Antitrust in artificial intelligence infrastructure – between regulation and innovation in the EU, the US, and China

Kena Zheng

Article - Version of Record

Suggested Citation:
Zheng, K. (2025). Antitrust in artificial intelligence infrastructure – between regulation and innovation in the EU, the US, and China. Computer Law & Security Review , 59, Article 106211.
https://doi.org/10.1016/j.clsr.2025.106211

Computer Law & Security Review 59 (2025) 106211

Contents lists available at ScienceDirect

Computer Law & Security Review: The International Journal of Technology Law and Practice

journal homepage: www.elsevier.com/locate/clsr

Antitrust in artificial intelligence infrastructure – between regulation and innovation in the EU, the US, and China

Kena Zheng [1,*]

ARTICLE INFO

Keywords:
Artificial intelligence infrastructure
Access data
Assess computational resources
Antitrust law
Innovation
Regulation

ABSTRACT

The enormous amount of data and the substantial computational resources are crucial inputs of artificial intelligence (AI) infrastructure, enabling the development and training of AI models. Incumbent firms in adjacent technology markets hold significant advantages in AI development, due to their established large user bases and substantial financial resources. These advantages facilitate the accumulation of enormous amounts of data, and the establishment of computational infrastructure necessary for sufficient data processing and high-performance computing. By controlling data and computational resources, incumbents raise entry barriers, leverage advantages to favour their own AI services, and drive significant vertical integration across the AI supply chain, thereby entrenching their market dominance and shielding themselves from competition. This article examines regulatory responses to these antitrust risks in the European Union (EU), the United States (US), and China, given their leadership in digital regulation and AI development. It demonstrates that the EU's Digital Markets Act, and China's Interim Measures for the Management of Generative Artificial Intelligence Services introduce broadly framed yet applicable rules to address challenges related to data and computational resources in AI markets. Conversely, the US lacks both AI regulations and digital-specific competition laws, instead adopting innovation-centric policies aimed at ensuring its AI dominance globally. Given the strategic importance of AI development, all three jurisdictions have adopted a cautious approach in investigating potential abusive practices.

1. Introduction

The launch of ChatGPT by OpenAI in November 2022 and subsequent developments of AI have demonstrated its potential as a significant technological advancement.[2] By 2023, the global AI market was valued at over €130 billion, with projections suggesting exponential growth to nearly €1.9 trillion by 2030.[3] Generative AI, a subset of AI, describes algorithms that can be used to create new content, including audio, code, images, text, music, and videos,[4] and involves popular applications like ChatGPT, DeepSeek, Grok, and Gemini, often developed under the branding of foundation models.[5]

AI development has been characterized by rapid innovation, with several players pushing technological frontiers.[6] As AI becomes prevalent, it has the potential to reshape the nature of economic activity and redefine competitive dynamics.[7] This could occur as increasing automation changes how parameters of competition are set or prompts the

* Corresponding author.
E-mail address: kena.zheng@hhu.de.

[1] Postdoctoral researcher in the project 'Shaping Competition in the Digital Age' (SCiDA) Project at the Heinrich Heine Universität Düsseldorf, Chair of Civil law, German and European Competition Law. Heinrich Heine Universität, Universitätstraβe 1, 40204, Düsseldorf, Germany. I would like to thank the anonymous Reviewers, and participants of the 2024 Digital Law Hub. All errors are mine.

[2] Q. Mei et al., 'A Turing Test of Whether AI Chatbots are Behaviourally Similar to Humans', (2024) 121 Proceedings of the National Academy of Sciences of the United States of America, 1-8.

[3] The European Parliament, AI investment: EU and Global Indicators, (2024).

[4] McKinsey, 'What is Generative AI?', (2 April 2024) <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai>, (accessed 4 September 2024). See also A. de Moncuit et al, 'AI Challenges in Competition Law – How are Regulators Responding?', (2024) Practical Law: PLC Magazine, 1-6. Nvidia, Glossary: What is Generative AI?, <https://www.nvidia.com/en-us/glossary/generative-ai/>, (accessed 4 September 2024). See also OECD, Artificial Intelligence, Data and Competition, (2024) 12.

[5] R. Bommasani et al., 'On the Opportunities and Risks of Foundation Models', Arxiv (2021) <https://arxiv.org/abs/2108.07258>.

[6] UK Competition & Markets Authority (CMA), AI Foundation Models: Update Paper, (2024) 12.

[7] OECD, *supra* note 4, 14.

https://doi.org/10.1016/j.clsr.2025.106211

Available online 7 October 2025

emergence of new supply chains.[8] Given the missed or underappreciated dynamics of digital markets that have led to 'winner-take-most-or-all', there is a concern that inadequate or delayed interventions may fail to address the potential tipping of emerging AI markets. The regulatory response that lags may fail to effectively address challenges, while overly swift or extensive regulations could damage a wellspring of innovation or even reinforce existing problems.[9]

AI services are developed by pre-training machine learning algorithms, producing foundation models that can be refined and deployed across a wide range of downstream applications.[10] In the supply chain of AI services, a number of steps are required to train, improve and deploy these models.[11] AI infrastructure, as the first layer of the AI supply chain, provides a foundation for AI development with computational resources and data.[12] These two inputs are not merely technical prerequisites but have emerged as key determinants of market power within the AI supply chain. Training AI models is a complex and resource-intensive process, requiring vast amounts of data and significant computational resources.[13] In this context, ensuring equitable access to both data and computational resources is essential for fostering innovation and ensuring contestability in AI markets.

However, the requirements for these key inputs may also raise significant antitrust risks, particularly due to their linkages with existing digital platform markets and the potential barriers big technology firms (big techs) may create for new entrants seeking access to data and computational resources for AI development,[14] while shielding themselves from market pressure.[15] Meanwhile, the dominance held by big techs also enables them to leverage their advantages to favor their own AI services. AI developers have to collaborate with these powerful incumbents through vertical integration, partnerships, or strategic agreements to access essential resources.[16] Consequently, anti-competitive effects may emerge as both input markets become increasingly concentrated, and powerful big techs vertically integrate across the AI supply chain. Over time, the development of (generative) AI may reinforce the entrenched market power of a small number of big techs that already dominate digital platform markets. More importantly, these incumbents could also profoundly shape the development of AI-related markets, potentially reducing innovation and harming consumer welfare.[17]

Additionally, certain economic characteristics that have justified regulatory intervention in digital platform markets do not fully translate to AI markets. For instance, while zero-price strategies, user profiling based on personal data, and strong network effects are central to platform economics, their influence appears more limited in the context of AI markets, particularly in the training phase of foundation models. These distinctions suggest that directly applying digital regulations designed for digital platform markets to AI markets may result in regulatory mismatches.

Antitrust laws and regulatory tools must therefore be enforced to prevent anti-competitive practices, which could limit AI development to a few superstar firms.[18] Therefore, several competition authorities have publicly expressed a range of concerns around AI, from the advantage that big techs have in their access to data used to train foundation models, to partnerships between companies potentially being used to sidestep required merger review processes.[19]

The EU, the US, and China hold global leadership in digital regulation and AI development. To illustrate, the EU continues to build and refine its regulatory framework - including the Digital Markets Act (DMA), the Artificial Intelligence Act (AI Act) - to ensure digital transformation in its Digital Single Market. China adopted the Interim Measures for the Management of Generative Artificial Intelligence Services (Interim Measures) in July 2023, imposing a wide range of obligations on the provision and use of generative AI technologies. However, the US has responded to emerging risks by adopting fragmented approaches to AI policies, prioritizing innovation to secure its global AI dominance. Notably, all three jurisdictions have launched investigations into certain big techs, such as Nvidia and OpenAI, for alleged abuse of dominance in AI infrastructure markets. Against this backdrop, the article evaluates the regulatory responses and recent initiatives adopted by the EU, the US, and China to address antitrust risks in AI markets, particularly those related to ensuring equitable access to data and computational resources.

Therefore, this article examines the competitive dynamics surrounding two critical inputs for AI development - data and computational resources - and assesses regulatory responses in the EU, the US and China to address the associated competitive concerns. Section 2 provides essential context on the role of data and computational resources as key inputs for AI technologies, and the various sources potentially accessing them. Building on this foundation, it analyses market powers within each market. Section 3 identifies key antitrust risks, including increasing entry barriers, leveraging market powers and vertical integrations via partnerships, arising from the potential abuse of dominance in data and computing markets by big techs. Subsequently, Section 4 assesses regulatory responses and the ongoing efforts adopted by the EU, the US, and China to address the challenges discussed in Section 3. Finally, Section 5 concludes the analysis.

2. AI infrastructure

This section provides an overview of how data and computational resources are allocated across the AI supply chain and analyses their impact on development and innovation within current AI markets.

2.1. Data as a key input

Data is an indispensable input for AI technologies, given the large

[8] Ibid.

[9] S. Hunt, et al. 'You Are What You Eat: Nurturing Data Markets to Sustain Healthy Generative AI Innovation', (2023) 1 CPI Tech Reg Chronicle.

[10] OECD, The Impact of Artificial Intelligence on Productivity, Distribution and Growth: Key Mechanisms, Initial Evidence and Policy Challenges, (2024) 45.

[11] For example, CMA explains three layers for the development of foundation models. See CMA, *supra* note 6, 10. Stucke and Ezrachi identified five layers in FM supply chain. M.E Stucke and A. Ezrachi, 'Antitrust & AI Supply Chains', (2024), SSRN Electronic Journal.

[12] Competition Bureau Canada, Artificial Intelligence and Competition: Discussion Paper, (2024), 9. See also CMA, *supra* note 6, 27. In this report, CMA clarified that the first layer of the AI supply chain is AI Infrastructure which includes compute, data and expertise. Additionally, more governments have clarified the importance of data and computational resources as essential inputs for AI development. For example, according to French Competition Authority, launching and developing generative AI require high computing power, a large amount of data, and a skilled workforce. See Press Release, Autorité de la Concurrence, 'Generative artificial intelligence: the Autorité starts inquiries ex officio and launches a public consultation open until Friday, 22 March', (2024)

[13] CMA, AI Foundation Models: Initial Report, (2023) 10. See also Stucke and Ezrachi, *supra* note 11. CMA, *supra* note 6, 6. Specialized engineering talent is also a key input for AI development. However, this article focuses on competition law and does not examine this aspect in detail.

[14] OECD, *supra* note 4, 3.

[15] Competition Bureau Canada, *supra* note 12, 14.

[16] CMA, *supra* note 6, 6.

[17] CMA, *supra* note 6, 12.

[18] OECD, *supra* note 10, 45.

[19] See European Commission, Speech by EVP Margrethe Vestager at the European Commission workshop on 'Competition in Virtual Worlds and Generative AI', (2024) < https://ec.europa.eu/commission/presscorner/detail/en/speech_24_3550>. See also CMA, *supra* notes 6 and 13. The US FTC, Generative AI Raises Competition Concerns, (2023). Autorité de la Concurrence of France, *supra* note 12.

requirements for training, testing, and evaluating models.[20] Generative AI, a subset of AI, has increasingly relied on larger datasets,[21] with recent large language models (LLMs) being trained on billions to trillions of tokens.[22] From a technical perspective, these AI models can be trained on various types of data, extracting and learning patterns inherent in the training data.[23] Consequently, the value of these models is directly correlated with the quantity and quality of the input data and user interactions, increasing in value with more extensive usage.[24] In other words, what is fed to models is important, and there is a need to ensure data quality, emphasizing the need to use trusted data sources.[25] However, the exact amount and sources of data required to train a foundation model remain unclear,[26] as many undertakings provide limited details beyond generally noting their use of a mix of public and proprietary data.[27]

In this context, this subsection examines the functions of data, as essential infrastructure, within the AI supply chain, comparing its impact on digital platforms (see Section 2.1.1) and identifying three primary sources of data used for the development of AI models (see Section 2.1.2).

