



Volume 13 Issue 2



RESEARCH  
ARTICLE



OPEN  
ACCESS



PEER  
REVIEWED

# Platform power in AI: The evolution of cloud infrastructures in the political economy of artificial intelligence

**Dieuwertje Luitse** *University of Amsterdam* [d.luitse@uva.nl](mailto:d.luitse@uva.nl)

**DOI:** <https://doi.org/10.14763/2024.2.1768>

**Published:** 26 June 2024

**Received:** 13 November 2023 **Accepted:** 31 January 2024

**Competing Interests:** The author has declared that no competing interests exist that have influenced the text.

**Licence:** This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License (Germany) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. <https://creativecommons.org/licenses/by/3.0/de/deed.en>  
Copyright remains with the author(s).

**Citation:** Luitse, D. (2024). Platform power in AI: The evolution of cloud infrastructures in the political economy of artificial intelligence. *Internet Policy Review*, 13(2). <https://doi.org/10.14763/2024.2.1768>

**Keywords:** Infrastructural power, Artificial intelligence, Evolution, Political economy, Cloud computing

**Abstract:** In recent years, Amazon, Microsoft, and Google have become three of the dominant developers of AI infrastructures and services. The increasing economic and political power of these companies over the data, computing infrastructures, and AI expertise that play a central role in the development of contemporary AI technologies has led to major concerns among academic researchers, critical commentators, and policymakers addressing their market and monopoly power. Picking up on such macro-level political-economic analyses, this paper more specifically investigates the micro-material ways infrastructural power in AI is operated through the respective cloud AI infrastructures and services developed by their cloud platforms: AWS, Microsoft Azure, and Google Cloud. Through an empirical analysis of their evolutionary trajectories in the context of AI between January 2017 and April 2021, this paper argues that these cloud platforms attempt to exercise infrastructural power in three significant ways: through vertical integration, their complementary innovation, and the power of abstraction. Each dynamic is strategically mobilised to strengthen these platforms' dominant position at the forefront of AI development and implementation. This complicates the critical evaluation and regulation of AI technologies by public authorities. At the same time, these forms of infrastructural power in the cloud provide Amazon, Microsoft, and Google with leverage to set the conditions of possibility for future AI production and deployment.

This paper is part of **Locating and theorising platform power**, a special issue of *Internet Policy Review* guest-edited by David Nieborg, Thomas Poell, Robyn Caplan and José van Dijck.

## Introduction

Powered by a corporate 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 three of the dominant developers of artificial intelligence<sup>1</sup> (AI) infrastructures and services (Srnicek, 2022). Beyond leveraging their vast amounts of data, these three big tech corporations have equipped themselves with the necessary technical knowledge through the attraction of AI expertise, the acquisitions of AI startups (e.g. Google/Deepmind), and the emergence of extensive business partnerships such as those of Microsoft and OpenAI (Murgia, 2023). 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. Accounting for 68% of the global cloud-computing market (Synergy Research Group, 2024), these platforms provide full-stack integrated tools and services for the production, training, and deployment of machine-learning systems and applications thriving on their proprietary cloud infrastructures that provide computing power at scale (Luitse & Denkena, 2021). The most recent developments involve the creation and provision of very large pre-trained models – 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 expectations and hype around AI’s potential, corporate cloud AI<sup>2</sup> infrastructures and services are increasingly being implemented across different economic and societal sectors. This has raised significant concerns about the bias and social ramifications of these systems as they risk exacerbating existing patterns of social inequality and discrimination (e.g. Bender et al., 2021; Miceli et al., 2022). In addition, following scholarly discussions on AI’s political economy (Luitse & Denkena, 2021; Widder et al., 2023), these rapid implementations allow cloud platforms to leverage their AI infrastructures as core “commercial computing assets” (Narayan, 2022). This is producing new relations of dependency for third-par-

1. AI encompasses various approaches rooted in different research traditions and disciplines. Using AI and machine learning in this paper, I refer to ‘narrow AI’ deep-learning systems.
2. I understand cloud AI as the assemblage of service-based machine-learning tools and systems, as well as the underlying data and computing infrastructures operated through cloud platforms.

ty developers as platforms keep control over the tools, as well as the development environments for AI production and deployment. Consequently, researchers and public commentators observe a rapid concentration of economic and political power (monopolisation) into the hands of a small set of corporations (Srnicek, 2022). This tendency is increasingly met with calls for action to confront and reclaim this power through regulatory intervention by, among others, the US Federal Trade Commission and the European Commission (Kak & Myers West, 2023). Contributing to this body of work on the political economy of AI, this paper provides an empirical case study into how AWS, Microsoft Azure, and Google Cloud have strategically been operationalising their power in AI production and deployment over time through the cloud AI ecosystems they operate.

Critical analyses of big tech's monopoly power in AI have put forward valuable insights on a macro-level, proving important entry points for regulation. However, relatively little attention has been paid to the specific micro-material ways in which power in AI is operationalised by their cloud subsidiaries through the *evolution* of their network of AI infrastructures and services. An explicit focus on infrastructure is important as "cloud computing arrangements [...] are foundational to platform expansion" (Narayan, 2022, p. 915). As evolving assemblages of hardware and software services they set the conditions for the development of AI systems (Rieder, 2022). Dominant actors who own and operate these infrastructures thus give shape to the AI's present in terms of "what we do (and do not) know about [it]" (Whittaker, 2021a, 55) as well as how it can be produced and deployed. At the same time, they predetermine the future trajectories of the technologies that are increasingly being implemented into societal domains such as cultural production, healthcare, and the security industry (Jacobides et al., 2021) with significant implications for these respective areas. AWS, Microsoft Azure, and Google Cloud are the cloud computing subsidiaries of lead firms in AI research and development (Rikap, 2023) as well as the cloud-infrastructure market (Synergy Research Group, 2024). Consequently, they warrant a deeper examination into the specific operations of the infrastructures and services for AI they have developed over time. Such an empirical investigation allows us to better understand the ways in which these platform corporations have strategically manifested themselves in the field and how they attempt to exercise power over industrial AI development and implementation.

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. I investigate the (often concealed) development of their cloud AI infrastructures and services over this period to show how these major cloud platforms have been operating *specific forms of infrastructural power* in AI's larger political economy. Adapting Khalili (2018) toward the context of cloud AI, I understand infrastructural power as (cloud) platforms' ability to forge the assemblage of computational infrastructures, AI development practices, discourses, and governing procedures with the strategic aim of (re)producing and enforcing capitalist relations. This type of platform power is relational, dispersed, and materially emerges through the large network of cloud AI infrastructures and services as they set the conditions of possibility for AI production and deployment with the strategic aim to strengthen the political and economic position of the companies that operate them.

Pursuing this inquiry, I first situate this research within the literature on the political economy of AI and the infrastructural power of cloud platforms. This is followed by a discussion of the methodology that I term *evolutionary platform technography*, as well as the (archived) materials I analysed to trace the evolution of three cloud ecosystems' AI-specific infrastructures and services. The remainder of the paper draws from these platform evolutionary trajectories to empirically analyse the manifestation of infrastructural power in AI in three substantial ways – through *vertical integration*, *complementary innovation*, and *abstraction* – and considers the implications for the development of the field.

## **Infrastructural power in the cloud and the political economy of AI**

The power of big tech corporations like Amazon, Microsoft and Google is often theorised in terms of market and monopoly power (e.g. Khan, 2018; van Dijck et al., 2019) and the contemporary development of AI is further driving their monopolising tendencies. Research on AI's political economy in business studies, critical AI studies, and platform studies has paid particular attention to these dynamics (e.g. Jacobides et al., 2021; Srnicek, 2022) and identified three aspects that steer this concentration of power in the field. First, tech corporations champion the collection and provision of vast amounts of valuable data that are key to AI development. Second, these companies have many more financial resources to invest in and retain highly skilled AI researchers to capture expertise in the field. Third, the concentrated ownership and control of big tech over computing resources such as Graphics Processing Units (GPUs) facilitates a "compute divide" (Ahmed & Wahed, 2020, p. 1) that provides significant advantages for these corporations and a small

set of partnering academic research institutions (Whittaker, 2021a). Taken together, these dynamics present a major obstacle towards democratisation of AI development – i.e. involving a wider consortium of people (e.g. academic researchers, public participants, and regulators) to contribute to AI development processes, including critical evaluation and auditing mechanisms (Seger et al., 2023).

Contributing to these observations, this paper considers the infrastructural power of AWS, Microsoft Azure, and Google Cloud within AI's wider political-economic structure. Here, I draw from and build on critical work in platform studies (e.g. van Dijck, 2020; Plantin 2020; Rieder, 2022) that has described how the transformation of platform architectures into service components and their accompanying technical procedures “may effectively alter market dynamics and, by extension, political-economic power relations” (Lomborg et al., 2024). These shifting power relations emerge as platforms set technical standards, terms and conditions, and criteria for AI production and deployment through a variety of so-called “technical boundary resources” (Ghazawneh & Henfridsson, 2013). Such resources include Application Programming Interfaces (APIs), Software Development Kits (SDKs), developer guidelines, data and software tools to build machine-learning systems, and applications on top of the pre-existing platform infrastructures. These features demand developers across domains to align and integrate their practices and data infrastructures with those of the larger platform ecosystem (Nieborg & Helmond, 2019). Consequently, this allows cloud platforms to not only infrastructurally extend themselves through deep integration, but also to profit from their intermediary position as the owner of the core infrastructures AI systems are built on.

Vertical integration has been identified as one of the (key) sources of infrastructural power held by big tech companies (Khan, 2018). This process entails the “seamless integration of platforms” (van Dijck, 2020, p. 8) into the underlying proprietary “stack” (Bratton, 2016) of scalable infrastructures that make up the various material and abstract layers of (planetary-scale) computation owned and operated by this small set of powerful corporations. This consequently allows these companies to further channel developer activities under their corporate control, creating path-dependencies that result in user lock-in and vendor lock-in (van Dijck, 2020). Cloud platforms like AWS, Microsoft Azure, and Google Cloud operate according to this dynamic by leveraging their vast and complex set of infrastructural technologies and products as integrated services across three layers of the cloud stack: 1) Infrastructure-as-a-Service (IaaS); 2) Platform-as-a-Service (PaaS); and 3) Software-as-a-Service (SaaS). IaaS refers to the on-demand delivery of various computing resources such as server capacity, virtualisation technology, and networking capabili-

ties (Narayan, 2022) by corporate clouds, including the provision of processing units for the training and inference of machine-learning models. The PaaS-layer consists of a collection of integrated development resources including operating systems, machine-learning frameworks, and SDKs. Via this layer, cloud providers also manage the underlying computing infrastructures for third parties within a respective cloud ecosystem. SaaS allows third-party developers or individual customers to source integrated proprietary technologies that are fully controlled and operated by cloud platforms. This service layer is also referred to as AI-as-a-Service (AlaaS) (Parsaeefard et al., 2019) where services can be used for advanced computing capabilities such as facial recognition or language generation. Following van Dijck (2020), the vertical integration of IaaS, PaaS, and AlaaS services across the cloud stack indicates a strategic move by corporate clouds to potentially expand their control over AI development and implementation processes, which may strengthen their (critical) intermediary position. From this perspective, I draw on these insights to more specifically outline how AWS, Microsoft Azure, and Google Cloud attempt to exert this source of power through the infrastructures and services for AI production and deployment they have developed.

The ability of cloud platforms to leverage their infrastructural “core “ of computing assets and micro-services to third-party institutions has been considered another important source of their infrastructural power as it allows them to expand and integrate themselves into various domains of application (Aradau & Blanke, 2022). As Rieder (2022) explains, this dynamic of spawning “complementary innovation” (Gawer, 2014) is – next to the cross-market utility of data – driven by the transversality of software, computing hardware, and their “capacity to articulate, structure, and automate processes in very different task environments” (p. 3). In other words, following research in business studies (e.g. Gawer, 2014), proprietary cloud computing infrastructures such as operating systems, processing chips in data centres, or software-service portfolios can strategically be applied to facilitate “innovation” by third-party institutions or companies in a variety of different domains. Thus far, this has afforded companies like Amazon, Microsoft, and Google, which are behind the development of these infrastructural technologies, to expand their activities to penetrate different markets and societal sectors through their cloud computing branches and solidify their dominant position at a brisk pace (van Dijck et al., 2019). The rapid provision of AI-specific infrastructures and services is further accelerating this process as machine-learning systems have the potential to serve many purposes (Rieder, 2022) and can be applied in a plurality of contexts. Considering this issue, this paper investigates how AWS, Microsoft Azure, and Google Cloud have specifically been developing and operating their AI infrastruc-

tures and services toward (new) domains of activity, thereby mobilising their infrastructural power over third-party developers looking to adapt AI-driven systems into their respective practices.

## **Evolutionary technographic analysis of AWS, Microsoft Azure, and Google Cloud infrastructures for cloud AI**

Empirical research into the evolution of cloud platforms is a challenge, as computational ecosystems are not only rather complex and opaque, but are also subject to continuous change. To overcome these difficulties and examine how these companies operationalise forms of infrastructural power in AI through their cloud ecosystems, I developed a methodological approach that I term *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 allows for the critical observation and description of the workings of technical systems, including entire platforms. It hence allows us to trace their development in an evolutionary manner to account for the "*economic growth and technological expansion*" (Nieborg & Helmond, 2019, p. 197, emphasis in original) and, ultimately, the exertion of power by the cloud platforms that operate these infrastructures.

A technographic analysis of primary and secondary sources such as (platform) documentation, press releases, media or industry reports, and corporate blog posts, Bucher (2018) explains, is a way of reading such materials to develop a "critical understanding of the mechanisms and operational logic of software" (p. 61). It concentrates on the suggestive qualities of sociotechnical systems and infrastructures, i.e. how these systems and infrastructures work, and more importantly, who they work for (Galloway, 2004). Furthermore, such primary industrial resources – the material traces that platforms strategically distribute – provide insights into the "evolving production, preferred usage and embedded politics of software objects" (Helmond & van der Vlist, 2021, p. 4) such as cloud infrastructures and services as well as big tech's larger infrastructural ambitions (Nieborg & Helmond, 2019). As such, these resources offer valuable entry points to critically analyse specific strategies behind the technological development of a platform within a particular field such as AI. It is, however, important to acknowledge the corporate nature of these materials as they are strategic and self-serving, produced and distributed by industry. Therefore, they are likely to obfuscate potential technological risks of the services platforms tend to promote. Additionally, these documents neither provide insight into the specific operations of control in AI development ecosystems which are likely to be buried in contracts and terms and conditions, nor do they allow for

an analysis of how companies are violating antitrust laws (e.g., Khan, 2018) to further exploit their monopoly power. Relying on the selected public documents, thus, requires me to remain attentive to the discursive strategies behind these materials. As such, these resources demand careful and critical distance to avoid shallow perpetuation of corporate claims that risk further amplifying AI-hype cycles.

