of AI, aligning with existing critiques of Big Tech's AI dominance (Kak and Myers West, 2023), which our empirical analysis further substantiates. While it is not the only conceivable path for the future of AI, the continuous provision of essential infrastructure services by Amazon, Microsoft, and Google positions them to benefit from the widespread expansion of AI across industries and sectors.

We explore the core components of 'Big AI' empirically. This includes different forms of support and investment as well as the cloud platform offerings from Amazon, Microsoft, and Google. Our comprehensive approach provides unique insights into the current state and evolution of Big AI. We gain a deep understanding of AI as both a product and a service category, a distinct technology stack and ecosystem, and an integral and mundane component of existing cloud computing arrangements. Furthermore, we highlight the developmental and deployment aspects of the supposed 'AI revolution' (e.g. Murgia, 2023). This illustrates the substantial role that Amazon, Microsoft, and Google play in 'convening' enterprises, organisations, and developers to participate in the creation, capture, and commercialisation of AI (cf. Egliston and Carter, 2022; Helmond et al., 2019).

The following section places our research in the context of existing critical literature regarding AI and machine learning, Big Tech, and cloud computing. We then outline our two-fold 'technographic' approach to further explore the convergence of AI and Big Tech. This includes initially analysing Big Tech's infrastructural support and investments in the AI sector and then conducting a comprehensive examination of all their cloud platform products and services. This approach allows us to uncover the dual role of the cloud as both a fundamental infrastructure and a marketplace for AI applications and solutions. When considering these dimensions together, we gain unique insights into the distinct characteristics of AI's industrialisation and the implications of dependence on (large-scale) cloud infrastructure.

AI as an existing technology stack and ecosystem

'AI' is a nebulous concept, characterised by an ambiguous and fluctuating meaning. Crawford (2021: 217) highlights the importance of 'how we define AI, what its boundaries are, and who determines them', shaping our perceptions and framing what can be contested and debated. In the emerging field of critical AI studies, Raley and Rhee (2023: 188–189) suggests viewing AI as 'an assemblage of technological arrangements and socio-technical practices, as concept, ideology, and dispositif' while staying 'attuned to actually existing socio-technical systems'. Jacobides et al. (2021) and Rieder (2022) similarly advocate for understanding AI as a large technical system, emphasising attention to its technical specifics and materialities to comprehend its operations and modes of power.

Aligned with these perspectives, we conceptualise AI as an *existing technology stack and (emerging) ecosystem* while also recognising its role as a discursive phenomenon (i.e. the politics of the term 'AI', cf. Gillespie, 2010). The AI technology stack includes infrastructure, models, and applications, along with the respective providers of these components. Infrastructure encompasses both cloud software and hardware platforms, with their manufacturers viewed as significant beneficiaries in the AI market. Drawing from the existing digital platform theory (Jacobides et al., 2021; de Reuver et al., 2018; Van der Vlist, 2022), the broader ecosystem of the AI stack consists of applications and solutions typically created by third-party developers and businesses. These elements complement the cloud platform stack technically or represent a collection of firms and organisations collectively engaged with its components from an organisational viewpoint. These AI stack components are associated with specific 'sectoral and national systems of firms and institutions that collectively engage in [AI]', evolving differently across China, the USA, and the EU, with 'a handful of firms building global AI ecosystems' (Jacobides et al., 2021: 412). The variations within these national or sectoral systems underscore the political economy of AI, its intricate supply chains, dependencies, and the role of third parties participating in the creation, capture, and commercialisation of AI.

Existing critical research has illuminated the intricate relationship between AI, machine learning, cloud computing, and major technology companies, emphasising the significant influence and power of Big Tech and platform companies (Whittaker et al., 2018; Whittaker, 2021). Based on this literature, three main drivers contributing to industry concentration within the AI sector can be identified: the integration of proprietary hardware and software components within platform ecosystems (Crawford, 2021; Van Dijck et al., 2018; Van der Vlist, 2022; Mackenzie, 2019; Rieder, 2022), the expansion of AI into various industries through products and services (Ferrari and McKelvey, 2022; Lehdonvirta, 2022; Luitse and Denkena, 2021), and the facilitation of AI value chains (Ferrari, 2023; Widder et al., 2023).

Concerning the first driver of industry concentration in the AI sector, the profound integration of hardware and software components, coupled with external software developers relying on cloud computing platforms and hardware, establishes Big Tech as a crucial infrastructure service provider in the digital platform economy (Jacobides et al., 2021; Narayan, 2023). Cloud platforms and their marketplaces play a central role in AI development and integration by offering infrastructure-as-a-service (e.g. compute capacity) and commodifying digital products (e.g. pre-trained models or software packages).

Critical scholars have demonstrated how platform companies extend their reach, influence, and user base by engaging third-party developers and businesses in their ecosystems. This is achieved by promoting the use of proprietary developer tools and infrastructure to create complementary apps and services (Egliston and Carter, 2022; Helmond et al., 2019; Jacobides et al., 2021; Narayan, 2023; Van der Vlist and Helmond, 2021). Facilitating technical integrations with the platform's ecosystem and providing consulting and implementation services enable this strategy, emphasising the necessity of studying the expansion and hyper-scalability of platforms on both business and infrastructure levels (Narayan, 2022; Van der Vlist and Helmond, 2021).

Second, the integration of AI into existing socio-technical arrangements signifies its infrastructuralisation (Dyer-Witheford et al., 2019; Jacobides et al., 2021). While Big Tech companies advocate for the 'democratisation of AI' by making AI tools accessible to a broad audience, developers often, perhaps unknowingly, contribute to Big Tech's infrastructural objectives. They do so by building applications and integrations using the infrastructure provided by these tech giants, assuming they are shaping the future of AI (Burkhardt, 2019; Dyer-Witheford et al., 2019; Luchs et al., 2023).

In response, scholars propose 'infrastructural thinking' and an 'infrastructural analysis of AI' to understand its political economy. This approach recognises the diverse socio-technical systems necessary for AI to operate globally and underscores the need to trace the 'full stack supply chain' and production networks of AI systems from various disciplinary perspectives (Crawford, 2021; Dyer-Witheford et al., 2019; Ferrari, 2023; Rella, 2023; Whittaker et al., 2018; Widder and Nafus, 2023). Despite discursive claims of AI being 'open' (Widder et al., 2023), it fundamentally relies on computational resources for both its initial training and ongoing operations, challenging the notion of 'openness' as it might obscure underlying infrastructure dependencies.

Lastly, the development and deployment of AI and large-scale machine-learning models have created intricate dependencies among companies within the AI value

chain. As Küspert et al. (2023) explain, this integration not only gives rise to new product categories but also results in complex supply chains and dependencies among AI stakeholders, influencing broader social and economic processes. They emphasise the complex lifecycle of general-purpose AI (GPAI) models, involving various actors responsible for different stages.

In our study, we illustrate how this interdependence operates on multiple levels, leading to strategic alliances, partnerships, and investments. These collaborations involve upstream providers of models and infrastructure, as well as downstream companies integrating these models into consumer-facing products and services. This reflects a broader trend of 'corporate financialization', strengthening Big Tech's infrastructural dominance within the economy and society (Klinge et al., 2023).

A technography of cloud AI

To explore Big Tech's role in the industrialisation of artificial intelligence and the implications of cloud infrastructure dependency, we employ methodologies drawn from critical platform and algorithm studies, with a particular emphasis on 'technography'. This descriptive and interpretive approach critically analyses the structural and operational aspects of technical systems (Bucher, 2016; Mackenzie, 2019; Van der Vlist et al., 2022). Technography encourages a detailed examination of the material aspects of technology by directly reading various publicly available documents generated by and related to technical systems. These documents include technical documentation, developer and product websites, financial reports, media articles, press releases, and blog posts that provide detailed insights into the intended objectives, functions, and capabilities of technical systems. This enables us to explain specific functionalities or aspects of the technology while maintaining a critical stance toward promotional language or industry jargon found in these materials.