2.1.1. The role of data in developing digital platforms and AI models

The characteristics of the digital platforms are summarised as: extreme returns to scale, network effects, and the role of data.[28] For digital platforms, multi-sidedness contributes to the significant profits of digital platforms and zero-price strategies for users. The benefit that one side derives from the platform depends on who participates on the other side: their numbers, but also on their identity.[29] For example, platforms that rely on advertising revenues will often provide content for a very low price, or even monetarily free, to consumers in order to attract them.

The use of data regarding the interests and behaviour of users forms an important means to attract customers on both sides of digital platforms.[30] The more detailed the profile that a provider of search engine, social network, or e-commerce platform has about its users, the more precise possibilities it can offer to advertisers for selecting their intended audience.[31] Specifically, advertisers benefit from better targeted advertising because of the higher probability that the advertised product is actually purchased by the users to whom the advertisement is displayed.[32] On the user side, the quality of the functionalities offered to users can be enhanced by using the collected data to increase the relevance of, for example, search results delivered by search engines and recommendations for future purchases made on e-commerce platforms.[33]

In this context, digital platform operators often directly collect personal data of end users for the purpose of providing online advertising services when end users use third-party websites and software applications.[34] The role of data in the competitive process between online platforms and undertakings is to draw profiles for users and thus assist advertisers and businesses in making decisions. Therefore, it is noted that personal data (information) of consumers is valuable for digital platforms that employ business models depending on the acquisition and monetization of personal data.[35]

However, this (zero-price) business strategy commonly adopted by platform operators may not be applicable in current AI markets. Notably, existing generative AI providers offer their models via chatbots, such as ChatGPT 4 and Grok-2, which charge users subscription fees.[36]

Therefore, the technological characteristics of AI markets feature both similarities and differences from digital platforms, regarding the roles of data. To illustrate, the role of the network effect is significantly smaller than for platforms, considering a given ChatGPT user does not derive significant direct benefits from others joining ChatGPT and subscription its AI services.[37] The exception is that data feedback loops are material for both platforms and AI technologies. With respect to digital platforms, data feedback loops are typically about users' profiles in improving matching efficiency. Regarding AI services, the feedback loops are about generating additional data to improve model performance,[38] although this type of user data is not the main source of quality gains.

Fundamentally, through techniques that allow the processing of large amounts of data, generative AI models learn to identify patterns and can 'predict' the best responses to queries based on probabilities.[39] Specifically, these models are essentially statistical models that predict

[20] Competition Bureau Canada, *supra* note 12, 11. See also D. Zha, Z. Pervaiz Bhat, K.He Lai, F. Yang, X. Hu, 'Data-Centric AI: Perspectives and Challenges', in Proceedings of the 2023 SIAM International Conference on Data Mining (SDM) (2023) 945-948. CMA, A Pro-Innovation Approach to AI Regulation: Government Response, (2024) 25.

[21] Competition Bureau Canada, *supra* note 12, 11.

[22] T.B. Brown et al. 'Language Models are Few-Shot Learners', in Advances in Neural Information Processing System, (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), (2020) 33 Curran Associates, 1877–1901.

[23] OECD, *supra* note 4, 18.

[24] T. Schrepel and J. Potts, 'Measuring the Openness of AI Foundation Models: Competition and Policy Implications', (2024), Sciences Po Digital, Governance and Sovereignty Chair, Working Paper.

[25] Singapore Infomcomm Media Development Authority, Model AI Governance Framework for Generative AI, (2024) 4.

[26] CMA, *supra* note 6. See also OECD, *supra* note 4, 19.

[27] For example, OpenAI explains that they used a mixture of publicly available data (such as from the internet) and licensed data from third-parties to train GPR-4. See OpenAI, J. et al. GPT-4 Technical Report, (2023), <https://arxiv.org/abs/2303.08774v6>.

[28] J. Crémer, Y.A de Montjoye and H. Schweitzer, Competition Policy for the Digital Era, (2019) Report to the European Commission, 19.

[29] Ibid.

[30] I. Graef, 'Market Definition and Market Power in Data: The Case of Online Platforms, (2015) World Competition, 473-505.

[31] Ibid. See also Crémer, de Montjoye and Schweitzer, *supra* note 28, 16. UNCTAD, UNCTAD Digital Economy Report 2019: Value Creation and Capture: Implications for Developing Countries (Geneva, 4 September 2019). OECD, Data Portability, Interoperability and Digital Platform Competition, OECD Competition Committee Discussion Paper (June 2021). H. Schweitzer et al, 'Data Access and Sharing in Germany and in the EU: Towards a Coherent Legal Framework for the Emerging Data Economy - A Legal, Economic and Competition Policy Angle', (2022) SSRN Electronic Journal, 71.

[32] Graef, *supra* note 30. See also Schweitzer et al, Ibid.

[33] Graef, *supra* note 30.

[34] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act), OJ 2022 L 265/1, para 36.

[35] H.A. Shelanski, 'Information, Innovation, and Competition Policy for the Internet', (2013) 161 University Pennsylvania Law Review, 1663-1678.

[36] See for example, OpenAI, ChatGPT, https://openai.com/chatgpt/pricing/. The launch of DeepSeek on 2 November 2023 seems challenge the current AI business models requiring subscription fees. Some of DeepSeek's models are available for free, others are offered through paid API services. See R. Richards, 'DeepSeek R1 Basically Replaces GPT O1 - for free', (2025), < https://www.notta.ai/en/blog/deepseek-r1-vs-openai-gpt-o1?utm_source=chatgpt.com>.

[37] A. Korinek and J. Vipra, 'Concentrating intelligence: Scaling and Market Structure in Artificial Intelligence', (2024) Working Paper No. 228, 19.

[38] Stucke and Ezrachi, *supra* note 11.

[39] OECD, *supra* note 4, 12.

sequences of tokens.[40] They need many occurrences of sequences of tokens in order to make robust predictions. That, in turn, requires large volumes of data. The predictive nature of these AI technologies plays a crucial role in anticipating future applications,[41] distinguishing them from digital platforms, which primarily aim to facilitate interactions within multi-sided markets. Consequently, some argue that generative AI models essentially function as knowledge pools and belong to a class of commons.[42] They utilize resources derived from the commons (training data) and generate value through interaction, such as training or prompting.

In this context, data is mainly used to train AI models and facilitate machine learning processes. To illustrate, the diversity, volume, and quality of data greatly affect models' ability to understand and generate contextually relevant, high-quality output.[43] Even with significant computational resources and top-tier talent, the models cannot generate meaningful output without sufficient rich, varied, and relevant data.[44] Higher-quality data can significantly enhance model performance per token of training data,[45] highlighting the significance of data quality in the development of effective AI services. Furthermore, higher-quality data is needed to reduce the risk that model outputs are themselves of low quality, for example, by exhibiting bias or producing incorrect information that is presented as true, known as 'hallucinations.'[46]

Unlike digital platforms, where data is often detailed to profile users, data used for pre-training AI models does not require such personalization. Instead, the quality of data is more crucial than its capacity for precise personalization. To develop effective models, sufficiently reliable data must be supplied to populate their parameters (or layers). Therefore, AI developers need to clean data by removing any undesirable elements and turning it into the appropriate format, such as tokenizing it or adding labels and structure, before initiating its process of training foundation models.[47]

Subsequently, in the pre-training stage of foundation models, it is common practice to use publicly available data, usually scraped from the web, as the majority of a pre-training dataset.[48] Likewise, proprietary data that could be bought from other sources, such as media and publishing companies or other owners of digital archives, might also have value for pre-training.[49] Additionally, data plays a different role in the process of fine-tuning and deploying models, compared to the training process. Specialized datasets are essential for the creation of fine-tuned models tailored to specific tasks or industries. For example, AI models for medical diagnosis require medical records for training.[50] The impact of data can be observed at the post-deployment of AI, when additional data may be used by AI models to execute a query.[51] In this context, data feedback loops are evident and useful. However, the roles of data in fine-tuning and deploying AI services are still not entirely similar to the function of profiling employed by digital platforms.

2.1.2. Three sources to access data

Generally, discussing access to data in the abstract is futile.[52] The significance of data and data access for competition will always depend on an analysis of the specificities of a given market and the type of data and data usage in a given case.[53] Given the different roles of data in the development of digital platforms and AI technologies, this subsection explores three potential sources of data, including (1) public data, (2) synthetic data, and (3) proprietary data, used in the phases of pre-training and fine-tuning AI models. This discussion serves as a foundation for the analysis of data access and potential antitrust risks in Section 3.

With respect to the pre-training phase, data scale and quality are the keys to the performance of foundation models.[54] Models, built by pre-training a machine learning algorithm on a broad dataset, seek to produce general-purpose, grammatically correct, and contextually coherent text output. Pre-training, thus, relies on a huge volume of high-quality data, described as peer reviewed and professionally written content, and major sources include books, news articles, scientific papers, and Wikipedia.[55] Specifically, most FM training relies heavily on data scraped from the web, a prime example being the datasets constructed by Common Crawl.[56] For example, Meta engineers have filtered Common Crawl using Wikipedia as the benchmark finding greatly improved performance.[57] Likewise, a number of high-profile models developed to date, including LLaMA (Meta), GPT-3 (OpenAI), and Stable Diffusion (StabilityAI) have been pre-trained entirely on data from publicly available sources.[58]

However, one may note that there is no unified definition or clear scope of publicly available data.[59] In other words, it is up to AI developers to define and disclose what they consider publicly available data for training their models. For example, on the website of X, it is explained that public X data refers to public posts, metadata associated with public posts (such as engagement and reposts), public Spaces, and public profiles (such as bio and display name).[60] Additionally, users' real-time interactions, inputs, and results may also be used to train and improve the performance of those generative AI models developed by X.[61]

Other big techs have announced that they pre-trained their AI models using publicly available data, but without providing explicit lists of datasets. For example, Meta initially claimed to use information that is

[40] B. Martens, 'Why Artificial Intelligence is Creating Fundamental Challenges for Competition Policy', (2024) Bruegel Policy Brief.
[41] OECD, *supra* note 4, 12.
[42] Schrepel and Potts, *supra* note 24.
[43] Hunt et al., *supra* note 9.
[44] Ibid.
[45] L. Gao et al, 'The Pile: An 800GB Dataset of Diverse Text for Language Modelling', (2020) arXiv < https://arxiv.org/pdf/2101.00027>. See also CMA, *supra* note 6, 28.
[46] OECD, *supra* note 4, 19.
[47] Ibid.
[48] CMA, *supra* note 6, 28.
[49] CMA, *supra* note 6, 30.
[50] Hunt et al., *supra* note 9.
[51] OECD, *supra* note 4, 9.
[52] Crémer, de Montjoye and Schweitzer, *supra* note 28, 73.
[53] Ibid, 74.
[54] Hunt et al., *supra* note 9.
[55] Ibid. See also Martens, *supra* note 40.
[56] Hunt et al., *supra* note 9.
[57] Hunt et al., *supra* note 9. See also G. Wenzek et al., 'CCNet: Extracting High Quality Monolingual Datasets from Web Crawl Data', (2019) ARXIV, < https://arxiv.org/pdf/1911.00359>.
[58] CMA, *supra* note 6, 28.
[59] Such differences perhaps stem from divergences in privacy and data protection frameworks across legal regimes. For example, in the EU, the General Data Protection Regulation (GDPR) classifies data into personal and non-personal categories. Specifically, Article 4(1) of the GDPR defines personal data as any information relating to an identified or identifiable natural person, including, for example, a person's name, location, physical attributes, mental state, economic circumstances, preferences, or website visits. By contrast, Article 6(2) of the DMA specifies that data that not publicly available includes both aggregated and non-aggregated data generated by business users, which may be inferred from or collected through the commercial activities of business users or their customers. This encompasses click, search, view and voice data collected on the relevant core platform services or associated services offered by the gatekeeper. Importantly, in some jurisdictions - such as China - there are currently no specific legal provisions that clearly delineate the types of data that may be used for AI training purposes. In light of these regulatory divergences, it appears that big techs are attempting to establish and shape a new concept –'publicly available data' – in a manner that preserves their broad discretion to collect and use such data, particularly for the training of AI systems.
[60] X, 'About Grok, Your Humorous AI Assistant on X', < https://help.x.com/en/using-x/about-grok>.
[61] Ibid.

publicly available online and licensed information.[62] Specifically, LLaMA 2 was trained on a new mix of publicly available data, excluding data from Meta's products and services.[63] However, two months later, Meta announced that it had used a combination of data sources for training AI models, including information that's publicly available online, licensed data, and information from Meta's products and services.[64] Meta further clarified that generative AI models (released after LLaMA 2) are pre-trained with publicly shared posts from Facebook and Instagram, including photos and text.[65] This inconsistency may also raise questions about Meta's non-transparency on the scope of 'publicly available data'.