The purposefully adapted evolutionary platform technography draws from a wide range of historical data from AWS, Microsoft Azure, and Google Cloud released between January 2017 and April 2021. As web archives are considered valuable for locating such material (strategic) platform resources (Helmond & van der Vlist, 2021), the selected materials for this study, include (1) archived product pages and documentation available via the Internet Archive Wayback Machine; (2) AWS, Microsoft Azure, and Google Cloud's corporate blog posts and press releases; and (3) relevant media or industry reports. The period was set primarily after Sundar Pichai's 2017 proclamation as it is considered to have started "another paradigm shift in the history of computing" (Burkhardt, 2019, p. 209). Following this moment, I trace the development of cloud AI infrastructures and services over time in the run-up to the release of foundation models. These models now dominate current AI discourses but are arguably a result of ongoing strategic developments in the AI industry.

For each of the three cloud platforms under study, the set of archived cloud AI product pages was systematically retrieved using the Wayback Machine. Additionally, I draw from corporate blog posts that specifically focus on the topics of AI and machine learning (Table 1). To identify and collect a relevant list of posts, all four blogs were queried for mentions of [AI], [artificial intelligence], and [machine learning]. In a similar vein, press releases on AI and machine learning products and services have been collected and added to the final dataset of platform resources that were considered for analysis. Lastly, as Bucher (2018) suggests, I considered a number of relevant media and industry reports. Following the set period of study, the dataset begins in 2017 and ends in April 2021. Appendices I and II provide a full overview of the collected (archived) cloud AI platform resources. These materials are organised per cloud platform and sorted by type and date to facilitate the analysis of these resources over time. Appendix III provides insight into the industry reports considered as secondary resources in this study, whereas relevant media sources are directly cited.

The analysis was carried out along two complementary stages of empirical inquiry. First, I examined the collection of archived product pages to reconstruct how AWS, Microsoft Azure, and Google Cloud have been developing their cloud AI infrastruc-

tures and services over time. This part of the research is substantiated by purposefully designed visualisations that present excerpts of these reconstructions per company (Figures 1–4 and 6–11).<sup>3</sup> The exploration and visualisation of the cloud platforms' evolution provided an overview of the respective levels of the stack toward various AI infrastructures and services that have been developed and operated – IaaS, PaaS, or AlaaS – as well as their domain of application. This part of the analysis paid specific attention to corporate development and operation of computing infrastructure, such as purpose-built hardware, that has proven to play a particularly important role in the production and deployment of contemporary machine-learning systems (Vipra & Myers West, 2023).

**TABLE 1:** Overview of corporate blogs from Amazon, Microsoft, and Google that were queried for mentions of [AI], [artificial intelligence], and [machine learning]

COMPANY	BLOG	URL
Amazon	AWS Machine Learning Blog	<a href="https://aws.amazon.com/blogs/machine-learning/">https://aws.amazon.com/blogs/machine-learning/</a>
Microsoft	Azure Blog and Updates	<a href="https://azure.microsoft.com/en-us/blog/">https://azure.microsoft.com/en-us/blog/</a>
Google	Google Cloud Blog   News, Features and Announcements	<a href="https://cloud.google.com/blog/">https://cloud.google.com/blog/</a>
Google	The Keyword   Google	<a href="https://www.blog.google/products/google-cloud/">https://www.blog.google/products/google-cloud/</a>

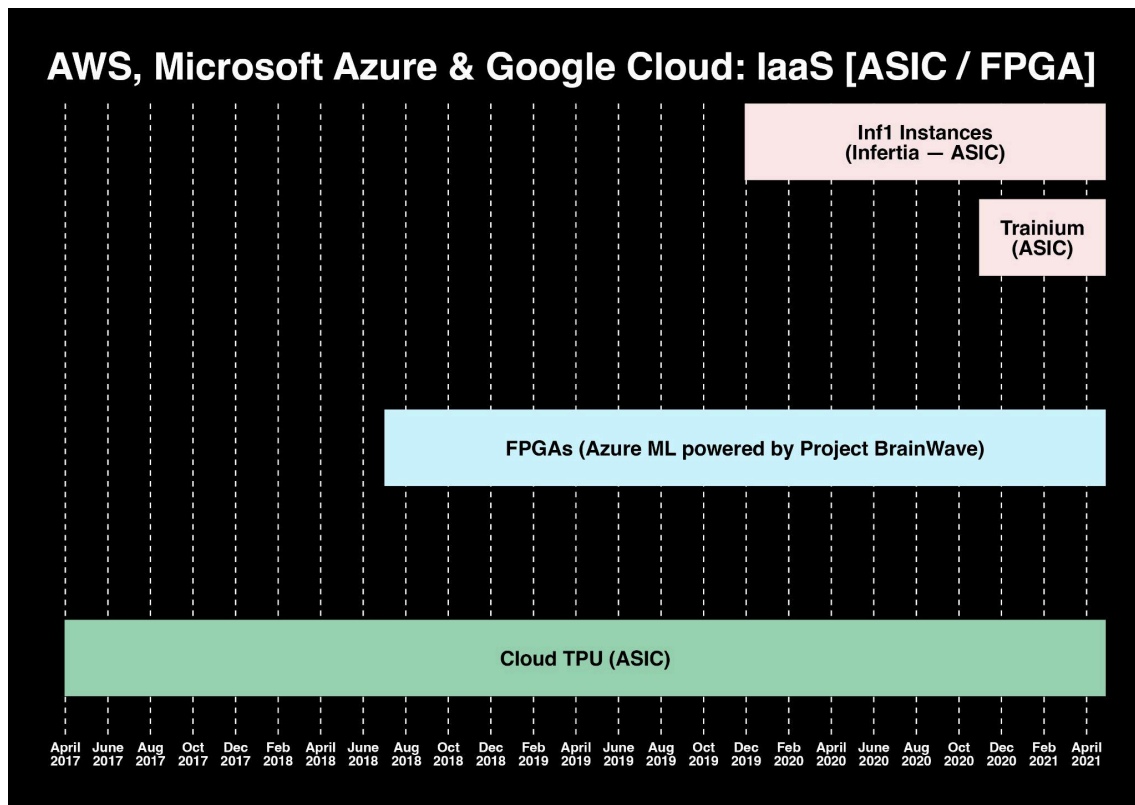
The second stage of the analysis builds on the previous level of inquiry. Following Bucher's (2012) statement that technography requires the critical observation, description, and interpretation of technical systems on their own material-discursive terms, this step involved a document analysis and close reading of the set of collected archived product documentation, blog posts, and press releases. In analysing these materials specific attention has been paid to distinguishing the strategic positions of AWS, Microsoft Azure, and Google Cloud in mobilising and operating their growing collection of AI-related infrastructures and services. This part is substantiated by an analysis of the collected media and industry reports that discuss economic strategies across different fields of AI. By outlining the corporate platform dynamics, this technographic exploration into the evolution of their platform-specific operations allowed me to gain empirical insight into the

3. Due to the size of the visualisations this article only includes image excerpts. An overview of the AI platform evolutions of AWS, Microsoft Azure, and Google Cloud (2017–April 2021) that support the findings of this study are openly available on the Open Science Framework (OSF) (Luitse, n.d.).

specific ways AWS, Microsoft Azure, and Google Cloud seek to exert infrastructural power within AI's larger political economy.

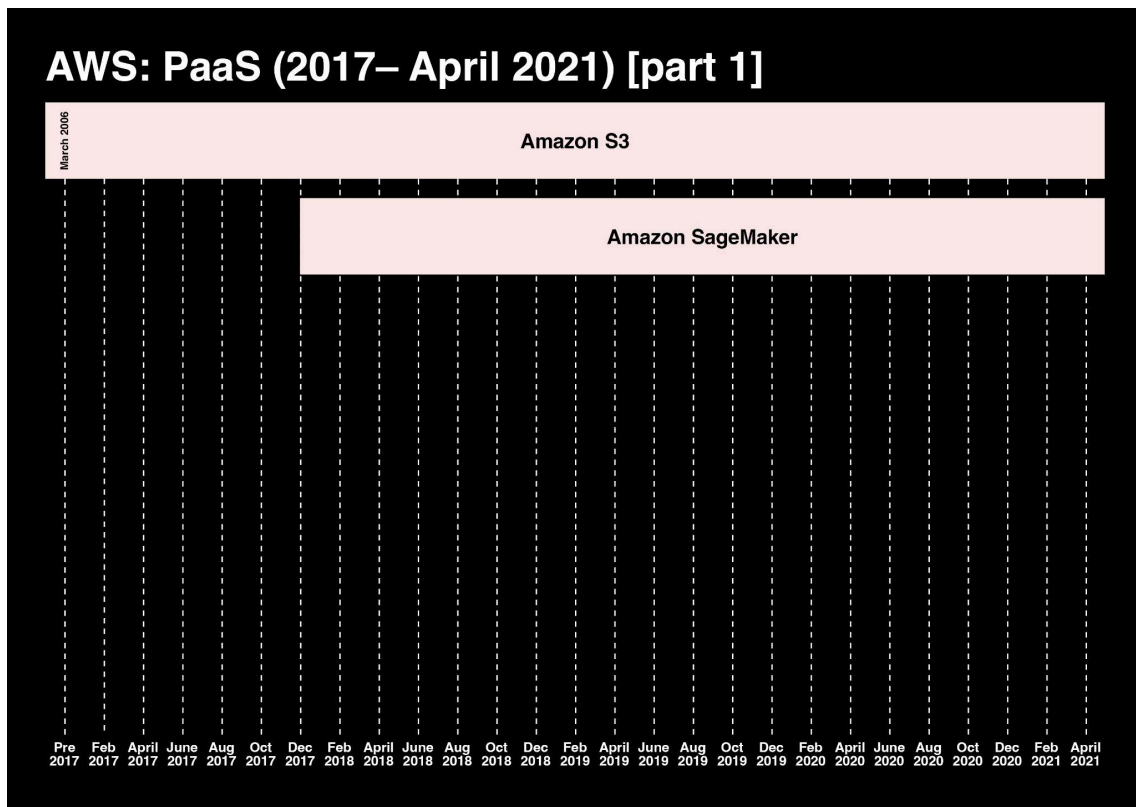
## Vertical integration

The evolutionary trajectories of AWS, Microsoft Azure, and Google Cloud demonstrate that these platforms strategically extended their AI capacities across the multiple layers of the cloud stack by vertically integrating their computing infrastructures into their respective PaaS and IaaS services (cf. Aradau & Blanke, 2022). Turning toward the bottom layer of infrastructures for computation first, the visualisations reveal that AWS and Google Cloud most specifically have been developing custom processing units for the training and inference of machine-learning systems and applications (Figure 1). It was Google's in-house AI research team that started this process in 2016 as the company needed new hardware that could better suit "the fast-growing computational demands of neural networks" (Sato & Young, 2017). This led to the development of an application-specific integrated circuit (ASIC) optimised for deep-learning processes: the *Tensor Processing Unit* (TPU). By now, this specialised chip is merely accessible as an infrastructural service through Google Cloud's *Computing Engine* and *Kubernetes Engine*, or its AI platform – now *Vertex AI* (GCP–2021c).

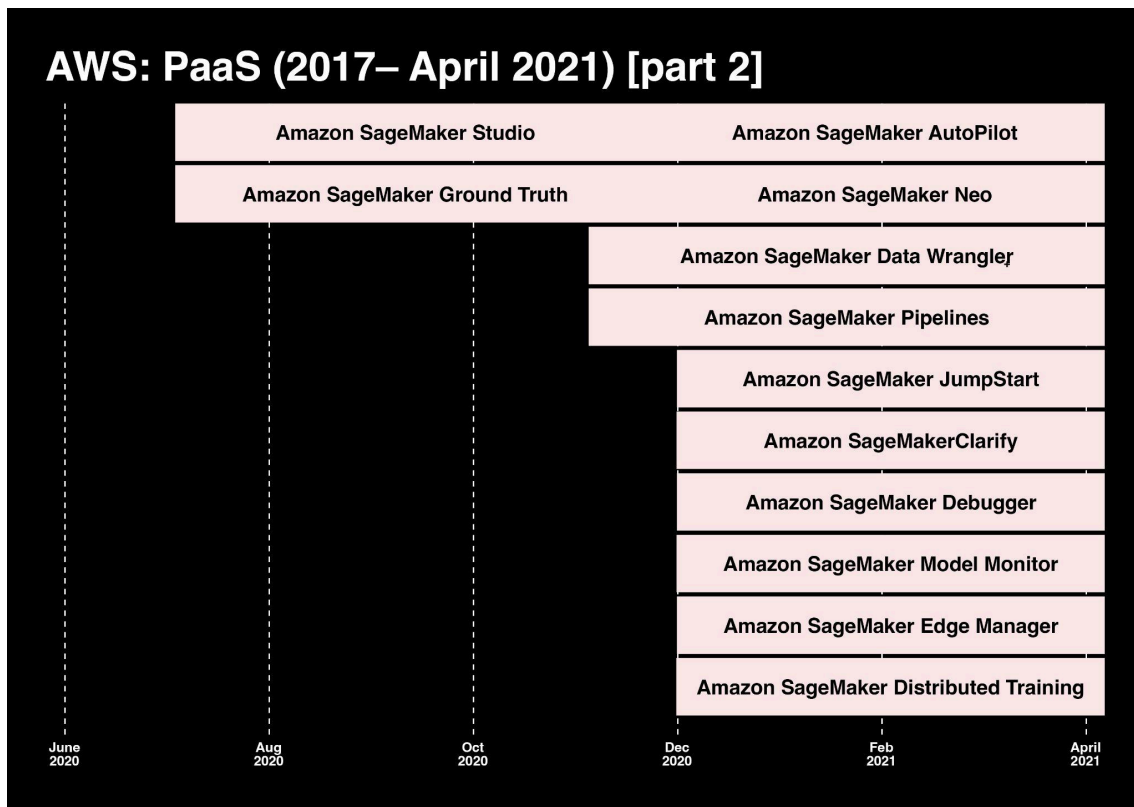


**FIGURE 1:** Custom application-specific integrated circuits (ASIC) and field-programmable gate arrays (FPGA) developed and operated by AWS [pink], Microsoft Azure [blue], and Google Cloud [green] between January 2017 and April 2021.