We start by scoping Big Tech's broader infrastructural support and investments in AI, which entails exploring their partnerships, acquisitions, and financial investments. This approach allows us to identify prominent examples illustrating the forms of infrastructural support and investments foundational to the industrialisation of AI. Amazon, Microsoft, and Google are our focal points due to their predominant market presence and significant influence in the field (Richter, 2023). While acknowledging both shared and unique features of their cloud platforms, our goal is not a direct platform-to-platform comparison.

Existing critical platform studies have identified three primary drivers of expansion and integration: consolidation through acquisitions, capturing and convening via partnerships, and expansion facilitated by software development kits (SDKs) and application programming interfaces (APIs) (Egliston and Carter, 2022; Helmond et al., 2019;

Van der Vlist, 2022; Van der Vlist and Helmond, 2021). We explore how these strategies of infrastructural support and investment by Big Tech play a pivotal role in driving the industrialisation of AI. The technographic approach draws from various publicly accessible documents, including corporate blogs, promotional materials on company websites tailored to AI startups, developers, and businesses (e.g. Microsoft for Startups), trade reports, as well as business information databases like Crunchbase and Tracxn.

Following this, we conduct an in-depth examination of the complete suite of cloud platform services offered by Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP). We categorise the diverse service offerings, uncover interrelationships among them, and initially explore solutions offered by third-party vendors on their respective online cloud marketplaces.[1] This second step yields valuable insights into how AI and machine learning are integrated within their cloud infrastructure setups and shows the intermediary role of their marketplaces that connect various industry sectors. This complements our document analysis in the first step.

Existing research offers guidance on employing technographic approaches to scrutinise the composition, governance, and evolution of these companies' development platforms, as well as the role and integration of specific elements within their stacks or ecosystems (Van der Vlist et al., 2022). In this context, platforms are viewed not as monolithic entities but as dynamic 'service assemblages' that adapt to developers' needs while shaping the economic dynamics of competition and monopolisation (Blanke and Pybus, 2020). This dynamic property extends to cloud platforms, where the 'AI as a service' ('AIaaS') business model exemplifies the simultaneous decomposition and integration of AI products and services across various industry sectors. Therefore, gaining insights into how cloud platforms have evolved into such service assemblages and understanding the pivotal role of developers in expanding and integrating AI across sectors are essential for exploring the multifaceted implications of reliance on cloud infrastructure.

The technographic approach relies on publicly accessible platform documentation, including comprehensive catalogues of cloud offerings, detailed product information, and technical specifications. These resources are primarily tailored for external (third-party) software developers and enterprises, offering guidance to individual and corporate software developers in implementing cloud platform tools, products, and services within their respective organisations. Thus, the technical documentation serves a purpose beyond mere marketing rhetoric, providing practical information about the platform's capabilities and integration processes. Any discrepancies or inaccuracies could lead to significant practical issues in development, necessitating a high level of reliability and detail that goes beyond the scope of promotional language.

To collect the individual product pages for each cloud platform, we developed custom web scrapers designed to capture all product pages linked from their comprehensive product directories.[2] This process also encompassed their marketplaces and industry-focussed subdirectories, all linked from their product directories.[3] Using this documentation, we manually categorised the diverse types of cloud offerings based on their existing product categories and subcategories across the three cloud platforms. Subsequently, we established connections between these offerings by detecting explicit mentions of product names across the corpus, including on-page hyperlinks (while excluding headers, navigation menus, and footers), using case-sensitive grep searches. These connections were then visually represented as network diagrams using visualisation and illustration software, showcasing the mentions or citations of these products within the corpus. Additionally, we employed this documentation to gain insights into how products are positioned and marketed to specific industry sectors, achieved by detecting mentions of product names in industry-specific subdirectories and counting the number of solutions available in each existing marketplace category and subcategory.

This two-fold 'technographic' approach enables us to critically analyse the structural elements underpinning the three dominant cloud infrastructure arrangements. It allows us to pinpoint the pivotal role played by AI and machine learning within these structures, offering insights into Big Tech's multifaceted involvement in the industrialisation of AI and the consequential implications tied to reliance on cloud infrastructure.

Big Tech's infrastructural support and investments in AI

Our analysis first delves into the crucial role of infrastructural support and investments in advancing the industrialisation of AI, focussing on corporate partnerships, acquisitions, and financial investments by Amazon, Microsoft, and Google (Alphabet). While AI initiatives like OpenAI and Stability AI gain attention, they heavily depend on substantial support and investment from the Big Tech. These companies have forged lasting partnerships, made significant investments, and expanded their influence by acquiring various AI companies and initiatives (Murgia, 2023). The financial and infrastructural requirements for scaling AI models and applications have led many AI firms to form exclusive or preferred cloud partnerships with Big Tech.

The investment activities of these Big Three companies are substantial. Data from Crunchbase and Tracxn shows Microsoft's involvement in approximately 211 investments and 214 acquisitions, with over $188 billion in acquisition spending—surpassing its counterparts. Google's activity includes 238 investments and 260 acquisitions, totalling an expenditure of $41.8 billion. Amazon's involvement is similarly substantial, with 132 investments and 99 acquisitions, amounting to an acquisition cost of $36.9 billion.

Microsoft's financial investments, acquisitions, partnership programmes, and startup funds cement its central role in AI development and deployment across industries. For example, Microsoft's long-term partnership with OpenAI follows a multi-year, $11 billion investment. Azure, Microsoft's cloud platform, is OpenAI's exclusive cloud provider, offering AI-optimised infrastructure and tools for their products and API services. OpenAI's models, including GPT-3, Codex, and Embeddings, are accessible through Azure's OpenAI Service, enabling enterprise developers to build AI applications on Azure (Microsoft Corporate Blogs, 2023). Microsoft also holds an exclusive licence to incorporate OpenAI's models into its own products and services, such as Microsoft 365, Edge browser, Bing search engine, GitHub Copilot, Power Platform, and Microsoft Designer. OpenAI's GPT-3 models are available through Microsoft's Azure Marketplace for integration into enterprise apps on Azure.[4]

In addition to OpenAI, Microsoft extends these integrations to its partners through the Cloud Partner Program and AI Partners (Dezen, 2023; Microsoft, 2023a). These partner programmes help Microsoft form strategic alliances with firms and organisations, assisting clients in incorporating advanced tools and services into their custom software and products. Microsoft's cloud and AI partners are instrumental in helping enterprise customers build AI and cloud-based software solutions in various industries.

Beyond OpenAI, Microsoft has established partnerships with companies like Wayve (deep learning for autonomous mobility), PrimerAI (intelligence and operational workflows), Novo Nordisk and Paige (AI-based diagnostics and drug development), and collaborates with NVIDIA to build AI-optimised supercomputers. These collaborations emphasise Microsoft's reliance on essential infrastructure providers within the stack, including GPU manufacturers. Microsoft's role in backing AI startups is evident through programmes like the AI Grant programme and Startups Founders Hub. These programmes provide investments, Azure cloud credits, technical advice, development tools, and other resources (Microsoft, 2023b).

Microsoft leads the Big Three in terms of investments made in AI-related firms (*The Economist*, 2023a). Beyond OpenAI, Microsoft has invested in AI startups like D-Matrix (efficient computing for data centres), Noble.AI (R&D technologies), and KudoAI (live speech translations, available on Microsoft Teams). The acquisitions of AI startups, such as Suplari (a corporate spending analysis tool) and Nuance Communications (acquired for $19.7 billion), demonstrate Microsoft's strategy to leverage its AI services across various industry sectors and establish itself as the primary infrastructure provider for enterprises.