In light of this, types of so-called 'publicly available data' may effectively become controlled by big techs, while access to new entrants may be restricted.

Additionally, certain researches suggest that publicly available data used for training foundation models could be fully exhausted,[66] with estimates indicating that high-quality data could be running out by 2024.[67] This is due to the growth rate of quality data, which is insufficient to sustain AI development.[68] Meanwhile, public data often contains repetition and is of lower quality, making data filtering a usual step in model training.[69] Other research predicts that if current LLM development trends continue, models will be trained on datasets roughly equal in size to the available stock of public human text data between 2026 and 2032, or slightly earlier if models are overtrained.[70]

Innovations in AI technologies and methods could decrease reliance on human-produced data for AI development.[71] For example, new methods for training or improvements in models or architectures could create efficiencies, requiring fewer data resources to achieve the same level of performance. More importantly, it is possible that small amounts of data can be extrapolated to create synthetic data for AI development.[72] Compared to the collection of public data, the use of synthetic data may be more easily accessible and less costly.[73]

However, researchers also explain that synthetic data seems unlikely to generate enough valuable data for training foundation models.[74] It shows that the use of synthetic data incurs risks in terms of data contamination and bias propagation, which could result in 'model collapse'.[75] Moreover, some argued that synthetic data generates substantial risk and leads to increasingly lower quality as the training loop progresses, unless there is enough real data.[76] It, therefore, appears likely that a reasonable amount of human produced data is still required to train foundation models for the foreseeable future.[77]

Consequently, as public data becomes increasingly exhausted and synthetic data presents certain risks, AI developers may turn more toward proprietary datasets. If developing the most competitive models requires obtaining large volumes of data from proprietary sources, then access to such data could become a crucial factor influencing competition in AI markets.[78] Several sources of proprietary data could exist, perhaps being held within internal databases of firms, or collected from users of services they host.[79]

In this regard, big techs may possess significant advantages in accessing proprietary data. On the one hand, they already have access to substantial quantities of proprietary data suitable for AI development, often collected through user interactions on their digital platforms or generated as a byproduct of their dominant positions in other digital markets.[80] For example, certain large datasets operated by big techs are proprietary and may provide unique insights that others struggle to replicate, such as Google's ownership of YouTube and the potential to control access to its video transcripts.[81] On the other hand, the high costs associated with accessing proprietary data further reinforce these advantages, as big techs have the financial capacity to pay gigantic licensing fees for such access.[82]

Pre-trained foundation models are not inherently tailored for specific tasks, and their outputs often reflect discrimination or biases present in the training data.[83] To enhance their utility and reliability, fine-tuning is employed – an additional training phase that adapts a model to perform a tailored task or function using targeted datasets and human feedback, thereby improving accuracy and reducing misinformation.[84]

During fine-tuning, data quality plays a crucial role, specifically in terms of its relevance to a certain application or domain, and its usability for specialized tasks.[85] Unlike the broad, diverse data used in pre-training, fine-tuning relies on narrower, task-specific datasets suited

[62] Meta, 'How Meta Uses Information for Generative AI Models and Features', <https://www.facebook.com/privacy/genai/>. Meta explains that the public content that Europeans share on their services and others, such as public posts or comments, then models and the AI features that power will not accurately understand important regional languages, cultures or trending topics on social medias.

[63] Meta, 'Llama 2: Open Foundation and Fine-Tuned ChatModels', (July 2023) 5.

[64] Meta, 'Privacy Matters: Meta's Generative AI Features', (2023) < https://about.fb.com/news/2023/09/privacy-matters-metas-generative-ai-features/>.

[65] M. Loh, 'Meta said it's been using your public Instagram photos and Facebook posts to train its AI', (2023) Business Insider <https://www.businessinsider.com/meta-train-ai-public-post-instagram-facebook-2023-10?international=true&r=US&IR=T>. See also Meta, 'Privacy Matters: Meta's Generative AI Features', (2023) < https://about.fb.com/news/2023/09/privacy-matters-metas-generative-ai-features/>.

[66] P. Villalobos et al, 'Will We Run Out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning', ARXIV, (2022), <https://www.debicker.eu/content/files/pdf/2211.04325.pdf>. See also Competition Bureau Canada, *supra* note 12, 11.

[67] Stanford University, Artificial Intelligence Index Report 2024 Human-Centered Artificial Intelligence, (2024).

[68] Hunt et al., *supra* note 9.

[69] T. Brown et al., 'Language Models are Few-Shot Learners', (2020) ARXIV < https://arxiv.org/pdf/2005.14165> ; a. Chowdhery et al., 'PaLM: Scaling Language Modeling with Pathways', (2023) ARXIV < https://arxiv.org/pdf/2204.02311>. H. Touvron et al., 'LLaMA: Open and Efficient Foundation Language', (2023) ARXIV, < https://arxiv.org/pdf/2302.13971>.

[70] Villalobos, *supra* note 66.

[71] Competition Bureau Canada, *supra* note 12, 15.

[72] See Amazon Web Services, 'What is Synthetic Data?', <https://aws.amazon.com/what-is/synthetic-data/>.

[73] Ibid.

[74] Hunt et al., *supra* note 9. See also OECD, *supra* note 4, 19.

[75] Model collapse occurs when a model fails to learn as intended. In the context of synthetic data, certain research has shown that use of this data to train AI models may result in a 'feedback loop' that leads to more repetitive, less diverse, and lower quality outcomes. See G. Martinez et al, 'Towards Understanding the Interplay of Generative Artificial intelligence and the Internet', (2023) ARXIV <https://arxiv.org/abs/2306.06130>. T. Marwala et al, 'The Use of Synthetic Data to Train AI Models: Opportunites and Risks for Sustainable Development', (2023) United Nations University. Competition Bureau Canada, *supra* note 12, 15.

[76] S. Alemohammad, et al. (2023), 'Self-Consuming Generative Models Go MAD', (2023) ARXIV < https://arxiv.org/abs/2307.01850>. See also OECD, *supra* note 4, 20.

[77] Stanford University, *supra* note 67.

[78] CMA, *supra* note 6, 43.

[79] OECD, *supra* note 4, 19.

[80] CMA, *supra* note 6, 30.

[81] T. Schrepel and A. Pentland, 'Competition between AI Foundation Models: Dynamics and Policy Recommendations', (2023) MIT Connection Science Working Paper. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4493900>.

[82] Ibid.

[83] As noted by a joint statement from several federal agencies. FTC, FTC Chair Khan and Officials from DOJ, CFPB and EEOC Release Joint Statement on AI (April 25, 2023), <https://www.ftc.gov/news-events/news/press-releases/2023/04/ftc-chair-khan-officials-doj-cfpbeeoc-release-joint-statement-ai>.

[84] Hunt et al., *supra* note 9.

[85] Ibid.

for applications such as use-cases like dialogue with users, legal advisories, customer service, or medical consultations.[86] For example, OpenAI's ChatGPT was tweaked from the GPT-3.5 to perform specialized chatbot functionalities using a narrower set of chat-specific training data.[87] Similarly, Morgan Stanley fine-tuned GPT-4 with proprietary financial data to create an internal tool – AI Financial Advisor.[88] In this regard, fine-tuning typically depends on labelled data,[89] and does not benefit from the volume of data as much as pre-trained models do.

Additionally, one of the most important methods used in the fine-tuning phase is reinforcement learning with human feedback (RLHF), aiming to enhance safety and reduce biases in AI models.[90] RLHF can either use data generated by humans directly or employ a reward algorithm trained on human feedback. For dialogue-based models such as ChatGPT, user interactions and feedback are important sources for fine-tuning. This leads to a network effect: the more the model is used, the better it becomes.[91] The fine-tuning phrase is, therefore, critical for reducing misinformation and driving significant improvements in accuracy, safety, and other performance metrics of models.[92]

Overall, both the pre-training and fine-tuning processes highlight the significant role of data in AI development. The analysis in this subsection reveals the distinct functions of data in AI model training and digital platforms. It further contributes to identifying three primary sources of data access by AI developers and underscores the significant advantages that big techs hold in accessing both public and proprietary data.

2.2. Computational resources

Computational resources appear likely to remain an important prerequisite for AI services.[93] AI developers generally have two options for accessing computational resources: purchasing hardware (often referred to as AI chips) to operate their own systems or accessing them via cloud computing services.[94] However, both cloud computing and AI chip markets are highly concentrated, with big techs dominating each.[95]

Compared to data, which is easily shareable, non-rivalrous intangible goods, making them inherently difficult to control, AI computing hardware is tangible and produced using an extremely concentrated supply chain.[96] Access to substantial computing hardware remains critical for AI training,[97] particularly for developing large-scale AI models, which require substantial computing capacity at a considerable cost.[98] For example, OpenAI's GPT-4 reportedly used $78 million in computing costs for training, while Google's Gemini required approximately $191 million.[99] In this context, global governments are investing in domestic computing capacity, controlling the flow of computational resources to competing countries (for example, from the US to China), and subsidizing access to computational resources for certain sectors.[100]

AI chips are specialized hardware units designed to train and operate AI technologies.[101] Currently, Graphics Processing Units (GPUs), such as Nvidia A100/H100, are the most widely used chips for AI development. With their ability to process large amounts of data at high speeds and perform in parallel at scale, GPUs are in high demand for training foundation models.[102]

Training large-scale AI models requires thousands of GPUs. For example, GPT-3 is estimated to have required over a thousand high-end GPUs, and Meta's LLaMA used over two thousand.[103] Notably, DeepSeek may be no exception, as the developers of the DeepSeek V3 model highlighted their use of a cluster of 2048 H800 GPUs to reflect a trade-off between time and cost.[104] However, the AI chip market is highly concentrated,[105] with one supplier, Nvidia, reported supplying up to 80 per cent of the global GPUs market.[106] This dominance, coupled with an ongoing shortage of GPUs intended for AI application, poses significant barriers to entry for new developers in the AI market.[107]