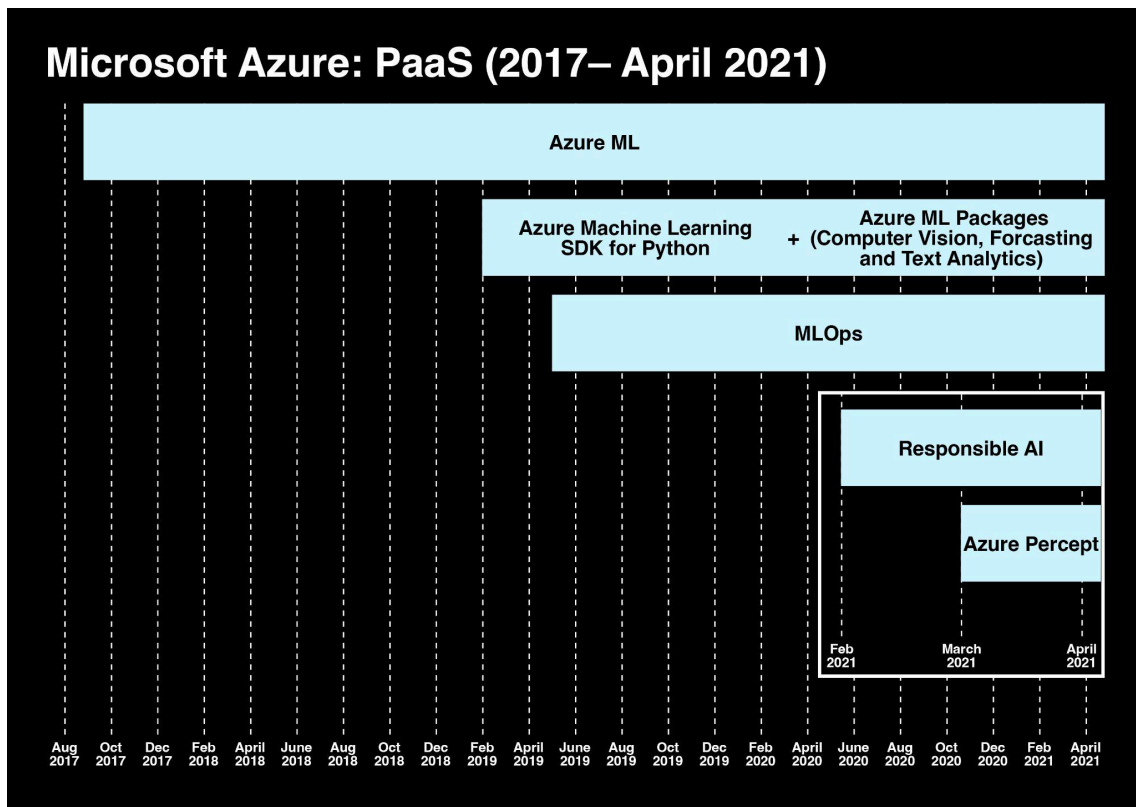
Additionally, Google has strategically tied the usage of its proprietary hardware infrastructure to its open-source machine-learning framework *TensorFlow*, which is specifically optimised for operating *TPUs* to reduce computing requirements (Srinicek, 2022). Like Google Cloud but differing from Microsoft Azure, AWS broke with the tradition of relying on Intel field-programmable gate arrays (FPGAs) and different types of Nvidia Graphics Processing Units (GPUs) through the release of its specialised processing chips in 2019: *AWS Inferentia* (Barr, 2019) and *Trainium* (Lardinois, 2020) (Figure 1). Respectively optimised for training and inference of deep-learning models at scale, these integrated processors power Amazon *Elastic Compute Cloud (EC2)* instances that are only accessible through the *Neuron SDK* (AWSN–2021) or development platform *SageMaker* (AWS–2017).



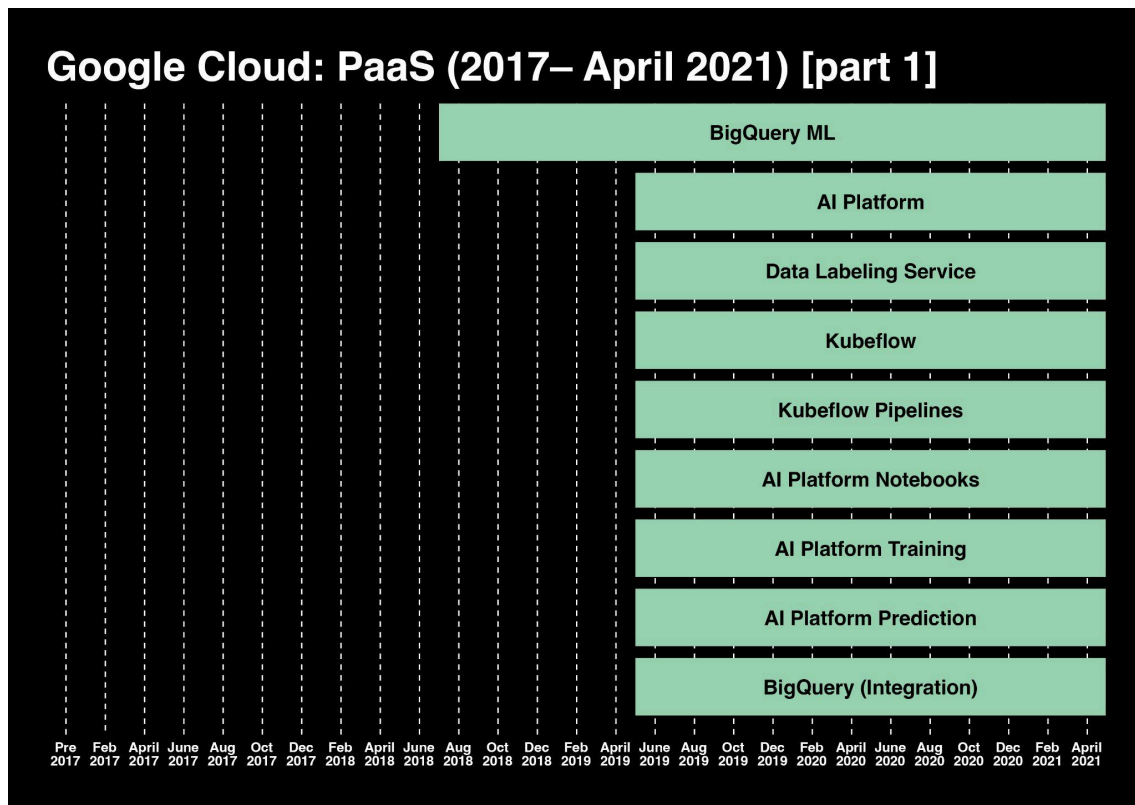
**FIGURE 2.1:** Excerpts of Cloud AI infrastructures and services operated by AWS at the platform-as-a-service (PaaS) level between January 2017 and April 2021 (part 1).



**FIGURE 2.2:** Excerpts of Cloud AI infrastructures and services operated by AWS at the platform-as-a-service (PaaS) level between January 2017 and April 2021 (part 2).



**FIGURE 3:** Excerpts of Cloud AI infrastructures and services operated by Microsoft Azure at the platform-as-a-service (PaaS) level between January 2017 and April 2021.



**FIGURE 4:** Excerpts of Cloud AI infrastructures and services operated by Google Cloud at the platform-as-a-service (PaaS) level between January 2017 and April 2021.

The power of vertical integration becomes particularly visible through the composition of such machine-learning development platforms. Next to Amazon *SageMaker* (Figure 1), Figures 3 and 4 show that this middle layer (PaaS) is occupied by *Azure Machine Learning* (2017) and Google Cloud's *AI platform* (2016). Over the last few years, these platforms have evolved into vast and fully managed development environments that provide third-party developers with the tools and services for data processing as well as model production, training, evaluation, deployment, and monitoring. Tables 2 and 3 provide an overview of the *SageMaker* and Google Cloud *AI platform* components. The specific set of Microsoft Azure platform services (Figure 3), however, remains underspecified even though these tools are part of integrated development environments (IDEs) such as *Visual Studio Code* (Lardinois, 2017). These integrated resources allow third parties to develop machine-learning models and deploy their applications all in one place. Yet, they also allow these cloud providers to increasingly manage, and thus control, a multitude of stages of machine-learning development pipelines, including the monitoring of applications through their corporate cloud ecosystems. This becomes visible through the collective push for services that facilitate managing entire machine-learning projects such as the MLOps tools illustrated in Figure 5. As an aggregate of machine learn-

ing and DevOps (Development and Operations), MLOps refers to a set of integrated practices that aim to shorten and simplify machine-learning development and operational cycles (Kepes, 2013). Yet, by doing so, MLOps also evoke further standardisation and automation, placing the platforms that operate them – AWS, Microsoft Azure, and Google Cloud – in the strategic position to exert their infrastructural power over these processes.

**TABLE 2:** Overview of *SageMaker* platform components as of April 2021. Source: Amazon Web Services (n.d.)

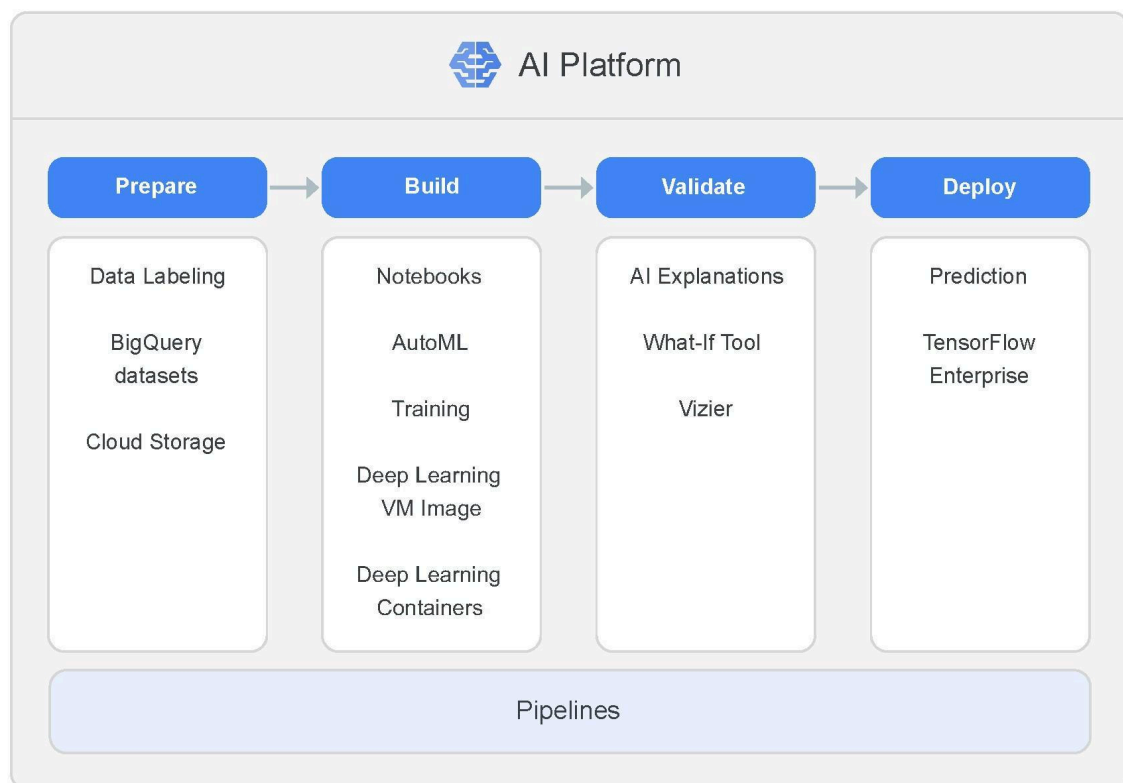
SAGEMAKER PRODUCT	DESCRIPTION
Automatic Model Tuning	Hyperparameter optimization
Built-in and Bring-your-own Algorithms	Dozens of optimized algorithms or bring your own
Distributed training libraries	Training for large datasets and models
Kubernetes & Kubeflow Integration	Simplify Kubernetes-based machine learning
Local Mode	Test and prototype on your local machine
Managed Spot Training	Reduce training cost by 90%
Multi-Model Endpoints	Reduce cost by hosting multiple models per instance
One-click Deployment	Fully managed, ultra low latency, high throughput
One-click Training	Distributed infrastructure management
SageMaker Autopilot	Automatically create machine learning models with full visibility
SageMaker Clarify	Detect bias and understand model predictions
SageMaker Data Wrangler	Aggregate and prepare data for machine learning
SageMaker Debugger	Debug and profile training runs
SageMaker Edge Manager	Manage and monitor models on edge devices
SageMaker Experiments	Capture, organize, and compare every step
SageMaker Feature Store	Store, update, retrieve, and share features
SageMaker Ground Truth	Label training data for machine learning
SageMaker JumpStart	Pre-built solutions for common use cases
SageMaker Model Monitor	Maintain accuracy of deployed models
SageMaker Pipelines	Workflow orchestration and automation
SageMaker Processing	Built-in Python, BYO R/Spark

SAGEMAKER PRODUCT	DESCRIPTION
SageMaker Studio	Integrated development environment (IDE) for ML
SageMaker Studio Notebooks	Jupyter notebooks with elastic compute and sharing

**TABLE 3:** Overview of Google Cloud *AI platform* components in April 2021. Source: Google Cloud (n.d.)

GOOGLE CLOUD AI PLATFORM PRODUCT	DESCRIPTION
AI Explanations	Understand how each feature in your input data contributed to model's outputs
AutoML	Easily develop high-quality custom machine learning models without writing training routines. Powered by Google's state-of-the-art transfer learning and hyperparameter search technology.
Continuous evaluation	Obtain metrics about the performance of your models in production. Compare predictions with ground truth labels to gain continual feedback and optimize model performance over time.
Data Labeling Service	Get highly accurate labels from human labelers for better machine learning models.
Deep Learning Containers	Quickly build and deploy models in a portable and consistent environment for all your AI applications.
Deep Learning VM Image	Instantiate a VM image containing the most popular AI frameworks on a Compute Engine instance without worrying about software compatibility.
Neural Architecture Search	Build application-specific models and improve existing model architectures with an automated service. Powered by Google's leading AI research, users can design models that are optimized for latency, accuracy, power consumption, and more.
Notebooks	Create, manage, and connect to VMs with JupyterLab, the standard data scientist workbench. VMs come pre-installed with deep learning frameworks and libraries.
Pipelines	Implement MLOps by orchestrating the steps in your ML workflow as a pipeline without the difficulty of setting up Kubeflow Pipelines with TensorFlow Extended (TFX).
Prediction	Easily deploy your models to managed, scalable endpoints for online or batch predictions.
TensorFlow Enterprise	Easily develop and deploy TensorFlow models on Google Cloud with enterprise-grade support and cloud scale performance.
Training	Train any models in any framework on any hardware, from single machines to

GOOGLE CLOUD AI PLATFORM PRODUCT	DESCRIPTION
	large clusters with multiple accelerators.
Vizier	Optimize your model's output by intelligently tuning hyperparameters.
What-if Tool	Visualize your datasets and probe your model to better understand its behavior with an interactive visual interface.



**FIGURE 5:** Overview of Google Cloud AI Platform components according to the different stages of the machine-learning life cycle proposed by Google by April 2021. Source: Google Cloud (n.d.).

Lastly, at the top layer of the cloud stack, infrastructural power is operated through the rapid availability of vertically integrated, ready-to-deploy pre-trained models for particular inference tasks such as Amazon *Recognition*; Azure *Vision API*, and Google *Vision API* for image recognition (AWS-2021; MA-2021a; GCP-2021a); automated speech recognition (Amazon *Lex* and Google *Speech API*); or content moderation (Azure *Form Recognizer* and *Immersive Reader*) (MA-2021a; MA-2021b). Respectively available as *AI services* (AWS), *Cognitive Services* (Microsoft Azure), and various APIs (Google Cloud) visualised in Figures 6, 7 and 8, these pre-trained models are integrated within proprietary ecosystems and leveraged on a service

basis through paid APIs<sup>4</sup> available on all cloud computing platforms (AWS–2021; MA–2021a; GCP–2021a). As such, these models are leveraged as closed systems that operate according to the standards of AWS, Microsoft Azure, or Google Cloud, running on infrastructures from different cloud AI branches ranging from hardware (e.g. *Inferentia*) to data storage (e.g. *Amazon S3*, Google’s *BigQuery*). The “gateway function” (van der Vlist & Helmond, 2021, p. 13) of APIs allows these companies to strengthen their position as powerful intermediaries by governing the accessibility of pre-trained models as well as their use by third-party developers for the development of specific applications.