Google and Amazon have adopted similar strategies to solidify their positions in AI development and deployment. In 2014, Google acquired DeepMind, specialising in deep learning algorithms and neural networks across multiple fields. This acquisition has enhanced Google's AI capabilities and strengthened its position in the AI landscape. Google has formed partnerships with AI-focussed companies like Anthropic, AI21Labs, Midjourney, Osmo, and Cohere, making Google their preferred cloud provider for training models using Google's optimised Cloud Tensor Processing Units (TPUs) processing chips (Ichhpurani, 2023). Cohere's language AI models are available through the Google Cloud Marketplace (Cohere Team, 2022). Google's partnership with Adobe integrates external AI into its own AI products, like Firefly for Bard, Google's conversational AI service (Greenfield, 2023). The expanded strategic collaboration between Salesforce and Google integrates Salesforce Data Cloud with Google's 'Vertex AI', its core AI platform, and 'BigQuery', a fully managed enterprise data warehouse with built-in machine-learning capabilities. This integration offers their customers 'seamless data access across platforms and across clouds, akin to having their data housed in a single location' essentially bringing 'two very large ecosystems of data together' and enabling customers to bring their AI models trained in Vertex AI into the Salesforce platform (Salesforce, 2023; Wiggers, 2023). Google's collaboration with NVIDIA to create AI-optimised hardware instances mirrors Microsoft's approach, emphasising the importance of strategic alliances in AI advancements (Ichhpurani, 2023).

Google's AI programmes and partnership initiatives are designed to increase accessibility to its infrastructure and AI capabilities for partners at all levels of the cloud AI stack. This includes chipmakers, developers, and consulting firms, who are provided access to Google's infrastructure, AI products, and so-called 'foundation models'. This accessibility allows enterprise customers to scale their technology implementations with support from specialised consulting firms, developers, and technology partners proficient in machine-learning model development and deployment (Ichhpurani, 2023). Google also collaborates with companies building foundation models, AI platforms, and generative AI application developers. These partnerships foster customer use-case creation and drive innovation across various industries (Ichhpurani, 2023). Furthermore, these technology partners are available on the Google Cloud Marketplace, offering their services on the Google Cloud platform to address the specific needs of enterprise-level businesses. Additionally, Google has expanded its Startups Cloud Program to include a dedicated AI programme for AI-first startups, providing cloud credits, technical training, and other resources to advance their AI projects (Yang and Gokturk, 2023).

Amazon has partnerships with Stability AI and Hugging Face, designating Amazon as their preferred cloud partner.

These companies use Amazon's cloud machine-learning platform, 'SageMaker', to develop and train models, with pre-trained machine-learning models available on the AWS Marketplace for deployment on SageMaker. Amazon's Bedrock AI platform provides access to a range of models, including Amazon's Titan foundation models and models from other companies, such as Stability AI's Stable Diffusion, Anthropic's Claude, and AI21 Lab's Jurassic-2 (Sivasubramanian, 2023). AWS Partners assist businesses with their AI and infrastructure needs.[5] While Amazon does not directly invest in many AI startups, it runs the AWS Generative AI Accelerator programme for startups, offering credits, technical mentors, and networking opportunities with investors and enterprise customers (AWS Startup Loft, n.d.).

In summary, partnerships, acquisitions, and financial investments are propelling the industrialisation of AI, offering advanced infrastructural support, expertise, and resources. This accelerates AI development and integration across industries, with Amazon, Microsoft, and Google establishing themselves as premier cloud providers through strategic initiatives. As the AI market evolves, these companies actively foster partnerships and industry alliances via dedicated AI programmes, marking the next stage in the evolution of the AI ecosystem (Helmond et al., 2019; Jacobides et al., 2021).

Mapping cloud AI: Infrastructure, models, and applications

Our technographic focus now shifts to the features of cloud infrastructure services provided by Amazon, Microsoft, and Google across their extensive product and service offerings. Notably, Big Tech companies have seamlessly incorporated all three components of the cloud AI stack—infrastructure, models, and applications—into their cloud services. They also offer industry-specific solutions and marketplaces aimed at attracting third-party developers and businesses, fostering expansive ecosystems.

Our technographic analysis unveils a diverse range of both shared and distinctive attributes defining cloud AI across each platform's offerings. Moreover, we gain valuable insights into the seamless integration of AI and machine learning within their respective cloud platform stacks, revealing connections to various products, infrastructure elements, and industry sectors. This subsequent analysis underscores the profound interdependence between AI and cloud infrastructure, highlighting the industry-specific nature of cloud AI.

Surveying cloud offerings

The technical documentation associated with AWS, Microsoft Azure, and GCP's extensive cloud offerings

reveals their perspectives on AI and machine learning and their connections to other elements of cloud infrastructure.

Some cloud products and services include terms such as 'AI', 'ML', and 'Deep Learning' in their titles or descriptions, particularly those in the AI and machine-learning category,[6] such as Microsoft's 'Azure Machine Learning' and 'Azure AI Content Safety', Amazon's 'AWS Deep Learning AMIs', and Google's 'Document AI' and 'Vision AI'. However, many titles describe the service's intended purpose without explicitly using these terms, like AWS's 'Fraud Detector' and 'Transcribe', Azure's 'Bot Services' and 'Speech to text', and GCP's 'Recommender'. The category of AI and machine learning also includes hardware infrastructure services like Google's 'Cloud GPUs', 'Cloud TPUs', and 'Deep Learning VM Image', offering processing units and preconfigured virtual machines for machine-learning applications. This demonstrates that AI and machine learning are not isolated services but rather essential components integrated across various categories, highlighting the need for a comprehensive perspective.

The extensive cloud offerings from Amazon, Microsoft, and Google can be classified into five main categories with nineteen subcategories {Figure 1}. These classifications encompass a total of 852 distinct cloud offerings.

First, the Application Development and Integration category (273 offerings) focusses on developer tools, application integration, and management tools. Notable examples include 'AWS Elastic Beanstalk', a fully managed service for deploying and running applications, Google's 'Cloud Functions', a serverless execution environment for building and connecting cloud services, and Azure's 'Logic Apps', a cloud service for automating workflows and integrating systems. Second, the Computing and Infrastructure category (240 offerings) encompasses products and services such as compute, containers, storage, serverless computing, databases, and networking. Notable offerings include 'Amazon Elastic Compute Cloud' (EC2), Azure's 'Virtual Machines', and 'Google Kubernetes Engine' (GKE). The Security and Compliance category (88 offerings) covers services related to security, identity and access management, compliance, and cryptography. Key examples include 'AWS Identity and Access Management', 'Azure Active Directory', and GCP's 'Identity and Access Management'. Additionally, the Industry-Specific Solutions category (82 offerings) offers tailored products for industries like financial services, healthcare, media and gaming, and the Internet of Things (IoT). Examples include Google's 'Cloud Healthcare API' for healthcare data management and Azure 'Media Services' for encoding, streaming, and protecting media content. The Other category (54 offerings) covers services such as 'AWS Ground Station' for satellite communication and control, 'AWS RoboMaker' for robotic applications, and 'Azure Lab Services', which allows users to set up and manage cloud-based virtual machine labs for teaching, training, and testing purposes.

Lastly, the Data and Analytics category (115 offerings) focusses on products and services related to data analytics, data management, machine learning, and AI solutions. The Artificial Intelligence (AI) and Machine Learning (ML) subcategory (86 offerings) contains the majority of products and services in this category. Among these offerings, notable examples from Amazon include 'Amazon SageMaker', for building, training, and deploying machine-learning models at scale; 'Rekognition' for image and video analysis for facial recognition and object detection; and 'Comprehend' for natural language processing, sentiment analysis, and entity recognition. Microsoft contributes with 'Azure Machine Learning', a comprehensive platform for developers and data scientists for building, deploying, and managing machine-learning models; 'Cognitive Services', for pre-trained AI models and APIs that enable developers to add intelligent capabilities like vision, speech, and language understanding to their own applications; and 'Azure Databricks' for a unified workspace for Big Data and machine learning. Google offers notable products and services like 'Vertex AI', for building, deploying, and managing machine-learning models; 'TensorFlow Enterprise', for machine learning in enterprise environments; 'AutoML' for developing models with limited machine-learning expertise; and 'Dialogflow' for integrating natural language understanding in conversational agents. These examples demonstrate the diverse range of AI and machine-learning solutions within the subcategory.

While AWS, Microsoft Azure, and GCP offer a diverse array of services, there are notable similarities in the types and titles of these services across their cloud platforms, highlighting common essential elements of cloud infrastructure. Despite each platform encompassing a broad spectrum of service categories, Microsoft and Google distinguish themselves with offerings in Computing and Infrastructure, showcasing their expertise in this domain. Conversely, Amazon leads in the Application Development and Integration category, and Google leads in the Security and Compliance category. These focusses reflect their individual strengths and strategic priorities.