Computational resources can also be obtained via cloud computing services, which allow users to remotely access processing power and pay only for the resources they consume. However, cloud computing services are dominated by a small number of big techs.[108] For example, in the third quarter of 2024, Amazon Web Services (AWS) held 31 per cent of the market share, Microsoft Azure had 20 per cent, and Google Cloud accounted for 10 per cent of the global cloud infrastructure market.[109] Together, these 'Big Three' account for >60 per cent of the global cloud market, with the remaining competitors holding only a small share in the low single digits.[110] As a result, new AI entrants have to rely on cloud

[86] Ibid.
[87] Ibid.
[88] Morgan Stanley, 'Morgan Stanley Wealth Management Announces Key Milestone in Innovation Journey with OpenAI', (2023), < https://www.morganstanley.com/press-releases/key-milestone-in-innovation-journey-with-openai>. See also OpenAI, 'Morgan Stanley Uses AI Evals to Shape the Future of Financial Services', < https://openai.com/index/morgan-stanley/>.
[89] A type of structured dataset in which data elements are characterized with a data tag, for example pictures of animals (data) may be labeled with 'dog', 'cat', etc. See Taori et al., 'Alpaca: A Strong, Replicable Instruction-Following Model', Stanford University (2023).
[90] Hunt et al., *supra* note 9.
[91] Ibid.
[92] Ibid. For example, InstructGPT generated 37 per cent higher user satisfaction relative to GPT-3 and evaluators noted that the fine-tuned model generates truthful and informative answers about twice as often. See also L. Ouyang et al., 'Training Language Models to Follow Instructions with Human Feedback', (2022) ARXIV <https://arxiv.org/ pdf/2203.02155.pdf>.
[93] OECD, *supra* note 4, 21.
[94] OECD, *supra* note 4, 20. See also CMA, *supra* note 6. J. Vipra and S. Myers West, 'Computational Power and AI', (2023) AI Now Institute, 29.
[95] Vipra and Myers West, *supra* note 94.
[96] S. Girish et al, 'Computing Power and the Governance of Artificial Intelligence', (2024) ARXIV < https://arxiv.org/pdf/2402.08797>.
[97] Training in AI refers to the process of teaching an AI technology or system to learn from data and perform its intended task(s). Inference in AI refers to the process of an AI system providing an output. See Competition Bureau Canada, *supra* note 12, 9. CMA, *supra* note 6, 6. OpenAI, AI and Compute, (2018), <https://openai.com/index/ai-and-compute/>. N. Maslej et al. 'The AI Index 2023 Annual Report', (2023) AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.
[98] OECD, *supra* note 4, 20. See also Schrepel and Potts, *supra* note 24.
[99] Stanford University, *supra* note 67. See also Hunt et al., *supra* note 9.
[100] Girish et al, *supra* note 96.
[101] Competition Bureau Canada, *supra* note 12, 10.
[102] OECD, A Blueprint for Building National Compute Capacity for Artificial Intelligence, (2023).
[103] CMA, *supra* note 6, 28.
[104] DeepSeek, 'DeepSeek-V3 Technical Report', (2024) Arxiv < https://arxiv.org/html/2412.19437v1>.
[105] Vipra and Myers West, *supra* note 94. See also CMA, *supra* note 6.
[106] OECD, supra note 4, 20. See also The Economist, 'Just how rich are businesses getting in the AI gold rush?,' 17 March 2024, https://www.economist.com/business/2024/03/17/just-how-rich-are-businesses-getting-in-the-ai-gold-rush. Vipra and Myers West, *supra* note 94.
[107] Competition Bureau Canada, *supra* note 12, 15.
[108] CMA, *supra* note 6. See also Vipra and Myers West, *supra* note 94.
[109] Statista, 'Amazon Maintains Cloud Lead as Microsoft Edges Closer', (2024), < https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/>.
[110] Ibid.

services provided by these big techs for essential computing resources, even though these incumbents are also direct competitors in AI development.[111] Consequently, AI developers are increasingly forming partnerships with big techs to secure access to cloud computing services.[112]

Both the cloud computing service and AI chip markets, thus, exhibit high concentration, reinforcing big techs' dominating each market.[113]

3. Antitrust risks

At the outset of this section, it is important to underline that the law, particularly competition law, should recognize the significant benefits of generative AI, including its contributions to both static and dynamic efficiency (innovation) and its direct pro-competitive effects.[114] However, the structures and trends in AI infrastructure also warrant caution. Strong economic forces, driven by control over vast amounts of data and computational resources, may therefore push AI input markets toward monopoly or highly imperfect competition. In light of this, it is essential to examine the potential risks arising from the concentration of these critical inputs within the AI supply chain, given their irreplaceable roles in AI development.

3.1. Increased entry barriers

The scale of computational resources and data required to achieve competitive AI performance provides incumbent firms a significant advantage.[115] To illustrate, both data and computational resources markets exhibit high levels of concentration, with a few big techs, such as GAMMA, holding significant shares.[116] By controlling these resources, they can materially restrict access to essential inputs, insulating themselves from competition.[117] Such restrictions may reinforce their dominant positions in related markets, such as search and productivity software, by hindering rivals from developing or deploying effective AI models capable of powering next-generation competitive alternatives. This dynamic risks reducing consumer choice, lowering quality, and driving up prices for downstream business customers and end users.[118] Consequently, limited access to these essential inputs presents a significant barrier to entry for startups seeking to enter AI markets.[119] The dominance of incumbents could profoundly shape the development of AI-related markets, undermining fair, open, and effective competition, and ultimately harming innovation and consumer welfare.[120]

Notably, evidence suggests that access to data, particularly public data, is already a restrictive factor for AI development.[121] For example, 70 per cent of new entrants have reported experiencing an insufficient amount of training data.[122] On the one hand, several companies such as Reddit, Stack Overflow, X, and others have started licensing access to their databases for the purpose of training foundation models. For example, Reddit began charging for its Application Programming Interface (API) to prevent tech companies from scraping its data for free.[123] However, small AI developers are at a disadvantage, unable to secure the same access as incumbents due to the high costs associated with licensing data.

On the other hand, data-rich companies are restricting data access to web crawlers, thereby limiting the ability of other AI developers to train their models on public data.[124] These restrictions, imposed by big techs on web crawling, could potentially amount to an implicit refusal to supply, making entry into AI markets more challenging for new developers. This trend intensively raises antitrust concerns, as the actions of big techs restrict competition. For example, since the rise of the generative AI era in 2023, an increasing number of online content providers, including big techs, have blocked automated bots from collecting data used for training foundation models.[125]

Building on both practices of licensing available data and restricting web crawling, new entrants face particular challenges in accessing public data.

With respect to computational services, only a handful of firms can rely on their own resources, as discussed in Section 2.2. New entrants face significant challenges in accessing computational resources, whether through AI chips or cloud computing services. The high cost of hardware is prohibitive for startups, a challenge further exacerbated by the limited availability of AI accelerator chips. On the other hand, cloud computing services are dominated by big techs, which may have incentives to restrict access to these resources for potential competitors. Accordingly, there is growing concern that incumbent firms may control computational resources to shape AI-related markets in their own interests.

Therefore, if big techs possess market power in AI infrastructure, particularly concerning data or computational resources, they may both have the ability and the incentive to foreclose access to these critical inputs for (downstream) competitors.[126] As a result, contestability in AI infrastructure is reduced, particularly due to the existence of very high barriers to entry. In response to this challenge, certain AI developers have formed partnerships with big techs to secure access to computing services and data.[127] However, such collaborations raise additional concerns regarding vertical integration in the AI supply chain, which will be further examined in Section 3.3 below.

3.2. Leveraging advantages

Leveraging is a generic term that refers to the impact a practice

[111] Digital Platform Regulators Forum, Examination of Technology: Large Language Models, (2023).
[112] CMA, *supra* note 6, 6.
[113] Vipra and Myers West, *supra* note 94.
[114] N. Maslej et al, 'Artificial Intelligence Index Report 2024', (2024) AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, 19-20. See also M. Draghi, The future of European Competitiveness, Part A | A competitiveness strategy for Europe, (2024) 21. OCED, The Potential Impact of Artificial Intelligence on Equity and Inclusion in Education, (2024) 6.
[115] OECD, *supra* note 10, 32. See also CMA, *supra* note 6.
[116] Vipra and Myers West, *supra* note 94. See also CMA, *supra* note 6, 12.
[117] CMA, *supra* note 6, 12 and 14.
[118] CMA, *supra* note 6, 14.
[119] Competition Bureau Canada, *supra* note 12, 14.
[120] CMA, *supra* note 6, 9 and 12.
[121] Hunt et al., *supra* note 9.
[122] Mckinsey, 'The state of AI in early 2024: Gen AI Adoption Spikes and Starts to Generate Value', (2024), <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/>.
[123] R. Goswami, 'Reddit Will Charge Hefty Fees to the Many Third-party apps that Access its Data', CNBC (2023) <https://www. cnbc.com/2023/06/01/reddit-eyeing-ipo-charge-millions-in-fees-for-third-party-api-access.html>. See also K. Wiggers, 'Reddit will begin Charging for Access to its API', (2023) < https://techcrunch.com/2023/04/18/reddit-will-begin-charging-for-access-to-its-api/>.
[124] Korinek and Vipra, *supra* note 37, 19. See also R. Fletcher, 'How Many News Websites Block AI Crawlers?', Blog Post, Reuters Institute for the Study of Journalism, Oxford University (2024).
[125] K. Knibbs, 'Most Top News Sites Block AI Bots. Right-Wing Media Welcome Them', (2024), WIRED < https://www.wired.com/story/most-news-sites-block-ai-bots-right-wing-media-welcomes-them/>. It was confirmed that Facebook, Instagram, Craigslist, Tumblr, The New York Times, The Financial Times, The Atlantic, Vox Media, and the USA Today network are among the many organizations opting to exclude their data from Apple's AI training.
[126] OECD, *supra* note 4, 31.
[127] OpenAI, OpenAI Data Partnerships, (2023). See also Competition Bureau Canada, *supra* note 12, 12.

identified in one market may have on another.[128] This raises potential anti-competitive risks, particularly since big techs operate across upstream and downstream markets of the AI supply chain. To illustrate, big techs, which holds dominant positions in certain digital markets, can easily access large datasets and computing infrastructure at a significant scale.[129] These firms typically feature the ability to connect numerous business users with end users through their platform services, enabling them to leverage their competitive advantages, such as access to large amounts of data, from digital platforms to the training of AI models.[130]

A notable example of leveraging is the case of *Google Search (Shopping) v Commission*.[131] The General Court of the EU concluded that, through leveraging, Google had abused its dominant position in the market for general search services to favour its own comparison shopping service.[132] By leveraging its dominance, Google promoted the positioning and display of its own comparison shopping service and its results on general results pages, while demoting the results of competing comparison shopping services through algorithmic adjustment. Due to their inherent characteristics, competing comparison shopping services were particularly prone to being demoted on those pages by adjustment algorithms.

Georg Picht applied the concept of leveraging to AI markets, using Microsoft as an example. He suggested that if a market for big data-based AI generated digital content (the market for 'AI Content Generation Tools') exists or emerges, Microsoft could leverage its market power by integrating Chatpilot with its operating system and office productivity software, thereby strengthening its position in this new market.[133]

Other big techs with strong positions in digital platform markets can also leverage their dominance in data and computing resources to advance their AI development. For example, Google maintains dominant positions in the markets for general online search services and the Android operating system. Building on this, this article assumes that the integration of Google's AI model, Gemini, with its Android operating system and Google Search services could promote Gemini's performance and further extend Google's market power in AI markets. Meanwhile, as noted by the General Court in its judgment, Google relied on 'crawled' data and generic relevance signals derived from these data to generate its general search results.[134] This advantage is further reinforced by the extensive data collected across Google's ecosystem (including YouTube), which can be used to train its AI models and enhance its AI-driven services and results. This assumption is also supported by Google's vague statement regarding the scope of 'publicly available data' used for Gemini's pre-training, as discussed in Section 2.1.2. A further example of leveraging advantages in AI infrastructure can be observed with X, which has used user data from its social platform to develop its AI model, Grok.[135] By utilizing the vast amounts of data from real-time posts on the X platform, Grok is capable of providing updates and responses to user queries. Moreover, as more data is collected and used in training, the model generally becomes better at understanding and generating natural language.[136] These practices further demonstrate how big techs leverage their advantages in AI infrastructure to promote their AI models and services, potentially weakening competition in AI markets.