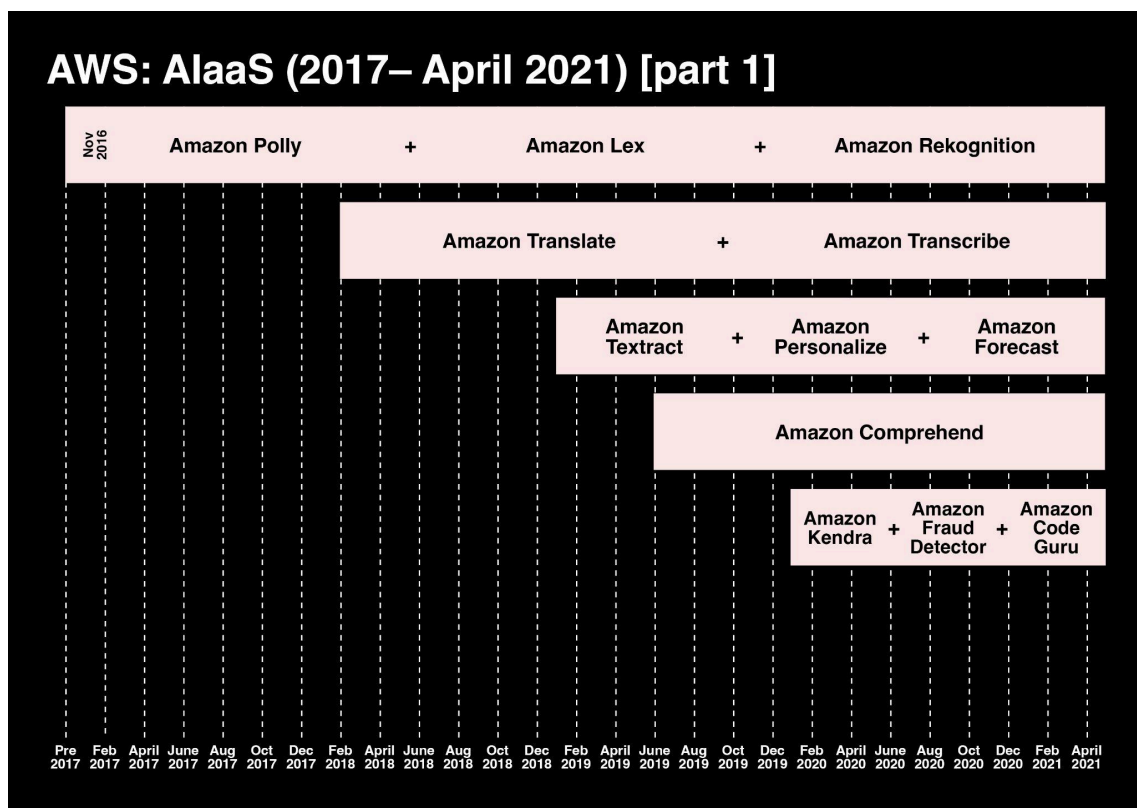
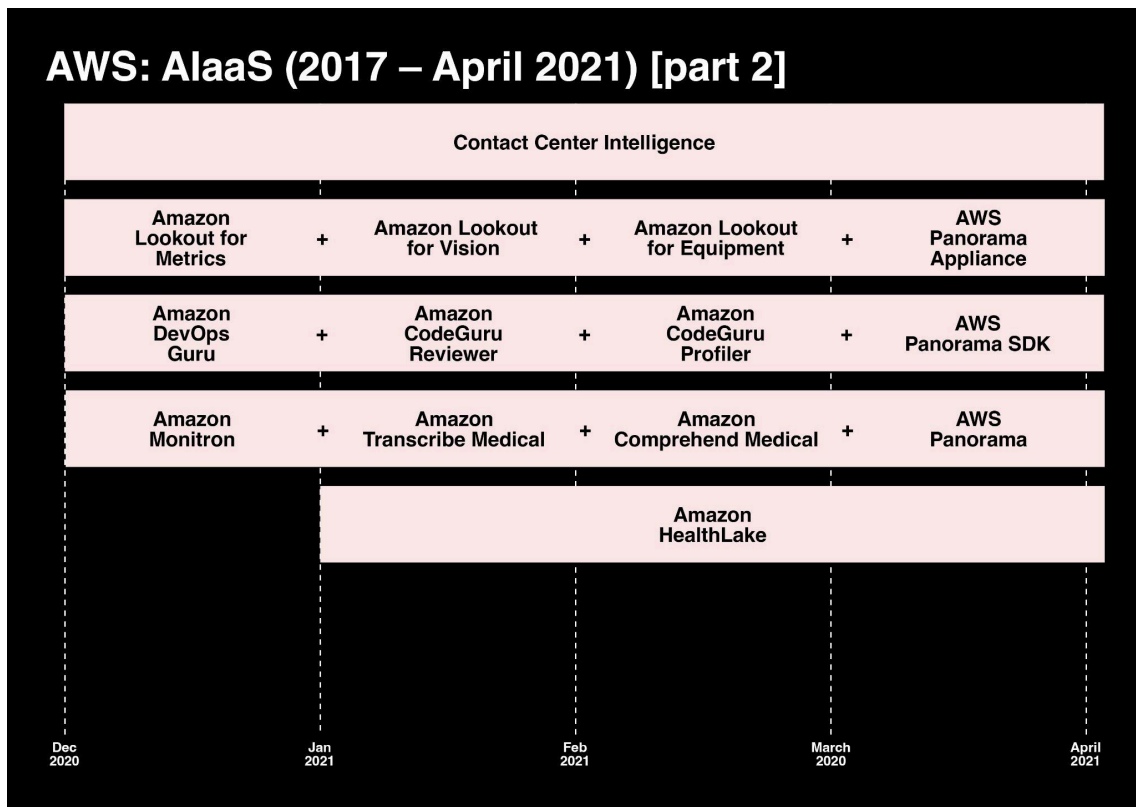


FIGURE 6.1

4. With paid APIs, I refer to APIs that operate on a pay-as-you-use basis, following a pricing plan set by cloud platforms. See for example Google Cloud, (2024)



**FIGURE 6.2. 6.1;6.2:** Excerpts of Cloud AI infrastructures and services operated by AWS at the AI-as-a-service (AlaaS) level between January 2017 and April 2021.