In conclusion, AI and machine learning are not isolated products or services; they are integral components of a broader technology stack and ecosystem of cloud infrastructure tools, products, and services. These diverse cloud offerings form the foundation for developing and deploying AI applications and solutions across enterprises in various industries, each with its unique requirements and objectives.

AI and cloud platforms

The extensive documentation on these tools, products, and services reveals the complex interconnections between
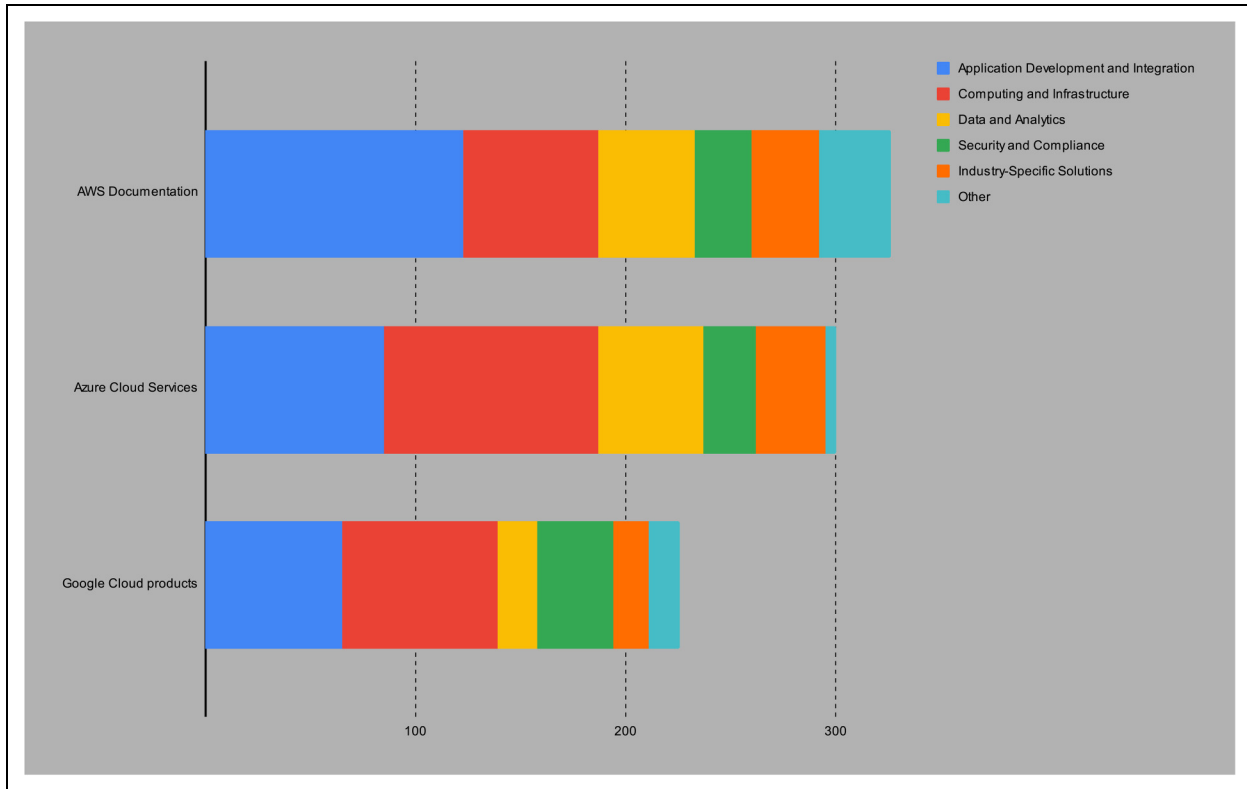
Figure 1. Classification of product and service offerings from Amazon Web Services, Microsoft Azure, and Google Cloud Platform shown by main category. Some products and services may appear in multiple categories.

various integral components of cloud platforms. As noted earlier, the comprehensive public technical platform documentation from these companies serves as an exhaustive catalogue. The documentation provides detailed specifications for each offering, with frequent technical citations and references to related offerings, aiding enterprise software developers in understanding and implementing these diverse services. For instance, GCP's 'Vertex AI' documentation details integration options and functionalities, such as exporting datasets from 'BigQuery', creating custom machine-learning models with 'BigQuery ML', using the integrated development environment in 'Vertex AI Workbench', generating data labels through 'Vertex Data Labeling', and optimising training time and costs with 'AI Infrastructure'. Similar structures are found in Azure and AWS's platform documentation. Analysing these citations and references within each product's documentation offers a comprehensive view and understanding of the AWS, Azure, and GCP cloud platform ecosystems.

Our analysis shows significant interconnectedness among elements within each stack (Figure 2). In the cases of Microsoft and Google, their AI platforms serve as the central pillars of their respective stacks. These stacks consist of multiple clusters that focus on specific elements related to infrastructure, including hardware and GPUs. Other clusters are centred around process-oriented

components, such as pattern recognition, data labelling, and optimisation. Additionally, there are dedicated clusters catering to enterprise solutions, highlighting their capacity to meet various industry requirements.

By applying colour-coding based on the frequency of 'AI' and machine-learning mentions (AI/ML) in product documentation pages,[7] we can effectively identify products (nodes) and clusters associated with AI/ML. This generates a visual 'heatmap' overlay that highlights the density of AI/ML mentions per offering, showing the distribution and prominence of AI/ML across the stack (Figure 2). This approach extends beyond products explicitly categorised as AI or machine learning, capturing any offerings that incorporate AI/ML terminology in their descriptions or technical specifications. In Azure's documentation, approximately 36.8% of their products mention AI/ML, indicating a significant presence. Conversely, AWS shows a lower percentage at 9%, while GCP stands out with a substantial 49.2% of their products referencing AI/ML.

In GCP's stack, the central network cluster is prominently centred around AI/ML, featuring products such as 'AI Infrastructure', 'Cloud Storage', 'Vertex AI', and 'BigQuery'. Among these, BigQuery plays a central role, being referenced frequently by many other products and containing a high number of AI/ML references in its

Figure 2. Cloud AI stacks: Citation networks illustrating the structural interconnections between cloud platform products and services for Amazon Web Services, Microsoft Azure, and Google Cloud Platform. The thickness of the lines represents the frequency of citations, indicating the strength of connections between different products and services. Colour scale reflects frequency of 'AI' and machine-learning mentions. Layout: Force Atlas 2; node scaling: by indegree (mention) count.

pages. Azure's stack revolves around their 'Applied AI services', 'Machine Learning', 'Cognitive Services', and 'Azure OpenAI', indicating their focus on AI-related solutions and services.

In AWS's central cluster, we observe a high frequency of AI/ML keywords, not only in products typically associated with AI/ML like 'SageMaker', but also in products that may not immediately seem connected to AI/ML. For example, 'AWS Solutions', a directory of vetted solutions and guidance for business and technical use cases, and 'AWS Quick Starts' (now 'AWS Partner Solutions'), a directory of automated deployment solutions, both feature a

notable presence of AI/ML keywords. Additionally, 'AWS Professional Services' offers cloud consultancy for enterprise customers, and 'AWS Training and Certification' provides training and certification programmes for aspiring AWS developers. Furthermore, 'AWS Deepracer' and 'AWS DeepComposer' offer developers a hands-on way to learn about machine learning, including using a musical keyboard. The prevalence of AI/ML keywords in these products demonstrates how Amazon not only focusses on dedicated AI/ML offerings but also integrates AI/ML elements across a broader range of products and services, including educational resources materials. This highlights the importance

of knowledge acquisition and dissemination through official training and certifications programmes, aimed at growing the developer community and expanding the ecosystem. It also underscores the important role of Amazon's play in assisting enterprises with the creation of their AI-driven products and services on Amazon's cloud platform.