3.3. Vertical integration via partnerships

Generally, a concentration – whether in the form of a merger, an acquisition of control or a full-function joint venture – may produce both anti-competitive and pro-competitive effects.[137] In the context of innovative markets, such a concentration is often considered pro-competitive when the combination of different ideas can generate unforeseen synergies that spur innovation.[138] However, a closer examination of the motives behind concentrations suggests the need for a more cautious approach, as certain transactions may have detrimental effects on social welfare.[139] While the antitrust risks associated with acquisitions are not exclusive to AI markets, this sector presents specific concerns that warrant particular attention.[140] Risks may emerge when markets become concentrated, and big techs engage in vertical integration along the AI supply chain,[141] often through partnerships rather than traditional acquisitions. A growing trend has emerged in which big techs establish partnerships with small AI developers,[142] effectively integrating vertically without triggering conventional merger scrutiny. For example, GAMMA are active across various levels of the AI supply chain to varying degrees, often through partnerships and agreements with FM developers.[143] Public reports from late 2023 revealed that major cloud service providers, Microsoft, Google, and Amazon, had invested billions in AI developers Anthropic and OpenAI.[144] Specifically, Amazon has invested $8 billion,[145] and Google has invested $2 billion in Anthropic.[146] Microsoft has invested close to $14 billion in OpenAI.[147]

Building upon vertical integration through partnerships, big techs can expand their influence within the AI supply chain and further solidify their dominant positions in digital markets. Researchers have observed that these partnerships often involve joint ventures, equity

[128] Case T-5/02 *Tetra Laval v Commission* [2002] ECLI:EU:T:2002:264, paras 156, 158 and 217.
[129] CMA, *supra* note 6, 9 and 12.
[130] Digital Markets Act, *supra* note 34, para 3
[131] Case T-612/17 *Google and Alphabet v Commission* (Google Shopping) [2021] ECLI:EU:T:2021:763.
[132] Ibid, para 167. Google's comparison shopping service is one of Google's specialised search services. In response to queries, it returns product offers from merchant websites, enabling users to compare them.
[133] P. Georg Picht, ChatGPT, 'Microsoft and Competition Law Nemesis or Fresh Chance for Digital Markets Enforcement?', (2023) SSRN.
[134] Case T-612/17 *Google and Alphabet v Commission* (Google Shopping) [2021] ECLI:EU:T:2021:763, para 272.
[135] X, 'About Grok, Your Humorous AI Assistant on X', < https://help.x.com/en/using-x/about-grok #:~:text=Grok's%20access%20to%20real%2Dtime,a%20wide%20range%20of%20topics.&text=X%20may%20share%20with%20xAI,and%20other%20generative%20AI%20models>.
[136] P. Azoulay, J.L. Krieger and A. Nagaraj, 'Old Moats for New Models: Openness, Control, and Competition in Generative AI', (2024), Working Paper 32474, 14.
[137] R. Van den Bergh, COMPARATIVE COMPETITION LAW AND ECONOMICS, 454 (Edward Elgar Publishing 2017). R. Whish and D. Bailey, COMPETITION LAW (Tenth edition, Oxford University Press 2021), 852-857. W. Frenz, HANDBOOK OF EU COMPETITION LAW, 1096 (Springer Berlin Heidelberg 2016).
[138] OECD, *supra* note 4, 28. R. Van den Bergh, Ibid.
[139] Van den Bergh, *supra* note 137.
[140] OECD, *supra* note 4, 29.
[141] Stucke and Ezrachi, *supra* note 11.
[142] European Commission, *supra* notes 19.
[143] OECD, *supra* note 4, 24.
[144] B. Jin and T. Dotan, 'Tech Giants Spend Billions on AI Startups—and Get Just as Much Back', (2023) Wall Street Journal, <https://www.wsj.com/tech/ai/ai-deals-microsoft-google-amazon-7f624054>. See also US FTC, Partnerships Between Cloud Service Providers and AI Developers, (2025) 4.
[145] D. Coldewey, 'Amazon Doubles Down on Anthropic, Completing its Planned $4B Investment', (2024) TechCrunch <https://techcrunch.com/2024/03/27/amazon-doubles-down-on-anthropic-completing-its-planned-4b-investment/>. See also CMA, Amazon.com Inc.'s Partnership with Anthropic PBC,(2024), 1. < https://assets.publishing.service.gov.uk/media/6710ba44e84ae1fd8592f52c/Full_text_decision.pdf>.
[146] K. Hu, 'Google Agrees to Invest up to $2 Billion in OpenAI Rival Anthropic', (2023) Reuters < https://www.reuters.com/technology/google-agrees-invest-up-2-bln-openai-rival-anthropic-wsj-2023-10-27/>.
[147] J. Novet, 'Microsoft CFO says OpenAI Investment will Cut into Profit this Quarter', (2024) CNBC <https://www.cnbc.com/2024/10/30/microsoft-cfo-says-openai-investment-will-cut-into-profit-this-quarter.html>.

stakes, or long-term collaboration agreements, which can effectively function as *de facto* mergers.[148] By structuring these partnerships, big techs, such as Microsoft and Google, can contribute to grow their AI capabilities while avoiding the scrutiny of lengthy regulatory reviews, particularly in terms of merger control.[149] In other words, as competition enforcers strengthen enforcement against anti-competitive mergers, big techs have increasingly turned to partnerships with smaller firms as a strategic means of consolidating their market power. These AI partnerships may represent a new strategy by big techs seeking to form relationships with AI model developers, and they reflect how incumbents have responded to the generative AI boom.

Notably, a key factor driving this consolidation is the concentration of AI inputs, primarily computational resources and data, which are essential for training advanced AI models.[150] These partnerships permit the sharing of these critical inputs,[151] allowing AI developers to access data and cloud computing servers that would otherwise be difficult or costly to obtain independently.[152] For example, through partnerships, Amazon AWS supplies compute capacity to Anthropic, including access to AWS Trainium and Inferentia chips for building, training, and deploying Anthropic's future models.[153] As a result, AWS has become the primary cloud provider of Anthropic.[154] In return, Anthropic makes its existing models available on Amazon Bedrock and has committed to providing AWS customers worldwide with access to future generations of its foundation models on the platform.[155] Similarly, OpenAI has relied on Microsoft as its 'exclusive cloud provider' since the start of their partnership in 2019, a relationship which was reaffirmed in 2023.[156] According to the FTC, these partnerships provide partners access to key resources, including computing resources, intellectual property, key personnel, performance and financial data, training data, and chip co-development opportunities.[157]

Therefore, these partnerships - which in many ways resemble mergers - pose significant threats to fair competition within the AI supply chain.[158] Often vertical relationships in nature, these partnerships risk giving big techs control over multiple layers of the technology stack, from the hardware level to foundation models, and eventually to downstream applications and devices.[159]

Furthermore, these partnerships may alter incentives and produce anti-competitive effects for non-partnership AI developers who rely on AI inputs offered by big techs. For example, a big tech could restrict access to data and computing resources for AI developers outside its partnership network, placing them at a competitive disadvantage.[160] Accordingly, these partnerships could affect access to key inputs, potentially influencing competition not only for the partners themselves but also for non-partner AI developers.[161]

Additionally, these partnerships significantly reduce the incentives for the partnered firms to compete against each other, given the extensive financial and technological interdependence involved.[162] For example, OpenAI, which relies heavily on Microsoft for a large proportion of its funding and computing power, has little, if any, incentive to compete aggressively with Microsoft if it risks putting that support at stake.[163] Similarly, one would expect Microsoft to be reluctant to pursue a commercial strategy that involves directly taking on OpenAI's technology by building its own competing frontier models, given the billions it has invested in the startup.[164] Indeed, we are already witnessing such dynamics in action. For example, Microsoft decided to phase out its Cortana virtual assistant following the launch of its OpenAI-powered 'Copilot' assistant.[165]

Overall, big techs exert significant influence across AI markets, particularly AI infrastructure, characterized by a notable increase in strategic collaboration.[166] While other firms also operate in the AI supply chain, big techs currently appear the best positioned to scale investments and foster key partnerships.[167] This privileged position in AI development is unique to big techs, setting them apart from new entrants, including successful ones, that, while competing with their models, still heavily rely on big techs for financial support, computational resources, or data.[168] These dynamics pose a risk of further entrenching big techs' dominance, widening the competitive moat that shields them from future disruption.[169]

4. Antitrust in the EU, the US, and China

While the increasing development of AI may introduce novel challenges in regulation, it does not inherently shield tech firms from the application of existing laws.[170] As such, AI developers, operators, and deployers should anticipate that competition law will be enforced to mitigate potential anti-competitive harms arising in AI markets.[171]

This section, however, focuses primarily on the ongoing regulatory efforts in the EU, the US, and China, given their leadership in digital regulation and AI development. It examines initiatives designed to address monopolistic practices and unfair competition, particularly in AI input markets, such as the EU's Digital Markets Act (DMA), which imposes *ex ante* obligations on gatekeepers to ensure fair access to critical inputs; China's Interim Measures for the Management of Generative Artificial Intelligence Services (Interim Measures), which establish comprehensive rules for generative AI providers; and the US's fragmented policies aimed at fostering innovation in AI. This section also

[148] M. von Thun and D. Hanley, 'Stopping Big Tech from Becoming Big AI: A Roadmap for Using Competition Policy to Keep Artificial Intelligence Open for All', (2024) Open Markets Institute 2024 < https://blog.mozilla.org/wp-content/blogs.dir/278/files/2024/10/Stopping-Big-Tech-from-Becoming-Big-AI.pdf>.

[149] S. Ahmed, 'Pseudo-Acquisitions in AI: Tech Giants' New Strategy-Partner, Poach, and Dominate AI', (2024), Medium < https://medium.com/@sahin.samia/pseudo-acquisitions-in-ai-tech-giants-new-strategy-partner-poach-and-dominate-ai-60ada1eca83f>.

[150] von Thun and Hanley, *supra* note 148.

[151] US FTC, Partnerships Between Cloud Service Providers and AI Developers, (2025) 3. These resources include, to varying degrees, discounted access to computing resources for AI developer partners; intellectual property, and performance and financial data for CSP partners; and engineering personnel and training data for both partners.

[152] Ibid. Many also control the consumer and business interfaces. the development of the foundation models and the conditions under which they are released to the market.

[153] CMA, Amazon.com Inc.'s Partnership with Anthropic PBC, (2024) 4.

[154] Amazon, 'Amazon and Anthropic Deepen their Shared Commitment to Advancing Generative AI', (2024) <https://www.aboutamazon.com/news/company-news/amazon-anthropic-ai-investment>.

[155] CMA, *supra* note 153.

[156] US FTC, *supra* note 151, 20.

[157] Ibid, 21.

[158] von Thun and Hanley, *supra* note 148.

[159] Autorité de la Concurrence, of France, *supra* note 12.

[160] US FTC, *supra* note 151, 30.

[161] Ibid.