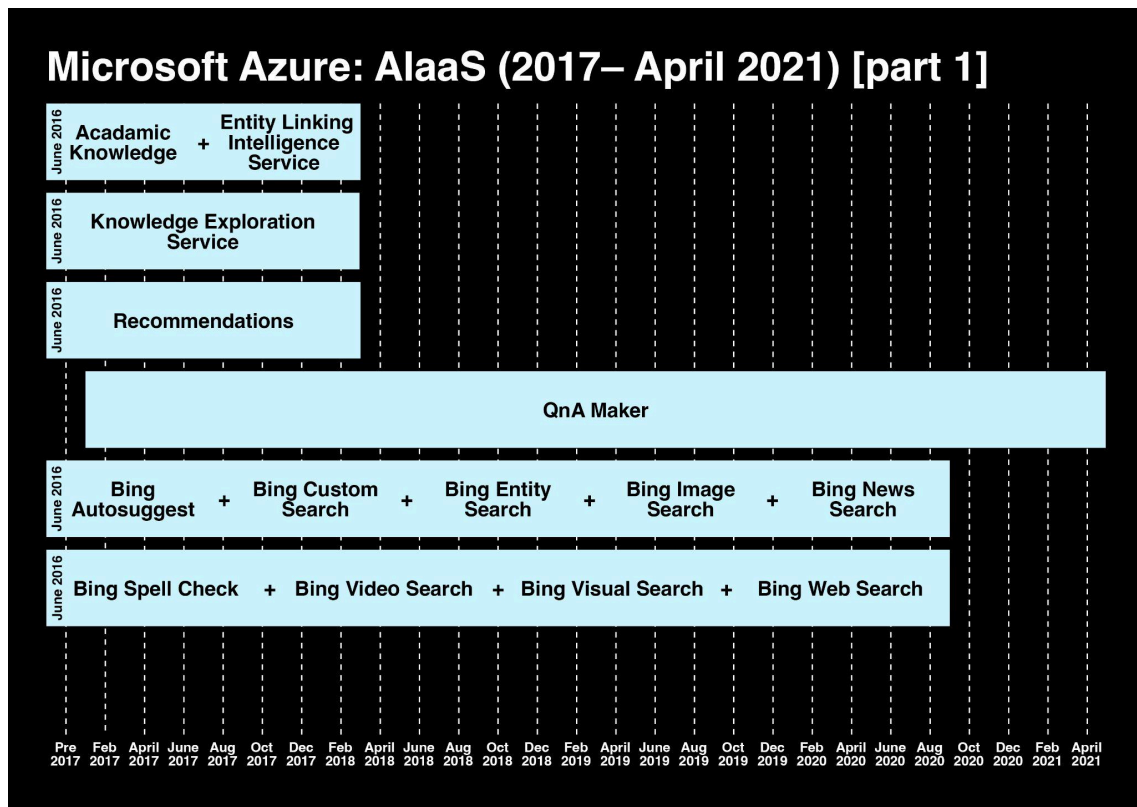


FIGURE 7.1

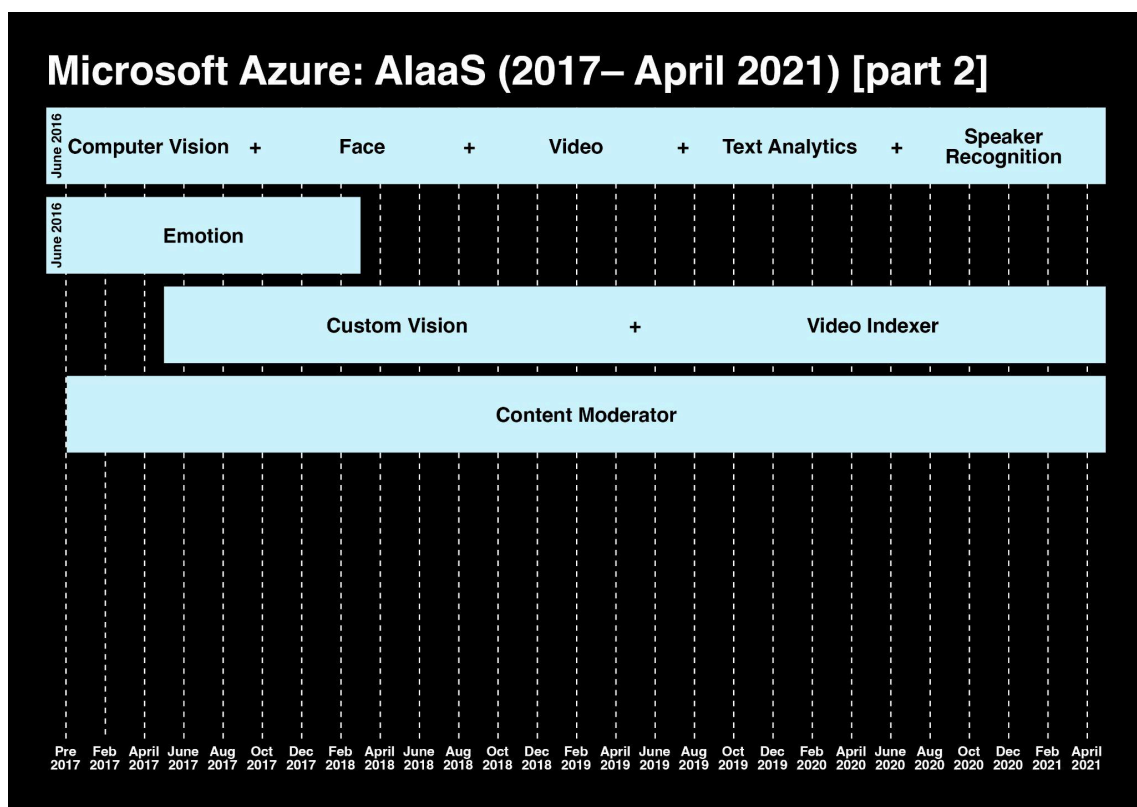


FIGURE 7.2

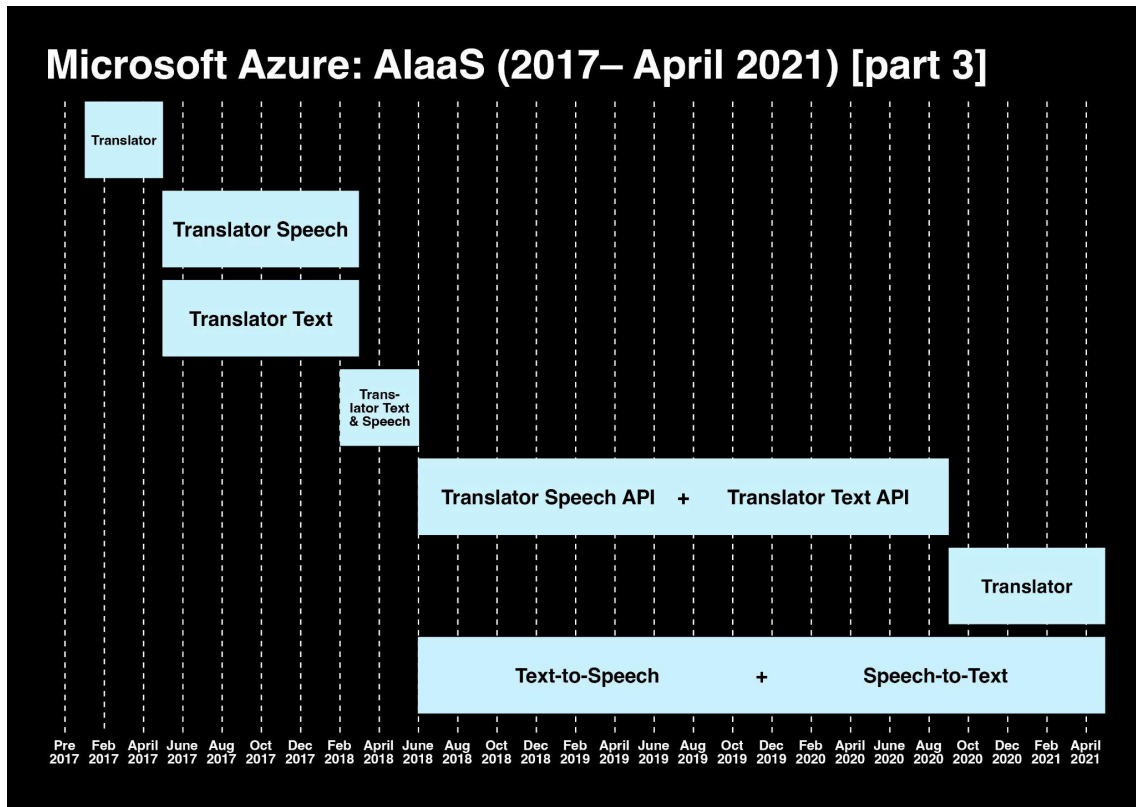


FIGURE 7.3

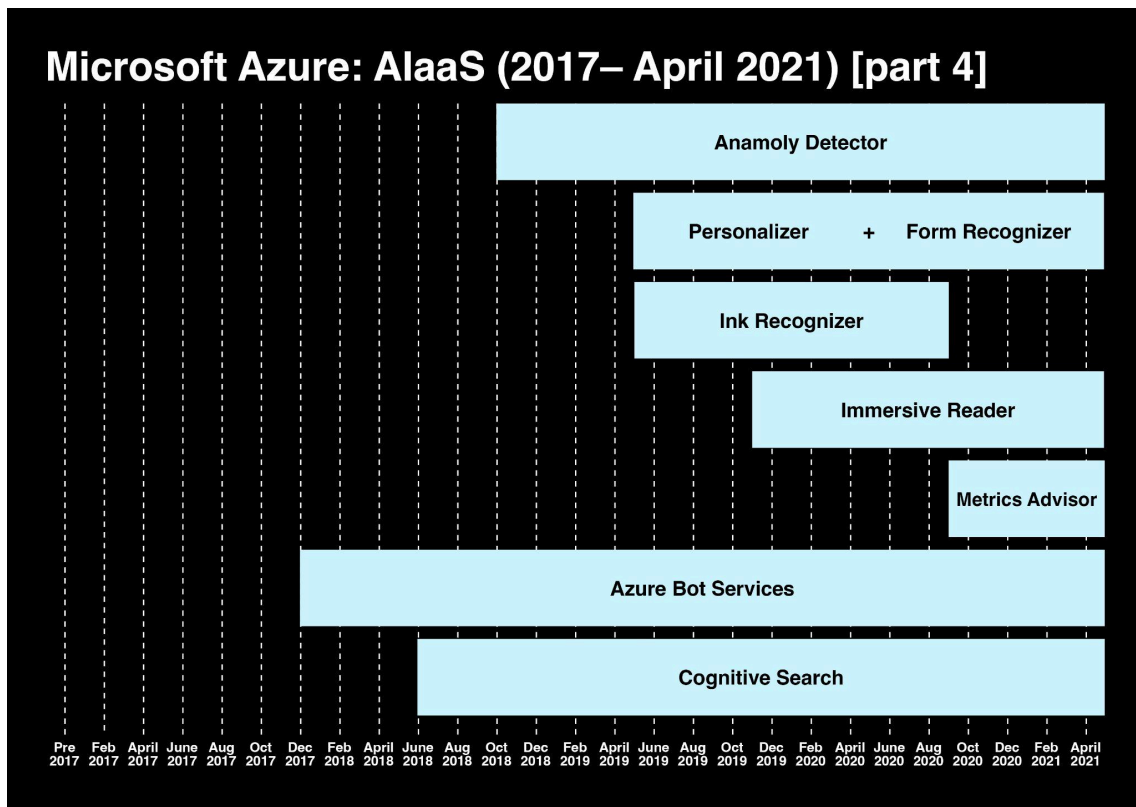
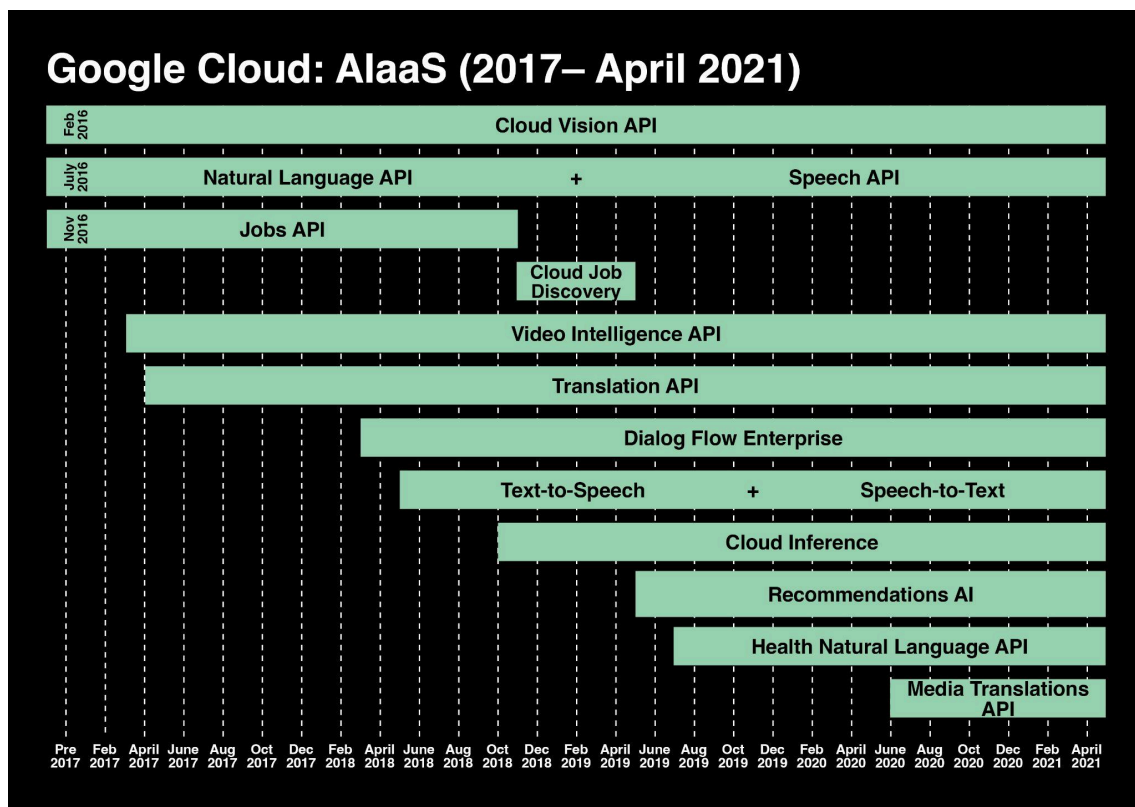


FIGURE 7.4. 7.1–7.4: Excerpts of Cloud AI infrastructures and services operated by Microsoft Azure at the AI-as-a-service (AlaaS) level between January 2017 and April 2021.



**FIGURE 8:** Cloud AI infrastructures and services operated by Google Cloud at the AI-as-a-service (AlaaS) level between January 2017 and April 2021.

Taken together, this growing operation of vertically integrated infrastructures and services, arguably allows AWS, Microsoft Azure, and Google Cloud to strategically evolve in ways to further expand their operative control over chains of valuable components – the means of production – for AI production and deployment. By doing so, corporate clouds do not only mobilise their infrastructural components to lower the access barriers to machine-learning resources which enable developers “to step further faster,” but actively condition the set of technologies in terms of “what can be considered possible in the first place” (Rieder, 2020, p. 16). Tying these elements together through pricing architectures (e.g. Khan, 2018), exclusive agreements and strategic partnerships (Vipra & Myers West, 2023), vertically integrated services facilitate further standardisation and privatisation of machine-learning workflows. From model production and evaluation to deployment and monitoring – all these steps can be configured and optimised through the ecosystems of AWS, Microsoft Azure, or Google Cloud. Like with other software tools operated in platform ecosystems (e.g. Foxman, 2019), such deep integrations of hardware and software risk leading to path dependencies and lock-in for developers looking to adopt machine learning into their applications. Once customers are invested in the tools and services operated by one of the major cloud providers, it

becomes very difficult to switch vendors without substantial investment in time, financial, and computational resources (Vipra & Myers West, 2023). The sets of vertically integrated infrastructures and services can thus be seen to function as strategically employed “engines of profit” (Whittaker, 2021b) for cloud platforms like AWS, Microsoft Azure and Google Cloud. These companies, in turn, not just attempt to operate infrastructural power through a compute divide in AI (Ahmed & Wahed, 2020) but through strategically tethered collections of integrated hardware infrastructures and software services that cover entire machine-learning development cycles.

## Complementary innovation

Locating the infrastructural power of AWS, Microsoft Azure, and Google Cloud in their ability to facilitate “complementary innovation” (Gawer, 2014), the evolutionary technographic analysis reveals that these cloud platforms operate this form of power in distinct ways. First, I find that the infrastructural “cores” of cloud platforms are continuously being leveraged by “breaking up, decomposing, and recomposing [these] existing digital components” (Aradau & Blanke, 2022, p. 104). These can then be further integrated into different societal domains such as healthcare. As Figures 6 and 8 show, AWS and Google Cloud have been harnessing their power in Natural Language Processing (NLP) by purposefully developing health-specific services such as Amazon *Comprehend Medical*, Amazon *Healthlake*, and Google *Healthcare Natural Language API*. These services for text analytics in the medical domain provide access to pre-trained language models that are specifically trained on health data. In contrast to its competitors, Google Cloud has taken this development even one step further by announcing controlled access to MED-PaLM 2 – an industry-tailored large language model (LLM) that aligns with the medical domain while relying on Google’s existing PaLM model (Gupta & Waldron, 2023). It is through such announcements that we can understand how Google Cloud is relying on its infrastructural power and expertise in AI development to enter the health-care industry (e.g. medical centres, pharmaceutical, and insurance companies) and shape the potential for applications from within the platform’s ecosystem.

In contrast to Google Cloud, AWS and Microsoft Azure have been positioning their AI infrastructures for the development of complementary services in the manufacturing and retail sectors. Visualised in Figure 6, Amazon *Monitron*, Amazon *Lookout for Equipment* and AWS *Panorama* represent a set of software and hardware services for assembly-line production, management, and remote operations originally designed for its warehouses (AWS–2021d). In addition, the platform launched AWS

*Panorama*, a hardware appliance and SDK to develop computer vision applications for workplace control (Shieber, 2020). Microsoft Azure launched a similar set of services more specifically for the retail industry through a platform called *Azure Percept* (MA–2021c; Figure 7). Complementary infrastructural expansions like these play to machine-learning systems' potential to serve a plurality of purposes (Rieder, 2022). As such, they allow AWS and Microsoft Azure to enter different legacy domains (cf. van Dijck, 2020) and arguably also provide these corporate clouds with the capacity to strategically condition the development and integration of AI systems in these respective areas.

In addition, cloud platforms arguably operationalise infrastructural power through complementary innovation by developing AI infrastructures and services that seamlessly operate with existing subsidiaries of Amazon, Microsoft, or Google platform ecosystems. This can be illustrated in the example of AWS which has emphasised that its *Monitron* service relies on the “same technology used to monitor equipment in Amazon Fulfilment Centers” (AWS–2020b). As Delfanti (2021) explains, Amazon distribution centres are heavily equipped with technologies to further optimise logistical operations and labour. In its attempt to optimise the maintenance of the in-house machinery, the company for example installed AI services like the *Monitron* system to monitor conveyor belts (Lee, 2020). Furthermore, AWS has been aligning the machine-learning services *SageMaker Groundtruth* and *Amazon Augmented AI* with the *Mechanical Turk* crowdwork marketplace. Launched in 2020 and 2021, these fully managed services allow third-party developers to outsource data labelling jobs or human review for (pre-trained or custom) model predictions to new or existing applications by connecting them to the *Mechanical Turk* workforce (Morton-Youmans & Gupta, 2020) or third-party vendors which are active on *AWS Marketplace*. While customers using the *A2I* service can also choose to work with their own employees, the service alignments with *Mechanical Turk* and the *AWS Marketplace* indicate another step by AWS to bring data processes within the platform's corporate control, thereby expanding its power over a significant part of the AI development pipeline.

Simultaneously, similar services have been developed by all cloud providers to address issues around AI ethics. Such tools include AWS' *Augmented AI*, Microsoft Azure's *Face API Transparency Note* (MA–2019), or Google Cloud's *What If Tool* (Robinson & Wexler, 2019) and are marketed with the promise to combat bias and to facilitate the development of “responsible systems”. However, as mentioned previously, the *Augmented AI* service is partly built on Amazon *Mechanical Turk* – the crowdworker marketplace that is known for the exploitation and the commodifica-

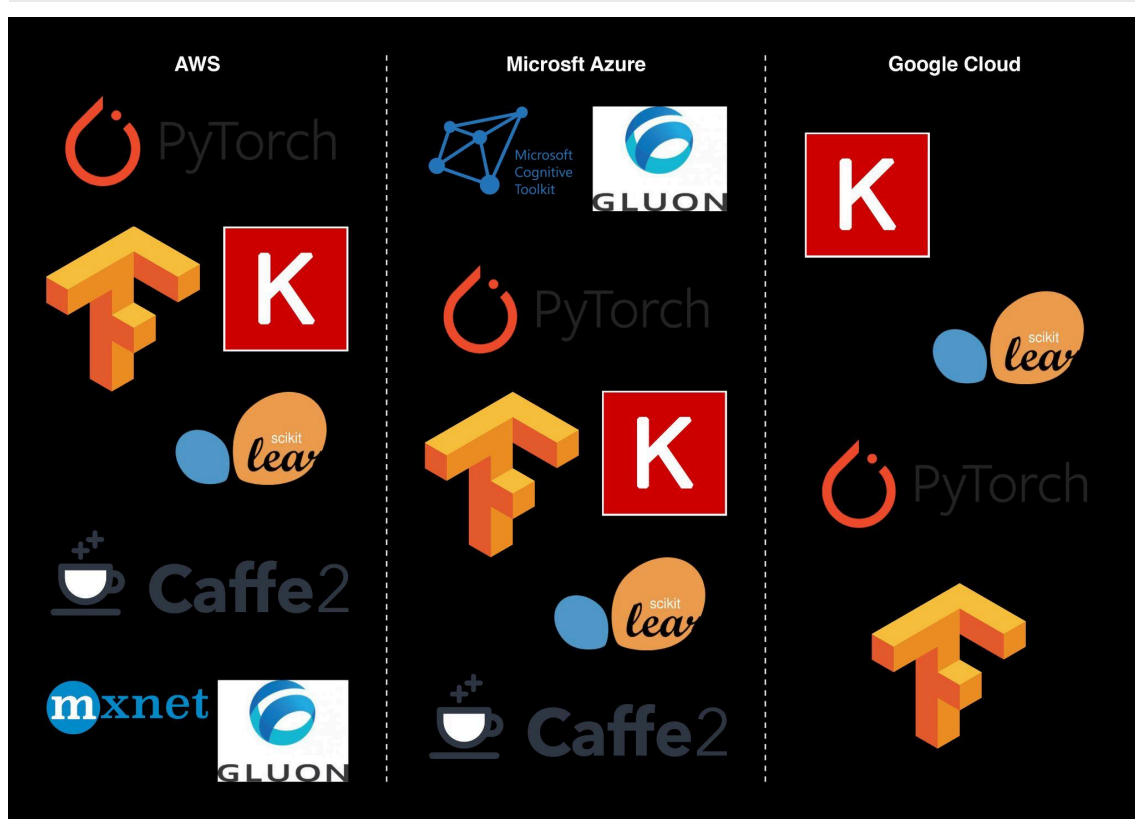
tion of data-related micro-labour in the field (e.g. Miceli et al., 2022). Microsoft Azure and Google Cloud frame their tools as technical solutions to facilitate model transparency but they can only be accessed through proprietary infrastructures controlled by these corporations. These examples hence substantiate critiques against corporate AI ethics initiatives circumscribed by the cloud platforms that dominate the field (Aradau & Blanke, 2022): the companies address issues of ethics and responsibility through self-regulating frameworks, but still choose how to deploy AI technologies and condition the meaning of ‘ethical’ and ‘responsible’ AI.