Industry-specific and marketplace solutions

Furthermore, we identified various specialised cloud offerings that cater to specific industry sectors with unique requirements. The documentation includes dedicated pages covering these industry-specific tools, products, and services. These target industries are diverse and include Government and Public Sector, Education, Healthcare, Gaming, Media and Entertainment, Advertising and Marketing, Energy, Manufacturing, Automotive, Retail, Financial Services, and Supply Chain and Logistics. By applying the same approach of detecting product citations within these industry-specific documentation pages, we can uncover associations between various products and industry sectors. This approach also offers another perspective on the interconnected nature of these cloud AI stacks.

The interconnectedness of the citation network illustrates the common industries targeted by these cloud platforms (Figure 3). For example, GCP offers healthcare-focussed products like the 'Cloud Healthcare API', 'Medical Imaging Suite', and 'Healthcare Natural Language AI' to facilitate AI solutions for medical diagnostics and data analysis. Azure provides 'Health Data Services', 'Azure Data Lake Storage', and a 'Health Bot' for developing AI-powered virtual healthcare assistants, while AWS offers 'AWS Health', 'HealthLake' for transforming medical and insurance data using machine-learning models.

In the retail sector, solutions such as 'Discovery AI for Retail' and 'Recommendations AI' aim to improve conversion rates and provide personalised product recommendations. For customer interaction and contact centres, 'Amazon Connect' and GCP's 'Contact Center AI' are available. Data management and analysis are supported by 'BigQuery', 'Amazon DocumentDB', and 'Azure Synapse Analytics'. Machine-learning capabilities are offered by 'Azure Machine Learning', 'Azure Machine Learning', and GCP's 'Document AI'. Specialised services like 'Amazon Neptune' and 'Google Workspace' cater to graph databases and productivity tools for schools and organisations.

'Amazon Personalize' and 'AWS Solutions' focus on real-time personalised recommendations and vetted solutions for enterprise use cases. 'AWS Supply Chain' and 'AWS Organizations' provide optimisation and management services for supply chain and account management, respectively. Additionally, AWS offers 'AWS Clean Rooms' for secure data operations across organisations, such as for digital marketing and advertising, 'AWS Data Exchange' for easy data sharing and monetisation, and 'AWS GovCloud' for government agencies' specific needs. These examples underscore the industry-focussed nature of cloud infrastructure services, tailored to the distinct requirements of individual sectors. By providing industry-specific tools and solutions, these cloud platforms and their partners empower enterprises in various sectors to employ the capabilities of AI.

The network reveals distinct industrial clusters in cloud AI across the three platforms, mirroring industry-specific needs and requirements. The first cluster spans the 'Healthcare and Life Sciences' and 'Government and Public Sector' sectors. The second cluster includes 'Financial services', 'Consumer Packaged Goods', and Supply Chain and Logistics'. The third cluster comprises 'Media and Entertainment' and 'Gaming'. The fourth cluster covers a diverse range of industries, including 'Manufacturing', 'Automotive', 'Education', 'Retail', and 'Telecommunications'. These clusters illustrate common needs and requirements shared across industries.

The 'AWS Marketplace' stands out in the cloud stack, gaining the highest citation count among the analysed industries. This online marketplace, similar to Azure and GCP's marketplaces, caters to a wide range of customers, including startups, large enterprises, government agencies, educational institutions, and non-profit organisations. Comparable to popular app stores like Apple's App Store and Google's Play Store for mobile applications, the AWS Marketplace offers a variety of software tools, business applications, services, and pre-trained (SageMaker) machine-learning models specifically designed for the cloud. It functions as a platform where customers can discover, purchase, and deploy apps and services from thousands of third-party vendors. The marketplace solutions complement the existing AWS cloud platform stack and acts as an important intermediation hub, providing access to a vast selection of pre-configured solutions in categories such as security, machine learning, data analytics, and more. These marketplaces play a crucial role in bridging the gap between infrastructure technology and the specific needs of individuals and organisations, regardless of industry, organisation type, or region.

A preliminary exploration of the solutions offered in the AWS, Azure, and Google Cloud marketplaces reveals intriguing insights. Figure 4 displays the significant differences between these marketplaces in terms of the number of marketplace apps across categories. Azure offers approximately 39,857 apps in total, followed by AWS with 32,987 apps, and GCP with 6387 apps.

AWS Marketplace is notable for its extensive offerings in categories such as (business) applications (7420), AI and ML (7051), computing and infrastructure (23,000), data and analytics (8655), industry-specific solutions (3777), IoT (1411), IT & operations (18,225), professional
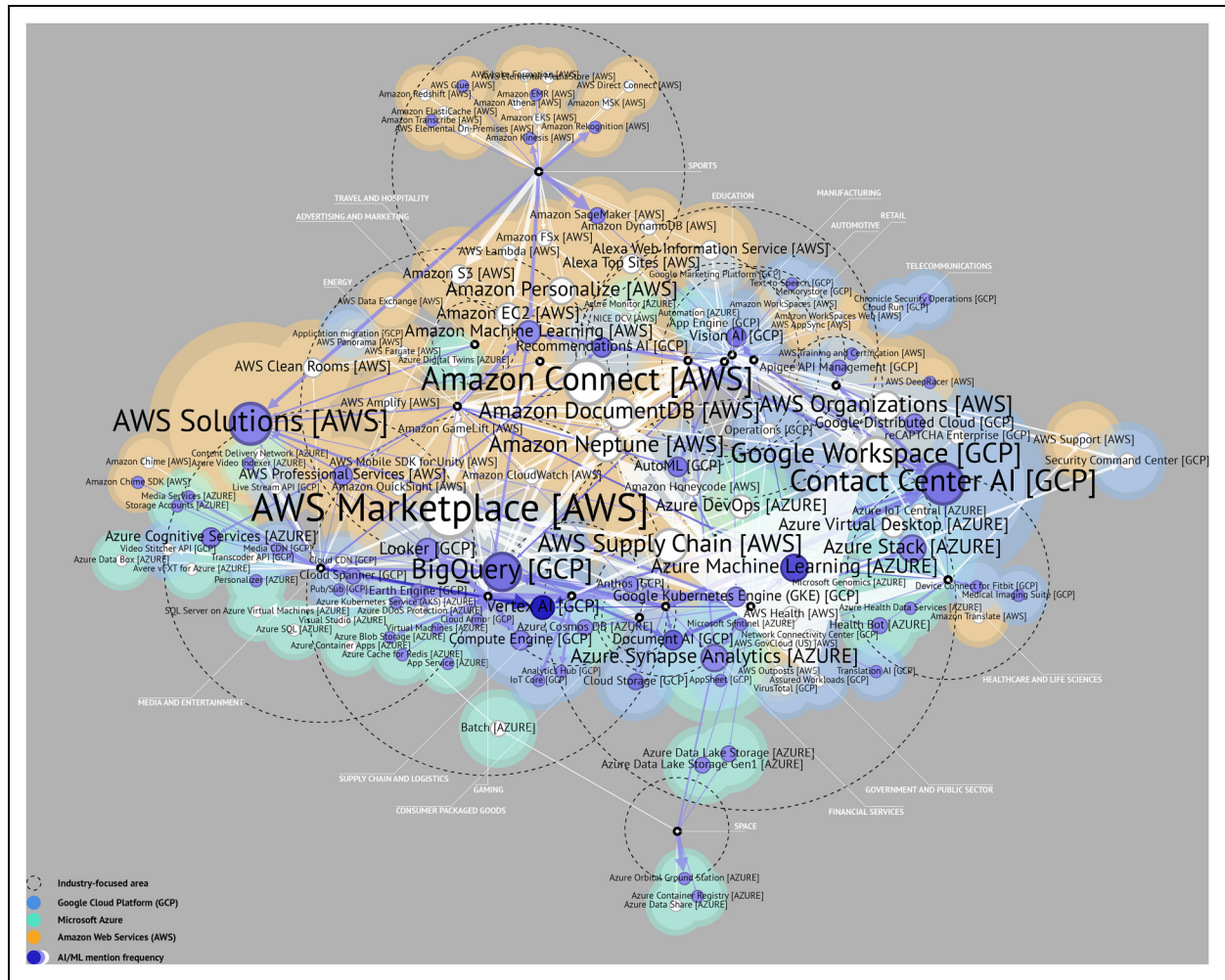
Figure 3. Cloud AI stacks: Citation bi-partite networks of cloud platform products and services mentioned across industry solutions pages, unified for Microsoft Azure, Amazon Web Services, and Google Cloud Platform. Dashed lines reveal overlapping industry-focussed areas. Colour scale reflects frequency of 'AI' and machine-learning mentions. Layout: Force Atlas 2; node scaling: by indegree (mention) count.

services (12,231), and security and compliance (3460). Azure Marketplace shows a strong presence in AI and ML (2410), computing and infrastructure (3768), data and analytics (4783), industry-specific solutions (1396), IoT (971), IT & operations (11,226), and professional services (5280). Google Cloud Marketplace focusses on data and analytics (1615), industry-specific solutions (876), IoT (218), and IT & operations (312). These findings highlight the diversity and depth of offerings across the Azure, AWS, and GCP marketplaces, demonstrating their capability to cater to various customer needs.