[162] von Thun and Hanley, *supra* note 148, 21. See also D.A. Hanley, 'Per Se Illegality of Exclusive Deals and Tyings as Fair Competition,' (2022) 37 Berkeley Tech. Law Journal 1057, 1073-74.

[163] von Thun and Hanley, *supra* note 148, 21.

[164] Ibid.

[165] Microsoft, 'End of support for Cortana,' Blog Post, <https://support.microsoft.com/en-us/topic/end-of-support-for-cortana-d025b39f-ee5b-4836-a954-0ab646ee1efa>.

[166] Stucke and Ezrachi, *supra* note 11.

[167] Ibid.

[168] Ibid.

[169] Ibid.

[170] Ibid.

[171] White & Case, 'AI Watch: Global regulatory tracker - United States', (2024), <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-states>. See also OECD, *supra* note 10, 45.

explores recent enforcement actions targeting potential abuse of dominance by big techs in AI infrastructure markets.

4.1. Antitrust in the EU

The EU continues to build and refine its regulatory framework to ensure digital transformation in the Digital Single Market.[172] This includes the DMA, the Digital Services Act,[173] the Artificial Intelligence Act,[174] and the Data Act.[175] Among these regulations, the DMA is the first comprehensive legal regime to regulate digital gatekeepers with the aim of making platform markets fairer and more contestable,[176] which introduces a system of accountability and obligations for digital platforms.[177]

The DMA's conduct rules do not apply to any digital platform service. They regulate the provisions of 'core platform services' (CPS) provided by 'gatekeepers' only.[178] With respect to the implementation of the DMA in AI services, two main scenarios should be considered. First, an AI developer may offer a CPS and meet the requirements of gatekeepers outlined in Article 3 of the DMA. Second, AI-powered functionalities may be integrated or embedded in existing designated CPSs and therefore be covered by the DMA obligations.[179] Additionally, the Commission can also reassess the services to evolve and integrate AI-powered services into CPSs in the future.

In the first scenario, as of January 2025, the Commission has designated seven gatekeepers - Google, Amazon, Apple, Booking, ByteDance, Meta, and Microsoft.[180] However, designated gatekeepers are not automatically considered gatekeepers across all their business activity.[181] Instead, the gatekeeper designation applies only to specific CPSs, as defined under Article 2(2) of the DMA. In line with these provisions, cloud computing services provided by Amazon, Google, and Microsoft - the three major providers discussed in Section 2.2 - could potentially be designated as CPSs, but none of these services have been designated as such to date.

Although Nvidia and OpenAI hold significant market power in AI markets, neither has been designated as gatekeepers by the Commission. Nvidia, despite controlling over 80 per cent of the GPU market, does not fall within the scope of the ten CPSs listed in Article 2(2) of the DMA. Additionally, while its market value exceeds the €75 billion threshold outlined in Article 3 of the DMA, its user base does not align with the required criteria.[182] Consequently, the Commission has launched an investigation into Nvidia under competition laws, which will be discussed below. Conversely, OpenAI is approaching the quantitative threshold criteria of the DMA to be designated as a gatekeeper,[183] and ChatGPT has over 100 million users, an estimated capital value of over $80 billion, and hundreds of thousands of business users developing specialised ChatGPT applications.[184] However, the chatbot services provided by ChatGPT may fall outside the existing CPS categories unless further clarification is provided. These examples highlight the challenges the Commission faces in the implementation of the DMA to AI markets.

In the second scenario, the DMA could regulate AI even though AI itself is not explicitly listed as a CPS. AI falls under the DMA's scope when embedded within designated CPSs, such as search engines, operating systems, and social networking services.[185] For example, Microsoft Windows, Apple iOS, and Google Android have been designated as CPS operating systems, and they already incorporate third-party AI model-driven services applications.[186] In this context, certain obligations outlined in Articles 5 and 6 of the DMA are relevant. Specifically, Article 5(2) may restrict gatekeepers from combining data collected across CPSs for AI development purposes.[187] Furthermore, Articles 6(9), 6(10), and 6(11) may facilitate data portability and provide indirect access to such data for AI developers, subject to the consent of end-user and business-user.[188] Notably, Article 6(11) imposes the data access obligation specifically in the context of search engine services. These date-related obligations could partially mitigate antitrust risks associated with gatekeepers' restrictions on data access. However, data-related practices remain relatively invisible, due to challenges in measuring the full scope and scale of data held by big techs within their ecosystems.[189] This makes it difficult to determine whether refusals of data portability or certain differences in the treatment of new entrants accessing data may violate the DMA.

[172] European Council and the Council of the European Union, Digital single market for Europe, (2020) < https://www.consilium.europa.eu/en/policies/digital-single-market/>. See also European Commission, A European fit for the Digital Age, < https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age_en>.

[173] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and Amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L 277.

[174] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L2024/1689.

[175] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on Harmonised Rules on Fair Access to and Use of Data and Amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act) [2023] OJ L, 2023/2854.

[176] A.C. Witt, 'The Digital Markets Act – Regulating the Wild West', (2023) 60 Common Market Review.

[177] Formally, the DMA is not classified as a competition law measure, as its legal basis is found in Article 114 TFEU. However, it complements the enforcement of the EU and national competition rules on restrictive agreements, abuse of dominance, and mergers. See I. Maher, 'Regulatory Design in the EU Digital Markets Act: No Solo Run for the European Commission', (2024) 12 Journal of Antitrust Enforcement, 273–279; A. Bacchiega and T. Tombal, 'Agency Insights: The First Steps of the DMA Adventure', (2024) 12 Journal of Antitrust Enforcement, 189–194.

[178] Digital Markets Act, Article 1(2). See also Witt, *supra* note 176.

[179] Digital Markets Act, Article 2.

[180] See the website of the European Commission: https://digital-markets-act.ec.europa.eu/gatekeepers_en.

[181] Georg Picht, *supra* note 133, 10.

[182] Nivida's market value is about €1.8 trillion as of 2025. See https://live.euronext.com/en/product/equities/US67066G1040-BGEM

[183] Digital Markets Act, Article 3.

[184] Martens, *supra* note 40.

[185] European Commission, *supra* note 19. See also Martens, *supra* note 40.

[186] Martens, *supra* note 40.

[187] Articles 5(2) of the DMA prohibits gatekeepers to combine CPSs data with personal data from any services unless there is explicit consent. See also G. Monti and A. de Streel, 'Data-Related Obligations in the DMA', (2024) <https://cerre.eu/wp-content/uploads/2024/01/Data-Related-Obligations-in-the-DMA_FINAL.pdf>.

[188] Article 6(9) of the DMA aims to prohibit gatekeepers from using data from end users, reflecting the obligations of the GDPR. Article 6(10) of the DMA states that gatekeepers should provide business users and third parties authorised by a business user with effective, high-quality, continuous and real-time access to, and use of, aggregated and non-aggregated data. Article 6(11) of the DMA asks gatekeepers to allow third party online search engine providers fair, reasonable and non-discriminatory access to the ranking, query, click and view data generated by a gatekeeper's search engine. It aims to address the asymmetry of information that exist between search engines acting as gatekeepers and other search engines, with the intention to feed fairer competition. See S. Stalla-Bourdillon and B. da Rosa Lazarotto, 'Search queries and anonymisation: How to read Article 6(11) of the DMA and the GDPR together?' (2024) European Law Blog. Martens, *supra* note 40.

[189] G. Sastry et al, 'Computing Power and the Governance of Artificial Intelligence', (2024) ARXIV, < https://arxiv.org/pdf/2402.08797>.

In addition to the DMA, EU competition law is likely to be a powerful tool in addressing the concerns discussed in Section 3.[190] The primary legislation in this area is the Treaty on the Functioning of the European Union (TFEU), specifically Articles 101 and 102, which prohibit anti-competitive agreements and abuse of dominant market positions. An important additional instrument of competition law is the EU Merger Regulation (EUMR),[191] which applies to concentrations between undertakings that have an EU dimension.[192] Although the AI Act does not impose specific obligations on AI developers to ensure contestable markets, it does refer to existing EU competition laws,[193] highlighting the interconnectedness of these regulations.

With respect to competition law enforcement in AI markets, the Commission initiated an investigation in January 2024 to assess the partnership between Microsoft and OpenAI under the EUMR.[194] However, this investigation was dropped after concluding Microsoft had not gained control over OpenAI.[195] Likewise, on 18 September 2024, the Commission, which had been reviewing the transaction following a referral by several Member States, announced that it would not take a decision on the matter due to a lack of jurisdiction.[196] Notably, although the Commission unconditionally approved the proposed acquisition of Run:ai by Nvidia under the EUMR,[197] in February 2025, Nvidia filed a lawsuit against the Commission for accepting the referral from the Italian NCA to review the transaction, despite it falling below both the EUMR and national merger control thresholds.[198] Additionally, the Commission launched a call for contributions to all interested stakeholders to gather insights on the level of competition in generative AI, and the potential role of EU antitrust authorities.[199] It is also worth noting that the Commission is investigating Nvidia for potential bundling and leveraging practices concerning Nvidia's dominance in the GPU market.[200] In addition to competition law and regulations that impose obligations on gatekeepers, the EU has adopted strategic policies and allocated billions in public funding to establish AI infrastructure, including public data centers and supercomputers, with the aim of promoting AI innovation and development. The Commission plans to establish 'AI Factories'.[201] These are open ecosystems formed around European public supercomputers and bringing together key inputs and human resources needed for the development of generative AI models and applications, including AI-dedicated supercomputers, and associated data centers in proximity or connected via high-speed networks.[202]

Likewise, according to the Data Act, the roll-out of European Common Data Spaces could improve interoperability and access to large volumes of high-quality data,[203] which are therefore key to providing a varied data ecosystem for AI startups. With regard to computational resources, the EU has created a large public infrastructure for computing capacity located across six Member States since the launch of the EuroHPC Joint Undertaking in 2018.[204] These policies suggest that the EU is also addressing potential risks related to data and computational resources access through a public policy lens.

Collectively, this subsection illustrates that the EU does not integrate the regulation and promotion of AI innovation into a single framework. Specifically, the DMA and competition laws focus on imposing obligations and regulatory measures to ensure fairness and contestability in AI markets. In contrast, broader EU policies aim to stimulate innovation and support AI development. This approach - separating regulatory constraints from innovation incentives – distinguishes the EU's strategy from those of the US and China, which will be explored in the following subsections.

4.2. *Antitrust in the US*

Currently, the US lacks comprehensive federal legislation or regulations governing AI markets.[205] Many of the proposed bills emphasize the development of voluntary guidelines and best practices for AI systems, reflecting a cautious approach to regulation aimed at fostering innovation without imposing strict mandates.[206]

Former US President Biden signed an Executive Order (2023 EO) on Safe, Secure, and Trustworthy Artificial Intelligence on 30 October 2023.[207] The 2023 EO outlines eight principles and priorities for the use of AI. The second principle emphasizes that the US should promote responsible innovation, competition, and collaboration in AI development. However, this document was rescinded on 20 January 2025.[208]

On 23 January 2025, President Trump signed a new Executive Order

[190] S. Hua and H. Belfield, 'Effective Enforceability of EU Competition Law Under Different AI Development Scenarios', in AAAI/ACM Conference on AI, Ethics, and Society, 1-11 (2023).

[191] Council Regulation (EC) No 139/2004 of 20 January 2004 on the Control of Concentrations Between Undertakings (the EC Merger Regulation), [2004] OJ 2004 L24/1.

[192] R. Whish and D. Bailey, Competition Law, (Tenth edition Oxford University Press 2021) 50.