## Abstraction

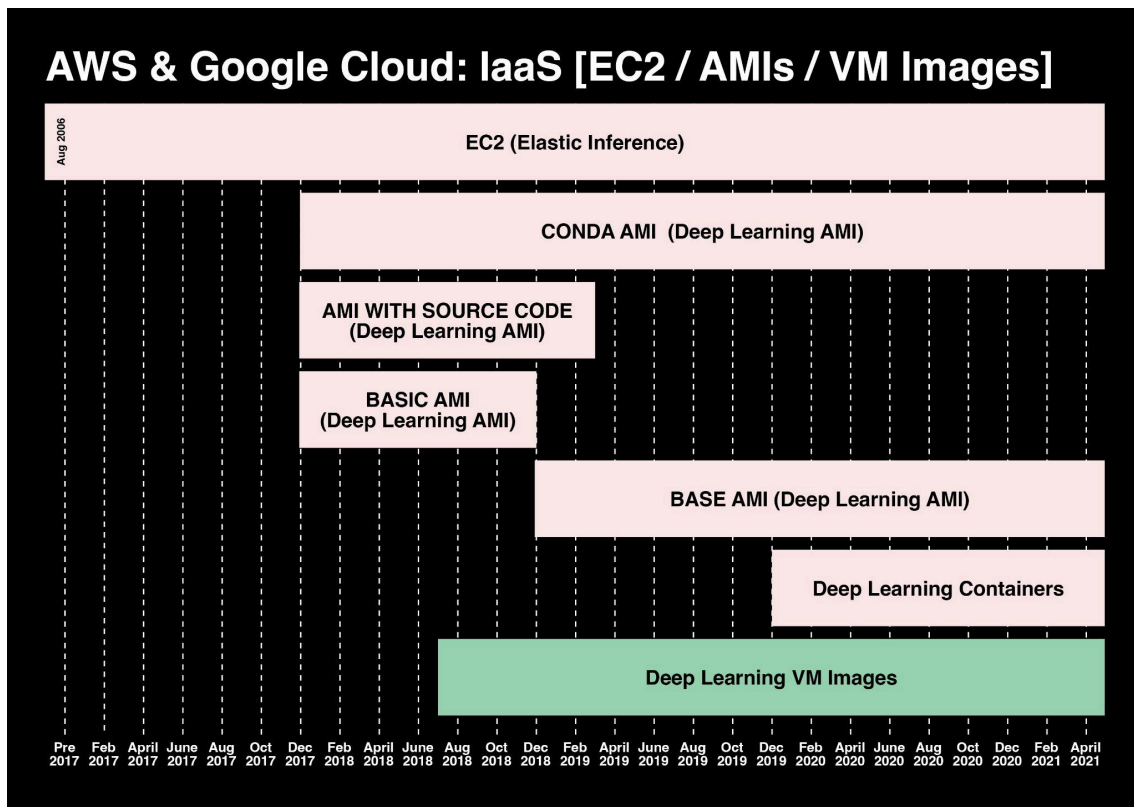
In addition to vertical integration and complementary innovation, AWS, Microsoft Azure, and Google Cloud attempt to exercise infrastructural power in AI’s political economy through the abstraction of AI infrastructures and services. In computing, the notion of abstraction refers to the practice of packaging complicated underlying operations into single commands, thereby “hiding” many technical complexities related to building, operating, and managing computational systems behind abstraction layers enabling developers to speed up working processes. As Selbst et al. (2018) explain: “abstractions are essential to computer science, and in particular machine learning” (p. 2) as they hide the domain-specific aspects of a machine-learning problem or task. By using abstraction layers, machine-learning tools remain modular and can be applied in proprietary as well as in open-source manners across different economic and social domains (Selbst et al., 2018). In the case of the variety of cloud AI infrastructures and services under study, different complex operations for the training and inference of machine-learning systems increasingly disappear behind abstraction layers to make it easier for third-party developers to adapt these functions as a service into their applications.

At the bottom and middle layers of the cloud stack, the evolutionary trajectories of AWS, Microsoft Azure, and Google Cloud show that these platforms operationalise the power of abstraction through the widespread provision of open-source machine-learning frameworks. As shown in Figure 9, AWS, Microsoft Azure, and Google Cloud all support their own set of frameworks which are pre-installed to their integrated machine-learning development platforms. Also visualised in Figure 10, AWS and Google Cloud offer developers direct access to frameworks through so-called Deep Learning Virtual Machine Images (*Deep Learning AMIs* and *Google VM Image*) (AWS–2021b; GCP–2021a). Open-source frameworks, Burkhardt (2019) explains, offer “predefined functions and functionalities [which] relieve de-

velopers from building software from the ground up” (p. 213) and have become crucial infrastructural elements for the development of AI systems today. Depending on developers’ individual preferences, frameworks offer multiple levels of abstraction that make it easier for developers to adapt to machine learning even if they do not have the specific expertise. The high-level framework *Keras* adds another layer of abstraction on top of *TensorFlow* and the *Microsoft Cognitive Toolkit*, supporting even more “easy and fast prototyping” (Rieder, 2020, p. 112). By offering integrated access to different frameworks with (multiple) built-in abstraction layers, AWS, Microsoft Azure, and Google Cloud thus simplify and speed up the AI production and deployment processes. However, these seemingly open-source infrastructures also enable these corporations to standardise and align such processes according to their proprietary interests to further drive AI’s deep-learning paradigm that relies on the cloud (Widder et al., 2023). Abstraction is hence strategically mobilised to exert infrastructural power over AI development processes to conform cloud platforms’ economic and political interests and shift attention away from other techniques and approaches that are less reliant cloud AI service models (Vipra & Myers West, 2023).



**FIGURE 9:** Open-Source machine-learning frameworks supported by the three major cloud platforms.

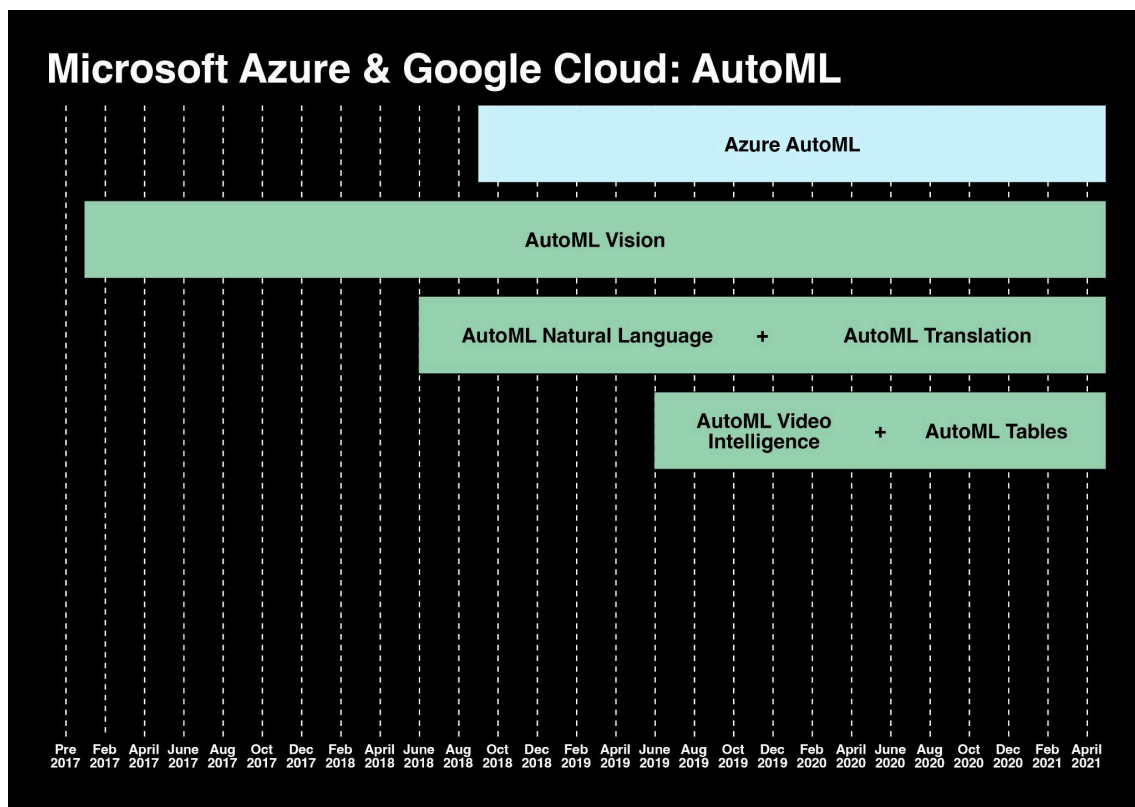


**FIGURE 10:** The evolution of Deep Learning Virtual Machine Images provided by AWS [pink] and Google Cloud [green] through the IaaS-layer of their cloud ecosystem.

Microsoft operates another layer of abstraction through its Azure Machine Learning Platform (*Azure ML*) (Figure 3), where it provides a no-code integrated development environment (IDE). This *Machine Learning Designer* IDE (MA-2020c) hides the underlying code layers behind a visual interface which customers can use to build, deploy, and manage machine-learning models. Instead of programming, customers without any code experience are encouraged to set up entire machine-learning pipelines by dragging and dropping datasets, pre-trained models, and development tool sets for data preparation, model training, and evaluation (Zhang, 2020). While the introduction of this IDE claimed to make machine learning more accessible for those with little or no programming experience (MA-2020c), Microsoft Azure arguably seeks to exercise power through the abstraction of machine-learning production and deployment processes: by hiding the code behind abstraction layers, customers can no longer modify or reprogramme the selected elements according to their specific preferences. Instead, they become dependent on the selection of machine-learning models and tools that Microsoft Azure provides without being able to critically verify the technical functioning of the individual elements at play. In addition, customers who rely on the *Azure Machine Learning Designer* are required to use the Azure Kubernetes Service to run their models (MA-2020c).

Hence, they lose the autonomy and flexibility of choosing and (re)configuring their deployment infrastructure as they are being locked into the Microsoft Azure cloud ecosystem.

The cloud providers' strategy to package complicated underlying technical operations of AI systems and infrastructures becomes even more clear observing the large-scale offering of large pre-trained models as AI services via paid APIs. The different services for natural language processing (NLP), for example, enable developers to directly run these pre-trained models for tasks such as text generation or translation or use them as essential building blocks for application development. However, while these technologies are programmable in the sense that developers can integrate machine-learning functionalities into their own applications, the pre-trained models themselves cannot externally be viewed, evaluated, or modified. Instead, their complex operations have disappeared behind abstraction layers hiding the technical functioning as well as the principles governing the use of these services. Consequently, developers who deploy pre-trained models are required to operate their applications within the tight boundaries specified and controlled by the operators of these services. This has been raising significant concerns among members of the critical AI research community as it further increases the power of cloud platforms to shape AI technologies while undermining calls to mitigate privacy issues (e.g. Powles & Hudson, 2017) as well as the bias and discrimination enforced by these systems (e.g. Bender et al., 2021).



**FIGURE 11:** AutoML services developed and operated by Microsoft Azure [blue] and Google Cloud [green].

The abstraction of complex machine-learning development mechanisms is even taken one step further through the development of services for automated machine learning (AutoML) in ways which are arguably even more particular to cloud AI spaces. Microsoft Azure and Google Cloud specifically evolved into this area by expanding their portfolio of proprietary AutoML services for the development of custom models into domains such as computer vision, video analysis, language generation, and translation (GCP–2021a; Figure 11). These services are claimed to automatically develop and test neural networks architectures (Thomas, 2018) and tend to abstract almost every aspect of machine-learning pipelines; from model construction, training and evaluation, to model deployment in specific settings and maintenance. Targeted at a developer base without the necessary expertise, these cloud platforms provide an abstracted plug-and-play version of machine learning. At the same time, AutoML systems remain limited in the scope of the problems they are claimed to solve and the amount of feedback they provide to the user. Google Cloud AutoML for example, is only available through an API which developers can access to query on new inputs (Liang et al., 2019). This means that developers have very little to no control over model production, training and valida-

tion processes. AutoML systems can thus fully benefit from the power of abstraction, imposing functional logic and practical affordances on developers that have been defined by Google's AI research team (Thomas, 2018).

In conclusion, the abstraction of AI systems and their underlying infrastructures arguably enables AWS, Microsoft Azure, and Google Cloud to further consolidate their infrastructural power in the political economy of AI. However, this form of infrastructural power is operationalised in different ways. First, the open-source frameworks that cloud platforms provide all offer multiple levels of abstraction that make it easier for developers to build, train, and deploy their machine-learning models which "appear to be applicable to problems in a variety of social settings" (Selbst et al., 2018, p. 4) at greater speed. Subsequently, AWS, Microsoft Azure, and Google Cloud mobilise this power of abstraction to attract more users from different sectors into proprietary AI ecosystems, channelling and structuring their activities according to their strengths and benefits. Second, the introduction of no-code IDEs by Microsoft and the growing availability of pre-trained models by all cloud providers shows how these platforms increasingly operate abstraction layers with the attempt to standardise and control a variety of processes across machine-learning pipelines. This, however, becomes even more specifically visible through the release of AutoML services which abstract entire machine-learning workflows into plug-and-play environments that fully rely on the cloud. As a result, third-party institutions or companies are required to develop their AI applications within the framework of these cloud providers and have very little ability to evaluate the systems they use to automatically produce or run for further application development. These organisations are drawn into a market without a basic understanding of the potential risks of the systems they operate and integrate into their respective applications even though there is significant evidence of the need to be critically aware (e.g., Miceli et al., 2022).

## Conclusion

The evolutionary technographic analysis of AWS, Microsoft Azure, and Google Cloud's AI infrastructures and services demonstrates that these major cloud providers operationalise infrastructural power in three substantial and complementary ways. First, the consistent vertical integration of AI infrastructures and services that are operated across the multi-layered stack of cloud architectures (IaaS, PaaS, and SaaS) shows that cloud providers attempt to exercise infrastructural power over entire AI systems and application development pipelines. This includes data storage and processing to model production, training, evaluation, de-

ployment, and the integration of systems for the production of specific applications. On this level, this research contributes to literature on the political economy of AI which emphasises that the power of big tech in AI is primarily established through the unequal distribution of computing resources (Whittaker, 2021a). Companies seek, not just to gain infrastructural power through a compute divide, (Ahmed and Wahed, 2020) but through the set of vertically integrated infrastructures and distributed services that strategically tether entire AI systems and application-development cycles within cloud platforms' respective ecosystems.