Further analysis of AI/ML mentions reveals the diffusion of AI and machine learning across all marketplace categories. An example from AWS Marketplace is Fusus' 'Real-Time Crime Center' platform, which transforms any video source into an AI-enabled camera for law enforcement and first responders.[8] This supports our previous

findings that AI/ML is integrated throughout the entire spectrum of apps and services. In terms of AI/ML offerings, AWS returns 4080 results (approx. 10.2%), Azure 2024 (6.1%), and GCP 483 (7.6%). These figures illustrate the diversity and depth of offerings in the Azure, AWS, and GCP marketplaces, emphasising their role as intermediaries that cater to a wide range of enterprise business needs, including industry-specific requirements.

AI's industrialisation and cloud infrastructure dependence

This study conducted an in-depth exploration of the roles played by the top-three cloud computing platform providers—Amazon, Microsoft, and Google (Alphabet)—in what we term the 'industrialisation of artificial intelligence'.
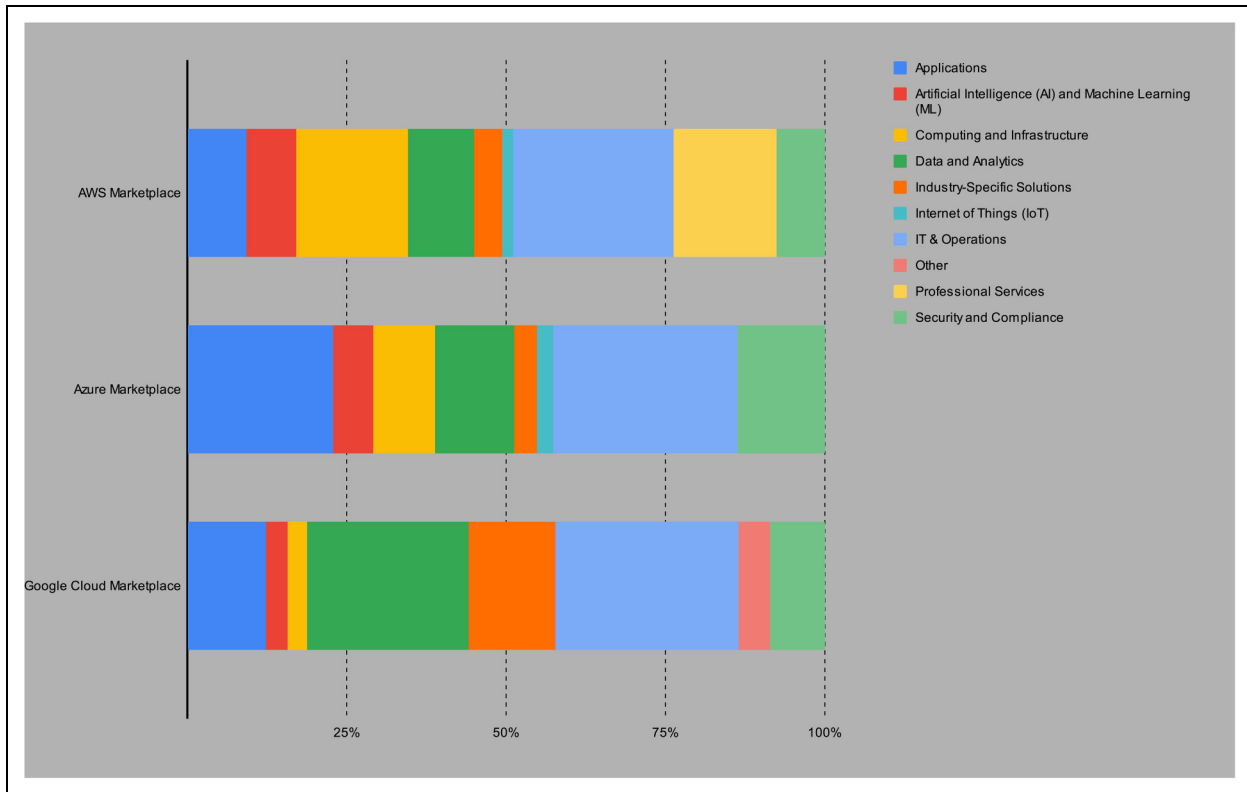
Figure 4. Classification of marketplace solutions in AWS Marketplace, Azure Marketplace, and Google Cloud Marketplace, shown by main category.

We employed a two-fold technographic approach to gain a profound understanding of their infrastructural support and investments within the AI sector, followed by a comprehensive examination of their cloud platform products and services. As AI transitions from research to real-world commercial products and services, we conceptualised AI as an *existing* technology stack and ecosystem, all the while acknowledging its discursive role. The following section discusses the implications of widespread reliance on cloud infrastructure.

In the first step of our analysis, we observed substantial infrastructural support and investments, evident through various partnerships, investments, and acquisitions. This underscores the need for smaller players to actively seek collaborations with Big Tech. These strategic alliances, partner programmes, and financial investments create new dependencies and 'lock-in' effects, tightly bound to the infrastructure and support provided by leading technology companies (Dyer-Witheford et al., 2019). Moreover, partner programmes and strategies are crucial in orchestrating platform ecosystems, attracting third-party developers and businesses to create applications and solutions, particularly those tailored to specific industries, thus driving the industrialisation of AI (Egliston and Carter, 2022; Helmond et al., 2019). These programmes are instrumental in consolidating the infrastructural and strategic power of

these platforms within the emerging AI ecosystem (Van der Vlist and Helmond, 2021).

All three companies have established AI startup programmes as a competitive battleground to attract and support emerging AI companies. They provide extensive resources, such as cloud infrastructure credits, specialised AI training, technical support, webinars, and access to experts. By offering these resources, often initially free, Big Tech effectively creates an environment where startups are incentivised to join their ecosystems. This strategy not only helps Big Tech attract startups but also positions these conglomerates as indispensable 'partners' in the startups' growth journey. The provision of cloud credits lowers the financial barrier for startups to access cloud infrastructure, while technical advice and development tools enable rapid development and deployment of AI solutions. In return, Big Tech benefits from fostering a network of innovators building on their platforms, further entrenching their infrastructure and services as foundational for the next wave of AI applications.

Consequently, the various partner, startup, and training programmes and their respective materials serve a dual purpose, as Burkhardt points out, to 'educate users *about* and recruit them *for* their products' (2019: 217, emphasis in original). Similarly, Luchs et al. highlight how Big Tech's AI and machine-learning online courses serve to

'consolidate and even expand their position of power by recruiting new AI talent and by securing their infrastructures and models to become the dominant ones' (2023: 1). This strategy leads to further dependencies and lock-in effects, with startups and developers becoming reliant on Big Tech's infrastructure and services. This entrenchment allows Big Tech to exert a standard-setting influence over AI development, as identified by Jacobides et al. (2021) and Luchs et al. (2023), shaping the future landscape of AI technology and its applications.

In the second step of our analysis, we observed that Amazon, Google, and Microsoft have integrated all three elements of the cloud AI stack—infrastructure, models, and applications—into their extensive cloud offerings. Concurrently, they have developed industry-focussed solutions and marketplaces aimed at attracting third-party developers and businesses, thereby fostering the expansion of their AI ecosystems. Our comprehensive mapping of Amazon, Microsoft, and Google's complete cloud AI stacks revealed a deep interdependence between AI and cloud infrastructure, emphasising the industry-specific characteristics of cloud AI, a convergence we refer to as 'Big AI'.