[193] Artificial Intelligence Act, Recital para 45, and Article 74(2).

[194] European Commission, *supra* note 19. See also Press Releases, European Commission, Commission Launches Calls for Contributions on Competition in Virtual Worlds and Generative AI (Jan. 8, 2024), <https://ec.europa.eu/commission/presscorner/detail/en/ip_24_85>.

[195] European Commission, *supra* note 194.

[196] European Commission, Press Releases, 'Commission takes note of the withdrawal of referral requests by Member States concerning the acquisition of certain assets of Inflection by Microsoft', (2024), < https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4727>.

[197] The European Commission, Press Release, Commission approves acquisition of Run:ai by NVIDIA, 20 December 2024, < https://ec.europa.eu/commission/presscorner/detail/en/ip_24_6548>.

[198] F. Yun Chee, 'Nvidia takes EU antitrust regulators to court for probing AI startup Run:ai bid', (2025) Reuters <https://www.reuters.com/technology/nvidia-takes-eu-antitrust-regulators-court-probing-ai-startup-runai-bid-2025-02-24/>. See also DigWatch, 'Nvidia takes legal action against EU antitrust investigation', (2025), <https://dig.watch/updates/nvidia-takes-legal-action-against-eu-antitrust-investigation#:~:text=Nvidia%20has%20filed%20a%20lawsuit,its%20powers%20over%20minor%20transactions>.

[199] European Commission, *supra* note 194.

[200] N. Khan, 'Nvidia Faces EU Antitrust Scrutiny as Bundling Practices Come Under Fire', (2024) Yahoo!Finance < https://finance.yahoo.com/news/nvidia-faces-eu-antitrust-scrutiny-170539193.html>. See also F.Y Chee, 'EU watchdog probes potential Nvidia hardware bundling as it scrutinises Run:ai deal', (2024) Reuters <https://www.reuters.com/technology/eu-watchdog-probes-nvidia-hardware-bundling-it-scrutinises-runai-deal-2024-12-04/>.

[201] The European Commission, AI Factories, < https://digital-strategy.ec.europa.eu/en/policies/ai-factories#:~:text=The%20Commission%20has%20identified%20the,into%20the%20%22AI%20continent%22>.

[202] European Commission, Communication on Boosting Startups and Innovation in Trustworthy AI, COM (2024) 28 final, (24.1.2024) 5.

[203] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act), OJ L, 2023/2854. The Data Act aims to unlock large volumes of IoT-generated data, empowering AI startups to utilise this untapped resource.

[204] Draghi, *supra* note 114, 30. LEONARDO in Bologna, Italy, LUMI in Kajaani, Finland, and MareNostrum 5 in Barcelona, Spain. According to the Commission, in 2024, through the EuroHPC JU, the Commission and Member States will invest a total amount of EUR 2.1 billion in acquiring new or upgrading existing EuroHPC supercomputers with AI capabilities, the creation of supercomputing services in AI and developing AI-oriented microprocessors and skills support. See also European Commission, Communication on Boosting Startups and Innovation in Trustworthy AI, COM (2024) 28 final, (24.1.2024) 18.

[205] IMF, The Economic Impacts and the Regulation of AI: A Review of the Academic Literature and Policy Actions, (2024) 61. See also White & Case, *supra* note 171.

[206] White & Case, *supra* note 171.

[207] The full text is available at: https://www.nist.gov/artificial-intelligence/executive-order-safe-secure-and-trustworthy-artificial-intelligence.

[208] Ibid.

(2025 EO) on Removing Barriers to American Leadership in Artificial Intelligence.[209] Notably, this policy focuses on revoking directives perceived as restrictive to AI innovation, paving the way for the 'unbiased and agenda-free' development of AI systems.[210] It is now the policy of the US to sustain and enhance its global AI dominance in order to promote economic competitiveness and revoke certain pre-existing AI policies that hinder the US AI innovation. While policy developments continue to evolve, the Trump administration's focus on technological leadership and reduced regulatory oversight is a significant shift from past approaches taken by the former Biden administration, for example, by prioritizing AI infrastructure investments such as the recent Stargate Project with OpenAI.[211]

Notably, a regulatory initiative has been taken at the state level - the California Assembly Bill 2013 (AB 2013) on Generative Artificial Intelligence: Training Data Transparency – which was signed into law on 28 September 2024.[212] This legislation mandates that AI developers publicly disclose information regarding the data used to train and test their generative AI models.[213] However, as it is a state-level measure rather than a federal regulation, it falls outside the scope of this paper and will not be examined further.

Overall, these policies reflect the US's ambition to prioritize innovation over regulation in the AI sectors. At the same time, the US competition authorities also employ various measures within competition laws to regulate AI infrastructure markets.

US antitrust laws are designed to prevent monopolistic practices and promote fair competition. The primary legislative instruments are the Sherman Antitrust Act[214] and the Clayton Act,[215] and the Federal Trade Commission Act.[216] These competition laws are enforced by the Federal Trade Commission (FTC) and the Department of Justice (DOJ). However, unlike the EU and China, the US currently lacks competition laws specifically tailored to the digital sector.[217]

Based on the US competition legal framework, the FTC has taken actions that are particularly relevant to the governance of data and computational resources. On the one hand, the FTC clarified that simply possessing large amounts of data is not inherently unlawful.[218] However, it further explained that even when companies adopt responsible data collection practices, their control over data may still create barriers to entry or expansion, thereby hindering the full flourishing of fair competition.[219] On the other hand, the FTC highlighted that the highly concentrated markets for specialized chips are, or could be, less competitive without appropriate competition policies and antitrust enforcement. In 2022, the FTC filed a law enforcement action to block Nvidia's acquisition of Arm,[220] alleging that the merger would have suppressed competition in multiple processor markets, including those for AI chips used by cloud service providers.[221] After more than two months into its litigation with the FTC, Nvidia announced its decision to abandon the acquisition of Arm.[222] It has also been noted that the FTC announced the opening of an antitrust investigation into Microsoft's hiring of key developers from the AI start-up, Inflection.[223] However, the outcome of the FTC's review, which has been ongoing since June 2024, remains pending.[224]

Notably, the FTC launched inquiries into generative AI investments and partnerships in January 2024.[225] The Agency's 6(b) inquiries targeted big techs, including Google, Amazon, Anthropic PBC, Microsoft, and OpenAI.[226] Following those inquiries, the FTC and the DOJ reached an agreement in June 2024 to divide investigations into certain companies for potential anti-competitive conduct. The DOJ has been investigating Nvidia and its dominant position in supplying high-end semiconductors essential for AI computing since September 2024, although no further details have been disclosed. Likewise, the FTC is set to investigate whether Microsoft and its partner, OpenAI, hold unfair advantages in the rapidly evolving AI technology sector, particularly concerning the development of LLMs.[227] Recently, on 17 January 2025, the FTC published a report reflecting its investigation into partnerships between cloud services providers and AI developers.[228] The report sheds light on key partnership terms and examines how these partnerships could potentially impact access to critical inputs for AI development, highlighting extensive exchanges between partners across the chip, data, model, and application layers.

Accordingly, the US lacks comprehensive AI legislation governing developers, users, operators, and deployers of AI markets.[229] Instead, its

[209] The White House, Removing Barriers to American Leadership in Artificial Intelligence, (2025) <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/>.

[210] Ibid.

[211] OpenAI, 'Announcing The Stargate Project', (2025) <https://openai.com/index/announcing-the-stargate-project/>.

[212] Assembly Bill 2013 (AB 2013) on Generative Artificial Intelligence: Training Data Transparency, <https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB2013>. According to this legislation, AI developers are required to post specific documentation about the training data on their public websites by 1 January 12026 (or prior to substantial modifications).

[213] Ibid. See also S.A. Choudary, 'Assembly Bill 2013: Generative Artificial Intelligence: Training Data Transparency', 25 February 2025, Securiti <https://securiti.ai/blog/assembly-bill-2013-generative-artificial-intelligence-training-data-transparency/>.

[214] The Sherman Antitrust Act regulates agreements and single-firm conduct.

[215] The Clayton Act primarily regulates mergers.

[216] The Federal Trade Commission Act largely runs in parallel to the other statutes, though may have a slightly broader reach.

[217] B. Hoffman et al, USA Digital Markets Regulation Handbook, (2024), <https://www.clearygottlieb.com/-/media/files/rostrum/dmrh/dmrh-usa_r1?utm_campaign=DMRH_April&utm_medium=pdf&utm_source=download&utm_content=usa#:~:text=effects%20on%20them.-,There%20are%20currently%20no%20digital%2Dspecific%20competition%20laws%20in%20the,2024%2C%20neither%20bill%20has%20passed>.

[218] FTC, *supra* note 19.

[219] Ibid.

[220] FTC, Nvidia/Arm, (2022), <https://www.ftc.gov/legal-library/browse/cases-proceedings/2110015-nvidiaarm-matter>.

[221] Ibid.

[222] Nvidia, NVIDIA and SoftBank Group Announce Termination of NVIDIA's Acquisition of Arm Limited, (2022), <https://nvidianews.nvidia.com/news/nvidia-and-softbank-group-announce-termination-of-nvidias-acquisition-of-arm-limited>.

[223] B. Fung, 'As AI Booms, Microsoft's Deal with a Startup Comes under Federal Investigation', CNN, 6 June 2024, <https://edition.cnn.com/2024/06/06/tech/ftc-microsofts-ai-investigation>.

[224] It should also be noted that this case focuses on the acquisition of experts, which lies beyond the scope of the present study focused on data and computational resources. Thus, the eventual outcome of this case is unlikely to affect the conclusion drawn in this article.

[225] The agency issued an order to Alphabet, Inc., Amazon.com, Inc., Anthropic PBC, Microsoft Corp., and OpenAI, Inc. to provide information about participations and investment to assess their impact on competition.

[226] FTC, FTC Launches Inquiry into Generative AI Investments and Partnerships, (2024), <https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships>.

[227] Debevoise & Plimpton, Navigating Antitrust in the Age of AI: Global Regulatory Scrutiny and Implications, (2024), <https://www.debevoise.com/insights/publications/2024/06/navigating-antitrust-in-the-age-of-ai-global

[228] FTC, Behind the FTC's 6(b) Report on Large AI Partnerships & Investments, (2025), <https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2025/01/behind-ftcs-6b-report-large-ai-partnerships-investments>. In January 2024, the agency launched a market study into these partnerships using its 6(b) authority to better understand the partnerships and their potential implications for competition and consumers.

[229] White & Case, *supra* note 171.

strategy is characterized by fragmented policies that aim to foster innovation,[230] while reinforcing 'traditional American values.'[231] Meanwhile, there are no specific competition laws tailored to the digital and AI sectors in the US, although the FTC and the DOJ have launched investigations into potential risks arising from markets concentration in AI infrastructure, including data and computational resources.

4.3. Antitrust in China

Unlike the US, China adopted the Interim Measures for the Management of Generative Artificial Intelligence Services (Interim Measures) in July 2023,[232] imposing a broad spectrum of obligations on generative AI service providers, spanning intellectual property (IP), data security, privacy, ethics, and competition.[233] Similar to the EU AI Act, the Interim Measures do not apply to generative AI technologies used for 'research purposes and not deployed to the market.[234]

Regarding antitrust in AI-related markets, Article 4(3) of the Interim Measures stipulates one of the principles that it is prohibited to exploit advantages in algorithms, data, platforms, or other similar resources to engage in monopolistic or unfair competitive practices. This principle closely resembles Article 9 of the Anti-Monopoly Law (AML).[235] Additionally, Article 6 sets out rules to encourage AI development, particularly by supporting the establishment of AI infrastructure. To illustrate, the second paragraph of Article 6 stipulates that: '…[E]fforts shall be made to advance the development of generative artificial intelligence infrastructure and public training data resource platforms. Collaborative sharing of computing power resources shall be promoted to enhance their utilization efficiency. Public data shall be classified, graded, and opened in an orderly manner to expand high-quality public training data resources. The adoption of secure and trustworthy chips, software, tools, computing power, and data resources shall be encouraged.' Furthermore, Article 7 imposes extensive obligations on AI service providers concerning data training processes, including pre-training and optimization training, ensuring compliance with relevant laws. However, the examination of these Interim Measures reveals that concrete rules imposing obligations on these providers or addressing these concerns regarding concentration within the AI infrastructure markets, as explored in Section 2, remain insufficient.