Second, AWS, Microsoft Azure, and Google Cloud attempt to consolidate their infrastructural power in AI through complementary innovation as they increasingly mobilise the transversal characteristics of the machine-learning systems they operate in two directions. On the one hand, they facilitate the strategic development of complementary specialised services to infrastructurally expand themselves into different application domains, such as healthcare, manufacturing, or retail. Through the development of their infrastructures and services AWS, Microsoft Azure, and Google Cloud actively set the conditions of possibility for AI development in these respective areas moving forward. On the other hand, the analysis showed that new machine-learning capabilities operate in complementary ways that seamlessly align with other branches in the larger ecosystems of Amazon, Microsoft, and Google such as *Amazon Mechanical Turk*. Most strikingly however, is that precarious workforces for data labour such as the micro-workers active on this Amazon platform (e.g. Miceli et al., 2022) are mobilised to substantiate services that are supposed to drive the development of "ethical and responsible AI". Power, in this case, is operationalised through infrastructure to set the standards for ethical frameworks that fully align with the interests of corporate cloud platforms such as AWS, Microsoft Azure, and Google Cloud (Aradau & Blanke, 2022).

Third, AWS, Microsoft Azure, and Google Cloud attempt to strategically exert 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 showed how these platforms mobilise and further develop their abilities to hide the complex operations of their cloud infrastructures and services across the various layers of the stack. More specifically, AutoML services abstract machine-learning workflows into plug-and-play environments that fully depend on the cloud. This creates another layer of significant advantages for cloud platforms to structure AI production in ways that further reinforce the creation of cloud dependent AI systems and applications. In addition, the widespread operationalisation of abstraction thwarts the critical

scrutiny and evaluation of AI systems – particularly in the case of AutoML – even though there have been ever more calls for critical oversight (e.g. Kak & Myers West, 2023). As such, the strategic mobilisation of abstraction can further strengthen AWS, Microsoft Azure, and Google Cloud in their position to operate infrastructural power. This allows them to shift the focus away from alternative resources and the development of new approaches that contribute to different understandings about AI technologies outside of the confining ecosystems of the cloud.

Taken together, *vertical integration*, *complementary innovation*, and *abstraction* as three central but interconnected sources of infrastructural power held by big tech in AI's political economy further stress the importance of recent calls for regulatory intervention by public authorities such as the FTC and the European Commission (Kak & Myers West, 2023). A growing number of steps have been taken in these directions (e.g. Khan, 2023), particularly since the launch of foundation models and the corporate ambitions to mobilise them in domains such as healthcare. However, focus on individual applications such as ChatGPT, or AI systems that have been developed for specific domains such as Google Cloud's MedPaLM risks overlooking how they are part of evolving infrastructures and services for cloud AI strategically operated in Amazon, Microsoft, and Google's ecosystems (see also van der Vlist et al., 2024). These systems thus require a deeper understanding of their operative politics and power in their integrated forms conditioned by corporate aims to (re)enforce capitalist relations at scale. The material-infrastructural approach to studying the evolutions of cloud platforms and the specific forms of infrastructural power in AI's political economy that I put forward in this paper provides ground to develop such comprehensive frameworks further.

There is a critical need for additional empirical research on the infrastructural power of cloud platforms and its specific implications for situated AI applications to substantiate such work and regulatory approaches. Acknowledging that AI operates in a dynamic ecosystem, such studies could further explore the strategies and operational logics of other significant firms such as Alibaba Cloud, Tencent Cloud, Oracle, IBM, and Nvidia. While scholars have extensively researched the implications of platform power for different areas of the economy and society (e.g. Lomborg et al., 2024), the immense scale at which these corporate clouds operate to govern and direct AI system and application development across domains and global contexts remains largely understudied. This is of particular concern considering the rapid implementations of generative AI. This paper therefore provides a critical empirical approach to further research AI and its political economy within

the context of corporate cloud platforms as vast and complex conglomerates of hardware infrastructures and software services that operate across the stack. Cloud platforms' (archived) materials and documentation afford valuable information into their technical operations as well as their governance structures which can be used to trace shifting relations of power in AI production and deployment. It is only through such a detailed understanding of how these infrastructures are strategically mobilised and operated that academic researchers and regulatory institutions can further deepen their capacities to effectively intervene.

---

## References

Ahmed, N., & Wahed, M. (2020). *The de-democratization of AI: Deep learning and the compute divide in artificial intelligence research* (arXiv:2010.15581). arXiv. <https://doi.org/10.48550/arXiv.2010.15581>

Amazon Web Services. (n.d.). *Amazon SageMaker*. <https://aws.amazon.com/sagemaker/>

Aradau, C., & Blanke, T. (2022). *Algorithmic reason: The new government of self and other*. Oxford University Press. <https://doi.org/10.1093/oso/9780192859624.001.0001>

Barr, J. (2019, December 3). Amazon EC2 update – Inf1 instances with AWS inferentia chips for high performance cost-effective inferencing. *AWS News Blog*. <https://aws.amazon.com/blogs/aws/amazon-ec2-update-inf1-instances-with-aws-inferentia-chips-for-high-performance-cost-effective-inferencing/>

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2022). *On the opportunities and risks of foundation models* (arXiv:2108.07258). arXiv. <https://doi.org/10.48550/arXiv.2108.07258>

Bratton, B. H. (2016). *The stack: On software and sovereignty*. MIT Press. <https://doi.org/10.7551/mitpress/9780262029575.001.0001>

Bucher, T. (2012). *Programmed sociality: A software studies perspective on social networking sites* [Doctoral dissertation, University of Oslo]. [https://tainabucher.com/wp-content/uploads/2009/08/Bucher\\_Ph.D.diss\\_.pdf](https://tainabucher.com/wp-content/uploads/2009/08/Bucher_Ph.D.diss_.pdf)

Bucher, T. (2018). *If... Then: Algorithmic power and politics*. Oxford University Press. <https://doi.org/10.1093/oso/9780190493028.001.0001>

Burckhardt, M. (2019). Mapping the democratization of AI on GitHub: A first approach. In A. Sudmann (Ed.), *The democratization of artificial intelligence: Net politics in the era of learning algorithms* (pp. 209–222). transcript-Verlag. <https://doi.org/10.1515/9783839447192-013>

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W.,

- Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, S., Maynez, J., Rao, A., Barnes, P., Tay, Y., Shazeer, N., Prabhakaran, V., ... Fiedel, N. (2022). *PaLM: Scaling language modeling with pathways* (arXiv:2204.02311). arXiv. <https://doi.org/10.48550/arXiv.2204.02311>
- Delfanti, A. (2021). Machinic dispossession and augmented despotism: Digital work in an Amazon warehouse. *New Media & Society*, 23(1), 39–55. <https://doi.org/10.1177/1461444819891613>
- Foxman, M. (2019). United we stand: Platforms, tools and innovation with the unity game engine. *Social Media + Society*, 5(4), 1–10. <https://doi.org/10.1177/2056305119880177>
- Galloway, A. R. (2004). *Protocol: How control exists after decentralization*. MIT Press. <https://doi.org/10.7551/mitpress/5658.001.0001>
- Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, 43(7), 1239–1249. <https://doi.org/10.1016/j.respol.2014.03.006>
- Ghazawneh, A., & Henfridsson, O. (2013). Balancing platform control and external contribution in third-party development: The boundary resources model. *Information Systems Journal*, 23(2), 173–192. <https://doi.org/10.1111/j.1365-2575.2012.00406.x>
- Google Cloud. (2024). *Google cloud pricing calculator*. <https://cloud.google.com/products/calculator-legacy/#tab=healthcare-api>
- Google Cloud. (n.d.). *AI platform*. <https://web.archive.org/web/20210408153644/https://cloud.google.com/ai-platform/#section-6>
- Gupta, A., & Waldron, A. (2023, April 13). A responsible path to generative AI in healthcare. *Google Cloud Blog*. <https://cloud.google.com/blog/topics/healthcare-life-sciences/sharing-google-med-pal-m-2-medical-large-language-model>
- Helmond, A., & van der Vlist, F. N. (2021). Platform and app histories: Assessing source availability in web archives and app repositories. In D. Gomes, E. Demidova, J. Winters, & T. Risse (Eds.), *The past web: Exploring web archives*. Springer. [https://doi.org/10.1007/978-3-030-63291-5\\_16](https://doi.org/10.1007/978-3-030-63291-5_16)
- Jacobides, M. G., Brusoni, S., & Candelon, F. (2021). The evolutionary dynamics of the artificial intelligence ecosystem. *Strategy Science*, 6(4), 412–435. <https://doi.org/10.1287/stsc.2021.0148>
- Kak, M., & Myers West., S. (2023). *AI Now 2023 landscape: Confronting tech power* [Report]. AI Now Institute. <https://www.ainowinstitute.org/2023-landscape>
- Kepes, B. (2013, December 18). *What does enterprise grade really mean?* *Forbes*. <https://www.forbes.com/sites/benkepess/2013/12/18/what-does-enterprise-grade-really-mean/>
- Khalili, L. (2018). The infrastructural power of the military: The geoeconomic role of the US Army Corps of Engineers in the Arabian Peninsula. *European Journal of International Relations*, 24(4), 911–933. <https://doi.org/10.1177/1354066117742955>
- Khan, L. M. (2018). Sources of tech platform power. *Georgetown Law Technology Review*, 2(2), 325–334.
- Khan, L. M. (2023, May 3). Lina Khan: We must regulate A.I. —Here's how. *The New York Times*. <https://www.nytimes.com/2023/05/03/opinion/ai-lina-khan-ftc-technology.html>
- Lardinois, F. (2017, September 25). Microsoft launches new machine learning tools. *TechCrunch*. <https://techcrunch.com/2017/09/25/microsoft-launches-new-machine-learning-tools/>
- Lardinois, F. (2020, December 1). AWS launches Trainium, its new custom ML training chip.

*TechCrunch*. <https://techcrunch.com/2020/12/01/aws-launches-tranium-its-new-custom-ml-training-chip/>

Lee, D. (2020, December 1). Amazon to roll out tools to monitor factory workers and machines. *Financial Times*. <https://www.ft.com/content/58bdc9cd-18cc-44f6-bc9b-8ca4ac598fc8>

Liang, J., Meyerson, E., Hodjat, B., Fink, D., Mutch, K., & Miikkulainen, R. (2019). Evolutionary neural AutoML for deep learning. *Proceedings of the Genetic and Evolutionary Computation Conference*, 401–409. <https://doi.org/10.1145/3321707.3321721>

Lomborg, S., Svendsen, K. S., Flensburg, S., & Lai, S. S. (2024). Monitoring infrastructural power: Methodological challenges in studying mobile infrastructures for datafication. *Internet Policy Review*, 13(2). <https://doi.org/10.14763/2024.2.1763>

Luitse, D. (n.d.). *The evolution of AWS, Microsoft Azure and Google 'Cloud AI' infrastructures and services (2017-2021)* [dataset]. Open Science Framework. <https://osf.io/7ut49>

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

Miceli, M., Posada, J., & Yang, T. (2022). Studying up machine learning data: Why talk about bias when we mean power? *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP), 1–14. <https://doi.org/10.1145/3492853>

Morton-Youmans, A., & Gupta, A. (2020, April 24). Amazon A2I is now generally available. *AWS Machine Learning Blog*. <https://aws.amazon.com/blogs/machine-learning/amazon-a2i-is-now-generally-available/>

Murgia, M. (2023, February 5). Big Tech companies use cloud computing arms to pursue alliances with AI groups. *Financial Times*. <https://www.ft.com/content/5b17d011-8e0b-4ba1-bdca-4fbfdbba10363>

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. <https://doi.org/10.1177/0308518X221094028>

Nieborg, D. B., & Helmond, A. (2019). The political economy of Facebook's platformization in the mobile ecosystem: Facebook Messenger as a platform instance. *Media, Culture & Society*, 41(2), 196–218. <https://doi.org/10.1177/0163443718818384>

Parsaeefard, S., Tabrizian, I., & Leon-Garcia, A. (2019). *Artificial intelligence as a service (AI-asS) on software-defined infrastructure*. 1–7. <https://doi.org/10.1109/CSCN.2019.8931372>

Pichai, S. (2017). *Google I/O'17 keynote* [Keynote address]. Google I/O, Mountain View, California. <https://www.youtube.com/watch?v=Y2VF8tmLFHw>

Plantin, J.-C. (2020, October). Platform logic and the infrastructural power of tech giants. *AoIR Selected Papers of Internet Research*. #AoIR2020: The 21st Annual Conference of the Association of Internet Researchers. <https://doi.org/10.5210/spir.v2020i0.11304>

Powles, J., & Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. *Health and Technology*, 7(4), 351–367. <https://doi.org/10.1007/s12553-017-0179-1>

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (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 University Press. <https://doi.org/10.2307/j.ctv12sdrv1>
- Rieder, B. (2022). Towards a political economy of technical systems: The case of Google. *Big Data & Society*, 9(2). <https://doi.org/10.1177/20539517221135162>
- Rikap, C. (2023). *Same end by different means: Google, Amazon, Microsoft and Meta's strategies to organize their frontier AI innovation systems* (Working Paper No. 2023-03). City Political Economy Research Centre (CITYPERC). <https://www.econstor.eu/handle/10419/280829>
- Robinson, S., & Wexler, J. (2019, July 19). Introducing the What-If Tool for cloud AI platform models. *Google Cloud Blog*. <https://cloud.google.com/blog/products/ai-machine-learning/introducing-the-what-if-tool-for-cloud-ai-platform-models>
- Sato, K., & Young, C. (2017, May 12). An in-depth look at Google's first Tensor Processing Unit (TPU). *Google Cloud Blog*. <https://cloud.google.com/blog/products/ai-machine-learning/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu>
- Seger, E., Ovadya, A., Siddarth, D., Garfinkel, B., & Dafoe, A. (2023). Democratising AI: Multiple meanings, goals, and methods. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 715–722. <https://doi.org/10.1145/3600211.3604693>
- Selbst, A. D., Boyd, D., Friedler, S., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. <https://doi.org/10.1145/3287560.3287598>
- Srnicek, N. (2022). Data, compute, labour. In M. Graham & F. Ferrari (Eds.), *Digital work in the planetary market* (pp. 241–261). MIT Press. <https://doi.org/10.7551/mitpress/13835.003.0019>
- Synergy Research Group. (2024). *Cloud market gets its mojo back; AI helps push Q4 increase in cloud spending to new highs* [Report]. <https://www.srgresearch.com/articles/cloud-market-gets-its-mojo-back-q4-increase-in-cloud-spending-reaches-new-highs>
- Thomas, R. (2018). *Google's AutoML: Cutting through the hype* [Report]. fast.ai. <https://www.fast.ai/2018/07/23/auto-ml-3/>
- van der Vlist, F., Helmond, A., & Ferrari, F. (2024). Big AI: Cloud infrastructure dependence and the industrialisation of artificial intelligence. *Big Data & Society*, 11(1). <https://doi.org/10.1177/20539517241232630>
- van der Vlist, F. N., & Helmond, A. (2021). How partners mediate platform power: Mapping business and data partnerships in the social media ecosystem. *Big Data & Society*, 8(1).
- van Dijck, J. (2020). Seeing the forest for the trees: Visualizing platformization and its governance. *New Media & Society*, 23(9), 2801–2819. <https://doi.org/10.1177/1461444820940293>
- van Dijck, J., Nieborg, D., & Poell, T. (2019). Reframing platform power. *Internet Policy Review*, 8(2). <https://doi.org/10.14763/2019.2.1414>
- Vipra, J., & Myers West, S. (2023). *Computational power and AI* [Report]. AI Now Institute. <https://ainowinstitute.org/publication/policy/compute-and-ai>
- Whittaker, M. (2021a). The steep cost of capture. *Interactions*, 28(6), 50–55. <https://doi.org/10.1145/3488666>
- Whittaker, M. (2021b, February 12). *AI and social control* [Keynote lecture]. <https://ruccs.rutgers.edu/t>

alks/list-of-past-events/icalrepeat.detail/2021/02/12/716/-/keynote-lecture-meredith-whittaker-on-ai-and-social-control