Developing foundational models like GPT using a 'bigger-is-better approach' (*The Economist*, 2023b) requires significant cloud computing resources, favouring larger companies (Jacobides et al., 2021; Kak and Myers West, 2023; Luitse and Denkena, 2021). The unparalleled (hyper-)scalability facilitated by cloud computing is a defining characteristic of Big AI (Narayan, 2022), resulting in new dependencies and corresponding investments in computational resources. However, it is important to critically evaluate innovation discourses that overly emphasise scaling as an end in itself. We should recognise the 'politics of scaling', as argued by Pfotenhauer et al. (2022), who contend that the 'fixation on "scaling up" has captured current innovation discourses and, with it, political and economic life at large' (3). Our study demonstrates that the industrialisation of AI extends beyond current LLMs and generative AIs, deeply embedding AI and machine learning in modern cloud computing stacks, and underpinning numerous activities and everyday processes across various industries. As a result, the industrialisation of AI signifies the becoming mundane of AI.

Cloud and AI partners serve as central intermediaries, connecting foundational cloud AI infrastructure with real-world AI applications for organisations and end-consumers. Conversely, Big Tech companies heavily depend on partnerships with semiconductor firms, chipmakers, and agreements for leasing data centre sites, emphasising the significance of these alliances in both directions. These dependencies extend beyond cloud platforms, involving 'enablers' like hardware manufacturers, chipmakers, and data management and processing (Jacobides et al., 2021: 415), further highlighting new dependencies in the AI landscape.

Strategic control over each component of the stack is crucial for optimising real-world AI applications, impacting performance and energy consumption, particularly at an industrial scale. This gives Big Tech substantial advantages over smaller players. However, the AI ecosystem is not solely shaped by Amazon, Microsoft, and Google; it also involves power dynamics with other major technology players like NVIDIA, which controls over 80% of the global GPU market. In response, Google introduced Cloud TPUs, Amazon developed their Inferentia chips, and Meta launched their Training and Inference Accelerator (MTIA) 'family of chips' as proprietary AI-optimised alternatives. Rella emphasises that this has transformed chip technology into a new battleground with its distinctive 'material political economy' (2023: 17).

Furthermore, Big AI presents substantial governance challenges that span various policy domains, including platform regulation and AI governance (Helberger and Diakopoulos, 2023; Lehdonvirta, 2022). The EU's Digital Markets Act (DMA) sets rules for digital service 'gatekeepers', aiming to ensure the openness of crucial digital services in the sector. Notably, the proprietors of Big AI provide multiple digital services simultaneously. Cloud platforms control not only computing infrastructure but also the marketplaces for AI models and applications, representing a dual power. In DMA terms, they offer both 'online intermediation services' (e.g. AWS Marketplace) and 'cloud computing services' (e.g. AWS SageMaker) for AI. This dual role cements Amazon, Microsoft, and Google in entrenched positions, often stemming from the creation of conglomerate ecosystems around their core platform services (Van Dijck et al., 2018; Van der Vlist, 2022), further fortifying existing barriers to entry.

Customers risk being locked into these ecosystems due to their reliance on Big AI offerings, with high switching costs exacerbating this risk. Understanding the politics of AI scalability is crucial for competition policy implications. For instance, partnerships between Big Tech and AI startups, like Microsoft and OpenAI, are likely to face scrutiny for potential hindrance to fair competition in the coming years.

Big AI beyond Amazon, Microsoft, and Google

In addition to Amazon, Microsoft, and Google, future research could explore the roles of other significant players in the AI landscape, such as Oracle, Adobe, IBM, and Salesforce. These companies also wield substantial influence within the broader ecosystem (Van der Vlist and Helmond, 2021) and specialise in various areas, including data management (e.g. 'Oracle Data Cloud' and 'Salesforce Data Cloud'), marketing clouds ('Adobe Experience Cloud'), and customer relationship management ('Salesforce Einstein GPT'). They offer tailored services and solutions, addressing the specific demands of their respective industries, and some feature hyperscale

computing architectures. These intermediaries play a crucial role in expanding the landscape of cloud and AI providers by providing industry-specific AI solutions. Market research using publicly accessible documentation to chart cloud implementations across industries and regions reveals distinct patterns in the clientele and market reach of cloud platform providers (Paraskevopoulos, 2023). For example, Oracle has a significant presence in South America, while Microsoft is favoured by large corporations and manufacturing entities, and Google is the preferred provider for smaller companies, particularly within the ICT sector. These observations highlight diverse strategies in the industrialisation of AI, with each platform carving out its own niche in different segments of the global AI market.

The roles of Chinese giants like Alibaba Cloud and Tencent Cloud, as well as the evolving roles of Meta and Apple in the AI ecosystem, also warrant further investigation. Apple and Meta, with their unique strategies targeting mobile devices, automated content moderation, digital marketing, AR/VR, and open-source initiatives, offer distinct advantages differentiating them from Amazon, Microsoft, and Google. This includes a shift towards smaller, specialised LLM models, such as Meta's open-source LLaMA, which are faster and cheaper to use, including on their own mobile devices (cf. *The Economist*, 2023b).

While the AI ecosystem extends beyond the Big Three, their support and investments remain critical. Consider Hugging Face, an influential open-source AI community platform that facilitates collaboration on machine-learning models and projects. Despite its prominence in the AI ecosystem, Hugging Face's strategic partnership with AWS illustrates the ongoing reliance on the infrastructure provided by major technology companies (cf. Kak and Myers West, 2023). Additionally, a diverse array of national and sectoral AI entities, including American, Chinese, and European companies and startups, collaborate with and contribute to the global AI landscape (Jacobides et al., 2021). Although these players often rely on or may be absorbed by Big Tech, their role in the industrialisation of AI through providing sector-specific solutions is significant and merits further attention.

Acknowledgements

Data availability

The data that support the findings of this study are openly available in the Open Science Framework (OSF) at https://doi.org/10.17605/osf.io/unvc2.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

ORCID iDs

F. N. (Fernando) van der Vlist https://orcid.org/0000-0003-1401-0325
A. (Anne) Helmond https://orcid.org/0000-0003-4327-4012
F. L. (Fabian) Ferrari https://orcid.org/0000-0003-0637-0232

Notes

1. 'AWS Marketplace', Amazon Web Services, https://aws.amazon.com/marketplace/; 'Azure Marketplace', Microsoft Azure, https://azuremarketplace.microsoft.com/marketplace/; and 'Google Cloud Marketplace', Google Cloud, https://console.cloud.google.com/marketplace.
2. 'AWS Documentation', Amazon Web Services, https://docs.aws.amazon.com/index.html; 'Directory of Azure Cloud Services', Microsoft Azure, https://azure.microsoft.com/products/; and 'Products and Services', Google Cloud Platform, https://cloud.google.com/products#general-availability.
3. All products and services were collected in January 2023; analyses were conducted on 15–19 May 2023. As such, data collection occurred prior to Amazon, Microsoft, and Google's rollout of new AI products in Spring 2023, capitalising on the AI frenzy. Marketplace solutions were collected on 30 June 2023.
4. Microsoft, 'Azure OpenAI', Azure Marketplace, https://azuremarketplace.microsoft.com/en-us/marketplace/apps/Microsoft.CognitiveServicesOpenAI.
5. 'Machine Learning Partner Solutions', AWS AI, https://aws.amazon.com/ai/partners/
6. 'AI + machine learning', https://azure.microsoft.com/en-us/overview/ai-platform/; 'Machine Learning' https://docs.aws.amazon.com/index.html; and 'AI and Machine Learning', https://cloud.google.com/products#general-availability.
7. The advanced search queries used encompassed relevant variations of these terms, including 'AI', 'artificial intelligence', 'machine learning', and 'deep learning'.
8. Fusus, 'Fusus Real-Time Crime Center In The Cloud (RTC3)', AWS Marketplace, https://aws.amazon.com/marketplace/pp/prodview-3xxezgl5la7ky.