Scholar argues that Interim Measures prioritize the growth of generative AI technology over regulatory constraints.[236] This article acknowledges that the document does emphasize the encouragement of AI development and innovation in several provisions.[237] However, it also explicitly sets out principles for regulating generative AI in Articles 4 and imposing obligations on AI service providers in Articles 7 to 15. Therefore, the extent to which it favors encouragement over constraint remains an open question, requiring further empirical investigation.

In the field of competition law, China has adopted and revised both hard and soft law instruments to regulate digital markets. For example, the Chinese competition authorities issued the Guidelines on Anti-Monopoly in the Field of Platform Economy (Platform Guidelines) first in February 2021.[238] Following a year of practice, the State Administration for Market Regulation (SAMR) proposed incorporating certain rules of the Platform Guidelines into the Amended AML to the National People's Congress and its Standing Committee, which exercise the legislative power of the State. An examination of these instruments reveals that the rules related to competition issues outlined in Articles 4 and 6 of the Interim Measures are, in fact, largely consistent with those under the Amended AML.

With respect to competition law enforcement, the latest development in China is that the SAMR announced its investigation of Nvidia for its suspected violation of the AML on 9 December 2024.[239] According to the announcement, Nvidia is under investigation in accordance with the law, on suspicion of violating the AML and the Announcement of the State Administration for Market Regulation on the Conditional Approval of Nvidia Corporation's Acquisition of Mellanox Technologies, Ltd. with Additional Restrictive Conditions (SAMR Announcement [2020] No. 16).[240] The decision of its investigation has not yet been adopted by the SAMR. Additionally, the SAMR announced another investigation of Google for its suspected violation of the AML on 4 February 2025.[241]

Overall, China has introduced a regulation governing generative AI, which simultaneously integrates provisions to encourage innovation within the same legal instrument. However, it is noted that the rules related to fair competition outlined in Interim Measures largely mirror those set out in the Amended AML and do not include specific provisions addressing concrete restrictions on access to data and computing resources. Nevertheless, the SAMR has recognized these potential antitrust risks and has initiated an investigation into Nvidia's dominance in the hardware market, signalling a growing focus on competition issues related to AI infrastructure.

[230] Software Improvement Group, AI Legislation in the US: A 2025 Overview, (2025) <https://www.softwareimprovementgroup.com/us-ai-legislation-overview/>.

[231] Ibid.

[232] Shengchengshi Rengong Zhineng Fuwu Guanli Zanxing Banfa (生成式人工智能服务管理暂行办法) [Interim Measures for the Management of Generative AI Services] (promulgated by the CAC, the NDRC, the Ministry of Education, the MOST, the MIIT, the Ministry of Public Security, and the National Radio and Television Administration, Jul. 10, 2023, effective Aug. 15, 2023) [hereinafter Interim Measures for Generative AI], <http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm>.

[233] See the Interim Measures, ibid, Articles 4 and 7.

[234] Article 2 of the Interim Measures: [...] These Measures do not apply where industry associations, enterprises, education and research institutions, public cultural bodies, and related professional bodies, etc., research, develop, and use generative AI technology, but have not provided generative AI services to the public."

[235] Anti-Monopoly Law of the People's Republic of China [中华人民共和国反垄断法] Standing Committee of the National

People's Congress, 24 June 2022. Article 9 states that: Undertakings shall not use data or algorithms, technology, capital advantages, or platform rules, etc., to engage in the monopolistic practices prohibited by this Law.

[236] Angela Huyue Zhang, 'The Promise and Perils of China's Regulation of Artificial Intelligence', Columbia Journal of Transnational Law (forthcoming).

[237] See Articles 3, 5, 6 of the Interim Measures. For example, Article 5: Encourage the innovative application of generative AI technology in each industry and field, generate exceptional content that is positive, healthy, and uplifting, and explore the optimization of usage scenarios in building an application ecosystem.[...]"

[238] Guidelines of Anti-Monopoly Platform Economy [国务院反垄断委员会关于平台经济领域的反垄断指南] the Anti-Monopoly Commission of the State Council, 7 February 2021.

[239] See Global Times, China's Top Market Regulator Launches Probe into Nvidia over Suspected Breach of Anti-Monopoly Law, (2024), < https://www.globaltimes.cn/page/202412/1324672.shtml>.

[240] 市场监管总局关于附加限制性条件批准英伟达公司收购迈络思科技有限公司股权案反垄断审查决定的公告》 (市场监管总局公告 [illegible]2020[illegible] 第16号) [Announcement of the State Administration for Market Regulation on the Decision Regarding the Anti-Monopoly Review of the Conditional Approval of NVIDIA Corporation's Acquisition of Equity in Mellanox Technologies, Ltd. (Announcement No. 16 [2020] of the State Administration for Market Regulation)], (2020), <https://www.samr.gov.cn/xw/zj/art/2024/art_ed4d3090401741a0894e475d35db652b.html>.

[241] SAMR, 谷歌公司涉嫌违反反垄断法市场监管总局依法决定立案调查 [Google is suspected of violating the AML - the State Administration for Market Regulation has decided to initiate an investigation], 4 February 2025, https://www.samr.gov.cn/xw/zj/art/2025/art_396a9ab3aa6d4c4bbd40833815afd245.html>.

4.4. Comparative analysis

The EU, the US, and China have each taken measures to investigate AI markets, with a particular focus on potential gatekeepers in AI infrastructure, including data and computational resources. Notably, all three jurisdictions have launched investigations into Nvidia for potential abuse of dominance in the computational resources market. Simultaneously, the EU and the US (and the UK) have also scrutinized the partnership between OpenAI and Microsoft according to merger control regulations. These public interventions by competition authorities underscore the potential risks of the concentration, particularly in data and computational resources controlled by big techs.

However, regarding competition legal frameworks, this article observes that none of the three jurisdictions has introduced specific competition legislation tailored to AI markets. Instead, the EU and China have enacted broader AI regulations: the AI Act in the EU and the Interim Measures for the Management of Generative Artificial Intelligence Services in China. While the EU's AI Act does not explicitly address AI-related antitrust risks, it indirectly references existing competition laws. Furthermore, an analysis of the DMA suggests that this regulation could potentially be enforced in AI markets to address challenges in AI input markets. In contrast, China's Interim Measures explicitly clarify fair competition as a legal principle, though the relevant provision largely mirrors the existing rules under the Amended AML. Neither framework, however, establishes antitrust provisions specifically targeting AI input markets. The US, unlike the EU and China, lacks comprehensive federal AI legislation and has not developed a competition law specifically addressing AI-related challenges. Instead, US policymakers have prioritized an innovation-centric strategy, focusing on policies aimed at fostering innovation in AI development and securing its global AI dominance.

Generally, the EU, the US, and China have adopted a cautious approach to imposing strict regulations on competition in AI infrastructure markets and broader AI-related markets. It is understandable, given the importance of AI development, as overly restrictive competition interventions could risk hampering innovation and reducing global competitiveness in this rapidly evolving sector. In this context, the different strategies adopted by these three jurisdictions in balancing innovation and regulation highlight the complexity of governing AI markets.

5. Conclusions

Over the past two decades, it has been observed that numerous digital markets are tipping toward dominance by one or two powerful firms, which have expanded their ecosystems to entrench their positions.[242] The lessons learned from the regulation of digital platforms underscore the importance of proactive responses from competition authorities to fulfil their missions. A failure to act swiftly - both substantively and in enforcement - risks repeating past policy shortcomings, particularly in the rapidly evolving AI sector.

This article examines the primary sources of access to data and computational resources for AI developers. It identifies that AI developers typically access data from three main sources: public data, synthetic data, and proprietary data. However, new entrants face significant barriers when attempting to access these data sources: public data is increasingly restricted by big techs and is becoming exhausted; synthetic data presents certain risks; and proprietary data is costly and often controlled by big techs. Similarly, computational resources can be accessed either through ownership of hardware, such as AI chips, or through cloud computing services. However, both of these sources are highly concentrated and largely controlled by big techs.

Against this backdrop, the article identifies three key antitrust risks arising from the market power of big techs in controlling these critical inputs. First, big techs may raise entry barriers to protect their dominance in digital platforms and AI markets. Second, they may leverage their advantages in data collection and computational resources to develop AI models and prefer their own AI services, thereby solidifying or extending their dominance across the AI supply chain. Third, new entrants are often compelled to collaborate with dominant firms through vertical integrations, exchanging AI technologies for access to data and cloud computing resources.

Building on this analysis, this article explores the legal responses and recent initiatives adopted by the EU, the US, and China in regulating AI infrastructure markets and broader AI-related markets. Firstly, it reveals that competition authorities in these three jurisdictions have learned from the emergence of digital markets at the beginning of this century and are striving to avoid repeating past mistakes in regulating AI markets. As a result, all three have intervened in addressing potential abusive practices by big techs in AI infrastructure markets. Secondly, both the EU and China have adopted broadly framed yet applicable rules that could potentially address antitrust risks in AI markets, namely the EU's DMA and China's Interim Measures for the Management of Generative Artificial Intelligence Services. However, neither of the two jurisdictions explicitly introduces competition law to tackle antitrust risks related to accessing data and computational resources for AI development. While both legal regimes aim to foster AI innovation, they employ different regulatory strategies: the EU adopts its AI policies in separate instruments from the DMA (and the AI Act), whereas China integrates its encouragement of AI innovation directly into the Interim Measures. Thirdly, the US, in contrast, lacks a comprehensive AI regulation and tailored antitrust framework for digital platforms and AI markets at the federal level. Instead, the US prioritises an innovation-centric strategy, focusing on policies designed to foster innovation in AI development and secure its global AI dominance.

The distinct regulatory strategies adopted by the EU, the US, and China reflect a cautious approach to imposing strict regulations, given the strategic importance of AI development. While there is no consensus among the three jurisdictions on what competition regulation of AI infrastructure should entail, each seeks to strike a balance between safeguarding competition and fostering innovation.

At least based on the current regulatory frameworks, the US leans more toward promoting AI innovation, particularly by encouraging global AI leadership – potentially reinforcing the power of existing tech giants. In contrast, the EU prioritises regulatory intervention, aiming to protect market contestability by restricting gatekeepers from data combination and by improving data access for new entrants. China adopts a middle ground approach, embedding support for AI innovation within a regulatory framework that imposes various obligations, thereby sending a clearer signal that AI innovation is expected to occur within defined regulatory boundaries (though these boundaries remain ambiguous). However, at this early stage – prior to the widespread adoption of AI – regulatory efforts in both the EU and China may be more indicative of underlying policy orientations than of actual enforcement effectiveness. Moving forward, empirical research assessing the effectiveness of those regulations in mitigating antitrust risks in AI infrastructure and the impact of AI innovation polices will be essential.

Declaration of competing interest

I declare that I have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

[242] Stucke and Ezrachi, *supra* note 11.