Widder, D. G., West, S., & Whittaker, M. (2023). *Open (for business): Big tech, concentrated power, and the political economy of open AI* (4543807). SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.4543807>

Zhang, L. (2020, September 30). Simplify and accelerate AI for the entire data science team with Azure Machine Learning designer. *Microsoft Azure Blog*. <https://techcommunity.microsoft.com/t5/azure-ai/>

## Appendices

### Appendix I

**TABLE 4:** Overview of the archived data sources. Pages are sorted by company and timeframe of information retrieval

CLOUD AI PLATFORM RESOURCE (TIMEFRAME: NUMBER OF (ARCHIVED) PAGES, ANALYSIS INTERVAL)	COMPANY	URL	TYPE
Machine Learning on AWS (January 2017–April 2021): 517 pages, monthly	Amazon	<a href="https://web.archive.org/web/20240000000000*/aws.amazon.com/machine-learning/">https://web.archive.org/web/20240000000000*/aws.amazon.com/machine-learning/</a>	Product page
Press release archive   Amazon.com, Inc. (January 2017–April 2021): 255 pages	Amazon	<a href="https://press.aboutamazon.com/press-release-archive">https://press.aboutamazon.com/press-release-archive</a>	Press Releases
Amazon SageMaker (November 2017–April 2021): 642 pages, monthly	Amazon	<a href="https://web.archive.org/web/20240000000000*/https://aws.amazon.com/sagemaker/">https://web.archive.org/web/20240000000000*/https://aws.amazon.com/sagemaker/</a>	Product page
AWS AI services (May 2019–April 2021): 104 pages, monthly	Amazon	<a href="https://web.archive.org/web/20240000000000*/https://aws.amazon.com/machine-learning/ai-services/">https://web.archive.org/web/20240000000000*/https://aws.amazon.com/machine-learning/ai-services/</a>	Product page
Azure Machine Learning   Microsoft Azure (January 2017–April 2021): 460 pages, monthly	Microsoft	<a href="https://web.archive.org/web/20240000000000*/http://azure.microsoft.com/services/machine-learning/">https://web.archive.org/web/20240000000000*/http://azure.microsoft.com/services/machine-learning/</a>	Product page
Cognitive Services   Microsoft Azure (January 2017–April 2021): 814 pages, monthly	Microsoft	<a href="https://web.archive.org/web/20240000000000*/azure.microsoft.com/services/cognitive-services/">https://web.archive.org/web/20240000000000*/azure.microsoft.com/services/cognitive-services/</a>	Product page
Microsoft news, features, events, and press materials (January 2017–April 2021): 336 pages	Microsoft	<a href="https://news.microsoft.com/">https://news.microsoft.com/</a>	Press Releases
Products and Services   Google Cloud (January 2017–April 2021): 1538 pages, monthly	Google	<a href="https://web.archive.org/web/20240000000000*/https://cloud.google.com/products/">https://web.archive.org/web/20240000000000*/https://cloud.google.com/products/</a>	Product page
Google Cloud Press Releases (January 2017–April 2021): 336 Pages	Google	<a href="https://cloud.google.com/press-releases/">https://cloud.google.com/press-releases/</a>	Press Releases
AI and Machine Learning Products   Google Cloud (July 2018–April 2021): 335 pages,	Google	<a href="https://web.archive.org/web/20240000000000*/cloud.google.com/products/">https://web.archive.org/web/20240000000000*/cloud.google.com/products/</a>	Product page

CLOUD AI PLATFORM RESOURCE (TIMEFRAME: NUMBER OF (ARCHIVED) PAGES, ANALYSIS INTERVAL)	COMPANY	URL	TYPE
monthly	ai/		

## Appendix II

**TABLE 5:** References to primary sources. Pages are sorted by company and date

CODE	COMPANY	TITLE	URL	ACCESS DATE
AWS-2015	Amazon	Introducing Amazon Machine Learning	<a href="https://aws.amazon.com/about-aws/whats-new/2015/04/introducing-amazon-machine-learning/">https://aws.amazon.com/about-aws/whats-new/2015/04/introducing-amazon-machine-learning/</a>	April 15, 2021
AWS-2016	Amazon	Introducing Amazon Lex, now in Preview	<a href="https://aws.amazon.com/about-aws/whats-new/2016/11/introducing-amazon-lex-now-in-preview/">https://aws.amazon.com/about-aws/whats-new/2016/11/introducing-amazon-lex-now-in-preview/</a>	April 15, 2021
AWS-2017	Amazon	Introducing Amazon SageMaker	<a href="https://aws.amazon.com/about-aws/whats-new/2017/11/introducing-amazon-sagemaker/">https://aws.amazon.com/about-aws/whats-new/2017/11/introducing-amazon-sagemaker/</a>	April 8, 2021
AWS-2020a	Amazon	Introducing AWS Panorama for computer vision at the edge	<a href="https://aws.amazon.com/about-aws/whats-new/2020/12/introducing-aws-panorama-for-computer-vision-at-the-edge/">https://aws.amazon.com/about-aws/whats-new/2020/12/introducing-aws-panorama-for-computer-vision-at-the-edge/</a>	April 10, 2021
AWS-2020b	Amazon	Introducing Amazon Monitron, an end-to-end system to detect abnormal equipment behavior	<a href="https://aws.amazon.com/about-aws/whats-new/2020/12/introducing-amazon-monitron/">https://aws.amazon.com/about-aws/whats-new/2020/12/introducing-amazon-monitron/</a>	April 8, 2021
AWS-2021a	Amazon	Machine Learning on AWS	<a href="https://aws.amazon.com/machine-learning/">https://aws.amazon.com/machine-learning/</a>	April 25, 2021
AWS-2021b	Amazon	AWS Deep Learning AMIs	<a href="https://aws.amazon.com/machine-learning/amis/">https://aws.amazon.com/machine-learning/amis/</a>	May 1, 2021
AWS-2021c	Amazon	AWS Neuron	<a href="https://aws.amazon.com/machine-learning/neuron/">https://aws.amazon.com/machine-learning/neuron/</a>	May 1, 2021
AWS-2021d	Amazon	AWS Announces General Availability of Amazon Lookout for Equipment	<a href="https://press.aboutamazon.com/2021/4/aws-announces-general-availability-of-amazon-lookout-for-equipment">https://press.aboutamazon.com/2021/4/aws-announces-general-availability-of-amazon-lookout-for-equipment</a>	May 1, 2021
AWS-2021	Amazon	Amazon Rekognition Developer Guide	<a href="https://docs.aws.amazon.com/rekognition/">https://docs.aws.amazon.com/rekognition/</a>	May 15, 2021

CODE	COMPANY	TITLE	URL	ACCESS DATE
			latest/dg/what-is.html	
AWSN-2021	Amazon	Amazon Neuron   Release Content	<a href="https://aws.amazon.com/machinelearning/neuron/">https://aws.amazon.com/machinelearning/neuron/</a> [data source no longer available]	May 3, 2021
MA-2016	Microsoft	Cognitive Services	<a href="http://web.archive.org/web/20160610225624/https://azure.microsoft.com/en-us/services/cognitive-services/">http://web.archive.org/web/20160610225624/https://azure.microsoft.com/en-us/services/cognitive-services/</a>	April 7, 2021
MA-2019	Microsoft	Transparency Note Azure Cognitive Services: Face API	<a href="https://azure.microsoft.com/mediahandler/files/resourcefiles/transparencynote-azure-cognitive-services-faceapi/Face%20API%20Transparency%20Note%20March%202019.pdf">https://azure.microsoft.com/mediahandler/files/resourcefiles/transparencynote-azure-cognitive-services-faceapi/Face%20API%20Transparency%20Note%20March%202019.pdf</a> URL no longer available, data source now available via: <a href="https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cplH">https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cplH</a>	April 7, 2021
MA-2020a	Microsoft	Bing Search APIs will transition from Azure Cognitive Services to Azure Marketplace on 31 October 2023	<a href="https://azure.microsoft.com/en-us/updates/bing-search-apis-will-transition-from-azure-cognitive-services-to-azure-marketplace-on-31-october-2023/">https://azure.microsoft.com/en-us/updates/bing-search-apis-will-transition-from-azure-cognitive-services-to-azure-marketplace-on-31-october-2023/</a>	April 7, 2021
MA-2020b	Microsoft	Azure Machine Learning updates Ignite 2020	<a href="https://azure.microsoft.com/en-us/updates/azure-machine-learning-updates-ignite-2020/">https://azure.microsoft.com/en-us/updates/azure-machine-learning-updates-ignite-2020/</a>	April 7, 2021
MA-2020c	Microsoft	What is Azure Machine Learning designer?	<a href="https://learn.microsoft.com/en-us/shows/ai-show/azure-machine-learning-designer">https://learn.microsoft.com/en-us/shows/ai-show/azure-machine-learning-designer</a>	April 8, 2021
MA-2021a	Microsoft	Azure Cognitive Services	<a href="https://azure.microsoft.com/en-us/services/cognitive-services/">https://azure.microsoft.com/en-us/services/cognitive-services/</a>	April 5, 2021
MA-2021b	Microsoft	What is Azure Content Moderator?	<a href="https://learn.microsoft.com/en-us/azure/ai-services/content-moderator/overview">https://learn.microsoft.com/en-us/azure/ai-services/content-moderator/overview</a>	April 7, 2021
MA-2021c	Microsoft	Learn how Microsoft AI is helping your industry	<a href="https://www.microsoft.com/enus/ai/business?activetab=pi">https://www.microsoft.com/enus/ai/business?activetab=pi</a>	May 5, 2021

CODE	COMPANY	TITLE	URL	ACCESS DATE
			vot1:primaryr2 [data source no longer available]	
MA-2021d	Microsoft	Azure Percept	<a href="https://azure.microsoft.com/enus/services/azure-percept/">https://azure.microsoft.com/enus/services/azure-percept/</a> [data source no longer available]	April 15, 2021
GCP-2019	Google	Products and services	<a href="http://web.archive.org/web/20190510045538/https://cloud.google.com/products/">http://web.archive.org/web/20190510045538/https://cloud.google.com/products/</a>	April 7, 2021
GCP-2021a	Google	Cloud AI building blocks	<a href="https://cloud.google.com/products/ai/building-blocks">https://cloud.google.com/products/ai/building-blocks</a>	April 7, 2021
GCP-2021b	Google	Getting started: training and prediction with Keras	<a href="https://cloud.google.com/ai-platform/docs/getting-started-keras">https://cloud.google.com/ai-platform/docs/getting-started-keras</a>	May 12, 2021
GCP-2021c	Google	Vertex AI	<a href="https://cloud.google.com/vertex-ai">https://cloud.google.com/vertex-ai</a>	May 18, 2021

## Appendix III

**TABLE 6:** Overview of the industry reports reviewed for this study

REFERENCE	YEAR	TITLE	AUTHOR(S)	URL
Benaich and Hogarth, 2018	2018	State of AI Report 2018	Benaich and Hogarth	<a href="https://drive.google.com/file/d/1rdPH1wf7d2Nx8Ax9sxd9eEypvMQu8cn7/view">https://drive.google.com/file/d/1rdPH1wf7d2Nx8Ax9sxd9eEypvMQu8cn7/view</a>
CBInsights, 2018a	2018	Amazon Strategy Teardown	CBInsights	<a href="https://www.cbinsights.com/research/report/amazon-strategy-teardown/">https://www.cbinsights.com/research/report/amazon-strategy-teardown/</a>
CBInsights, 2018b	2018	Google Strategy Teardown	CBInsights	<a href="https://www.cbinsights.com/research/report/google-strategy-teardown/">https://www.cbinsights.com/research/report/google-strategy-teardown/</a>
CBInsights, 2018C	2018	Microsoft Teardown	CBInsights	<a href="https://www.cbinsights.com/research/report/microsoft-">https://www.cbinsights.com/research/report/microsoft-</a>

REFERENCE	YEAR	TITLE	AUTHOR(S)	URL
				strategy-teardown/
CBInsights, 2020	2020	Big Tech In Healthcare: How Tech Giants Are Targeting The \$3T Industry	CBInsights	<a href="https://www.cbinsights.com/research/report/famga-big-tech-healthcare/">https://www.cbinsights.com/research/report/famga-big-tech-healthcare/</a>
McKinsey, 2018	2018	Artificial-intelligence hardware: New opportunities for semiconductor companies	McKinsey	<a href="https://www.mckinsey.com/~media/McKinsey/Industries/Semiconductors/Our%20Insights/Artificial%20intelligence%20hardware%20New%20opportunities%20for%20semiconductor%20companies/Artificial-intelligence-hardware.ashx">https://www.mckinsey.com/~media/McKinsey/Industries/Semiconductors/Our%20Insights/Artificial%20intelligence%20hardware%20New%20opportunities%20for%20semiconductor%20companies/Artificial-intelligence-hardware.ashx</a>
Benaich and Hogarth, 2019	2019	State of AI Report 2019	Benaich and Hogarth	<a href="https://drive.google.com/file/d/1RE0I4VMLN0xswXNnleAyoWsWFGexrZ1F/view">https://drive.google.com/file/d/1RE0I4VMLN0xswXNnleAyoWsWFGexrZ1F/view</a>
Benaich and Hogarth, 2020	2020	State of AI Report 2020	Benaich and Hogarth	<a href="https://docs.google.com/presentation/d/1ZUimafgXCBSLsgbacd6-a-dqO7yLyzl1ZJbiCBUUT4/edit">https://docs.google.com/presentation/d/1ZUimafgXCBSLsgbacd6-a-dqO7yLyzl1ZJbiCBUUT4/edit</a>
Baker et al., 2020	2020	Magic Quadrant for Cloud AI Developer Services	Van Baker, Elliot, Sicular, Mullen and Brethenoux	<a href="https://www.gartner.com/doc/reprints?id=11YCWP10B&amp;ct=200213&amp;st=sb">https://www.gartner.com/doc/reprints?id=11YCWP10B&amp;ct=200213&amp;st=sb</a> [data source is no longer available]
Baker et al., 2021	2021	Magic Quadrant for Cloud AI	Van Baker, Elliot, Sicular, Mullen	<a href="https://www.gartner.com/doc/">https://www.gartner.com/doc/</a>

REFERENCE	YEAR	TITLE	AUTHOR(S)	URL
		Developer Services	and Brethenoux	reprints?id=1255TRY6T&ct=210205&st=sb [data source is no longer available]

Published by



ALEXANDER VON HUMBOLDT  
INSTITUTE FOR INTERNET  
AND SOCIETY

in cooperation with



CREATE



centre  
— internet  
— et  
— **societe**



R&amp;I

IN3

Internet  
interdisciplinary  
Institute

Universitat Oberta de Catalunya



UNIVERSITY OF TARTU  
Johan Skytte Institute of  
Political Studies