References

AWS Startup Loft (n.d.) AWS generative AI accelerator. In: *AWS Startup Loft*. Available at: https://aws-startup-lofts.com/amer/program/accelerators/generative-ai.
Blanke T and Pybus J (2020) The material conditions of platforms: Monopolization through decentralization. *Social Media + Society* 6(4): 1–13.

Bucher T (2016) Neither black nor box: Ways of knowing algorithms. In: Kubitschko S and Kaun A (eds) *Innovative Methods in Media and Communication Research*. Palgrave Macmillan, 81–98.

Burkhardt M (2019) Mapping the democratization of AI on GitHub: A first approach. In: Sudmann A (eds) *The Democratization of Artificial Intelligence*. transcript Verlag, 209–222.

Cohere Team (2022) Cohere is available on the Google Cloud Marketplace. In: *Cohere*, 29 August. Available at: https://txt.cohere.com/cohere-is-available-on-the-google-cloud-marketplace/.

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

de Reuver M, Sørensen C and Basole RC (2018) The digital platform: A research agenda. *Journal of Information Technology* 33(2): 124–135.

Dezen N (2023) Microsoft creates new opportunities for partners through AI offerings and expansion of Microsoft Cloud Partner Program. In: *Official Microsoft Blog*, 22 March. Available at: https://blogs.microsoft.com/blog/2023/03/22/microsoft-creates-new-opportunities-for-partners-through-ai-offerings-and-expansion-of-microsoft-cloud-partner-program/.

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

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.

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

Ferrari F and McKelvey F (2022) Hyperproduction: A social theory of deep generative models. *Distinktion: Journal of Social Theory* 24(2): 338–360.

Gates B (2023) The age of AI has begun. In: *GatesNotes*, 21 March. Available at: https://www.gatesnotes.com/The-Age-of-AI-Has-Begun.

Gillespie T (2010) The politics of 'platforms'. *New Media & Society* 12(3): 347–364.

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/.

Greenfield E (2023) Adobe Firefly generative AI & Adobe Express to inspire millions to create in partnership with Google. In: *Adobe Blog*. Available at: https://blog.adobe.com/en/publish/2023/05/10/adobe-firefly-adobe-express-google-bard.

Helberger N and Diakopoulos N (2023) ChatGPT and the AI Act. *Internet Policy Review* 12(1).

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.

Hodgson C (2023) Tech giants pour billions into cloud capacity in AI push. *Financial Times*, 6 November. Available at: https://www.ft.com/content/f01529ad-88ca-456e-ad41-d6b7d449a409.

Holt J and Vonderau P (2015) 'Where the Internet Lives': Data centers as cloud infrastructure. In: Parks L and Starosielski N (eds) *Signal Traffic: Critical Studies of Media Infrastructures*. University of Illinois Press, 71–93.

Ichhpurani K (2023) Building the most open and innovative AI ecosystem. In: *The Keyword (Google)*, 14 March. Available at: https://cloud.google.com/blog/products/ai-machine-learning/building-an-open-generative-ai-partner-ecosystem.

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

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

Khan L (2023) Lina Khan: We must regulate A.I. Here's how. In: *The New York Times*, 3 May. Available at: https://www.nytimes.com/2023/05/03/opinion/ai-lina-khan-ftc-technology.html.

Klinge TJ, Hendrikse R, Fernandez R, et al. (2023) Augmenting digital monopolies: A corporate financialization perspective on the rise of Big Tech. *Competition & Change* 27(2): 332–353.

Küspert S, Moës N and Dunlop C (2023) The value chain of general-purpose AI. In: *Ada Lovelace Institute*, 10 February. Available at: https://www.adalovelaceinstitute.org/blog/value-chain-general-purpose-ai/.

Lehdonvirta V (2022) *Cloud Empires: How Digital Platforms Are Overtaking the State and How We Can Regain Control*. MIT Press.

Luchs I, Apprich C and Broersma M (2023) Learning machine learning: On the political economy of big tech's online AI courses. *Big Data & Society* 10(1): 1–12.

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.

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

Manyika J (2023) Why we focus on AI (and to what end). In *The Keyword (Google)*, 16 January. Available at: https://blog.google/technology/ai/why-we-focus-on-ai-and-to-what-end.

Microsoft (2023a) AI Partners: Working with Microsoft AI. In: *Microsoft AI*. Available at: https://www.microsoft.com/en-us/ai/partners.

Microsoft (2023b) Microsoft for Startups. Available at: https://www.microsoft.com/en-us/startups/.

Microsoft Corporate Blogs (2023) Microsoft and OpenAI extend partnership. In: *Official Microsoft Blog*, 23 January. Available at: https://blogs.microsoft.com/blog/2023/01/23/microsoftandopenaiextendpartnership/.

Murgia M (2023) Risk of 'industrial capture' looms over AI revolution. In: *Financial Times*, 23 March. Available at: https://www.ft.com/content/e9ebfb8d-428d-4802-8b27-a69314c421ce.

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.

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

Paraskevopoulos D (2023) *Global Cloud Projects Report and Database 2023*. IoT Analytics.

Pfotenhauer S, Laurent B, Papageorgiou K, et al. (2022) The politics of scaling. *Social Studies of Science* 52(1): 3–34.

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

Rella L (2023) Close to the metal: Towards a material political economy of the epistemology of computation. *Social Studies of Science* 54(1): 3–29.

Richter (2023) Big three dominate the global cloud market. In: *Statista*, 28 April. Available at: https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/.

Rieder B (2022) Towards a political economy of technical systems: The case of Google. *Big Data & Society* 9(2): 1–5.

Ropek L (2023) Amazon Web Services meltdown temporarily breaks the Internet. *Gizmodo*, 13 June. Available at: https://gizmodo.com/aws-amazon-outage-associated-press-mcdonalds-1850536296.

Salesforce (2023) Salesforce and Google Cloud announce expanded strategic partnership to unlock the power of AI, data, and CRM. *Salesforce*, 7 June. Available at: https://www.salesforce.com/uk/news/press-releases/2023/06/07/salesforce-google-ai-partnership/.

Sivasubramanian S (2023) Announcing new tools for building with generative AI on AWS. In: *AWS Machine Learning Blog*, 13 April. Available at: https://aws.amazon.com/blogs/machine-learning/announcing-new-tools-for-building-with-generative-ai-on-aws/.

The Economist (2022) Huge 'foundation models' are turbo-charging AI progress. In: *The Economist*, 11 June. Available at: https://www.economist.com/interactive/briefing/2022/06/11/huge-foundation-models-are-turbo-charging-ai-progress.

The Economist (2023a) Big tech and the pursuit of AI dominance. In: *The Economist*, 26 March. Available at: https://www.economist.com/business/2023/03/26/big-tech-and-the-pursuit-of-ai-dominance.

The Economist (2023b) 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.

van der Vlist FN (2022) *The platform as ecosystem: Configurations and dynamics of governance and power*. PhD Thesis, Utrecht University.

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.

van der Vlist FN, Helmond A, Burkhardt M, et al. (2022) API Governance: The case of Facebook's evolution. *Social Media + Society* 8(2): 1–24.

Van Dijck J, Poell T and de Waal M (2018) *The Platform Society: Public Values in a Connective World*. Oxford University Press.

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

Whittaker M, Crawford K, Dobbe K, et al. (2018) *AI Now Report 2018*. AI Now Institute.

Widder DG and Nafus D (2023) Dislocated accountabilities in the 'AI supply chain': Modularity and developers' notions of responsibility. *Big Data & Society* 10(1): 1–12.

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.

Wiggers K (2023) Google Cloud and Salesforce team up to bolster AI offerings. In: *Tech Crunch*, 7 June. Available at: https://techcrunch.com/2023/06/07/google-cloud-and-salesforce-team-up-to-bolster-ai-offerings/.

Yang Y and Gokturk B (2023) Google Cloud brings generative AI to developers, businesses, and governments. In: *Google Cloud Blog*, 14 March. Available at: https://cloud.google.com/blog/products/ai-machine-learning/generative-ai-for-businesses-and-governments.