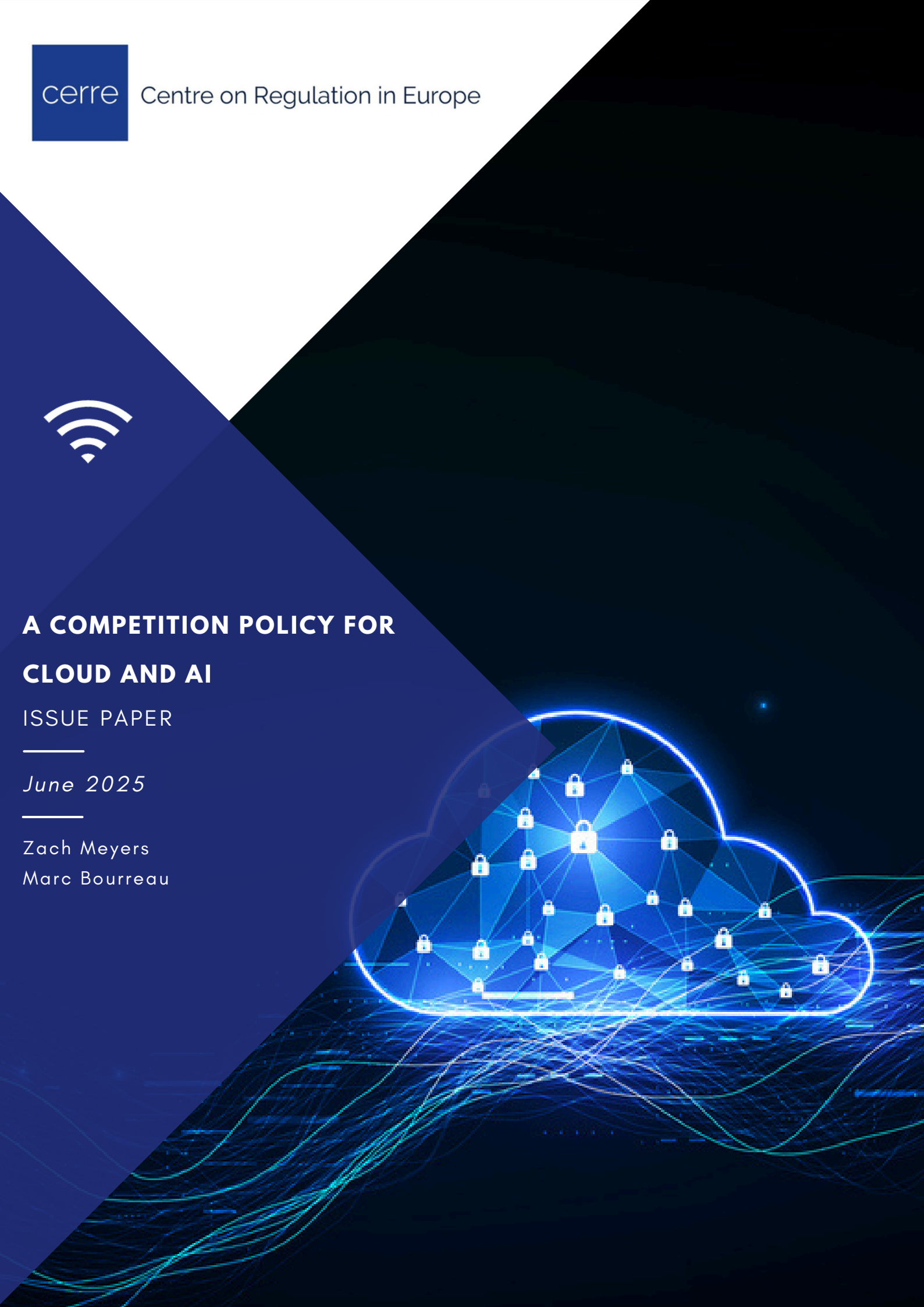
cerre Centre on Regulation in Europe

A COMPETITION POLICY FOR CLOUD AND AI

ISSUE PAPER

June 2025

Zach Meyers
Marc Bourreau



cerre

Issue Paper

A Competition Policy for Cloud and AI

Zach Meyers
Marc Bourreau

June 2025

As provided for in CERRE's bylaws and procedural rules from its "Transparency & Independence Policy", all CERRE research projects and reports are completed in accordance with the strictest academic independence.

The project, within the framework of which this report has been prepared, received the support and/or input of the following CERRE member organisations: Amazon, Arcep, BIPT, and Microsoft. However, they bear no responsibility for the contents of this report. The views expressed in this CERRE report are attributable only to the authors in a personal capacity and not to any institution with which they are associated. In addition, they do not necessarily correspond either to those of CERRE, or of any sponsor or of members of CERRE.

info@cerre.eu – www.cerre.eu

Executive Summary

As European leaders focus on boosting Europe's competitiveness, they must ensure businesses increase their productivity. Better deployment of new technology will be an important way to achieve that. Artificial intelligence (AI) could eventually boost competition across the economy – both by disrupting incumbents in the tech sector and by helping firms in many sectors become more productive, particularly services industries which have traditionally struggled to use technology to boost their output and efficiency.

To maximise the economic benefits of AI foundation models, competition authorities must ensure that competition thrives throughout the value chain – so that AI remains as cheap, high-quality and widely available as possible and incentives to innovate are maximised.

This issue paper provides an overview of how effective competition between providers of AI foundation models is functioning today – and how the sector could develop. We consider upstream inputs, in particular the provision of computing power and data for AI foundation models. This report does not consider in detail downstream uses of AI – such as competition between applications and devices that deploy AI, or future markets that might develop using AI, such as markets for AI agents.

Despite competition authorities' initial worries, **several potential barriers to entry for AI developers have proven less significant than feared.** Currently, there is a thriving ecosystem of diverse AI foundation models. However, there is uncertainty about the future trajectory of competition in the sector, thanks to the significant role of a few large firms across the AI value chain, the potentially growing dependence of smaller models on larger ones, a shift towards lower up-front costs and higher operational costs, and the growing importance of open AI models. At the moment, these shifts suggest the possibility of sustainable and intense competition. However, there remain some potential chokepoints – such as access to certain datasets, including a user's own usage history of a service – where targeted regulatory interventions may prove necessary in order to help ensure the sector remains contestable.

The European Commission and national competition authorities are particularly focused on AI developers' relationship with the largest cloud computing providers (called 'hyperscalers'). The relationship between these giants and AI developers is multifaceted given the vertical integration of the hyperscalers across the AI value chain. Hyperscalers may:

- compete to provide 'accelerated compute' (the specialised computing power needed to train AI foundation models and to allow those models to produce a respond to user requests) to AI developers;
- provide an important channel to market for AI foundation models;
- invest in many AI developers' firms;
- provide their own AI foundation models in competition with AI developers; and
- be significant users of AI, which is often integrated into their other digital services.

We suggest that:

- Given the diversity of AI models today, the significant growth in the number of models, and the ability of many diverse types of AI developers to attract investment, the AI sector looks competitive even on **traditional static metrics of market power.**
- **However, these metrics may understate the level of competition.** A dynamic competition analysis would point to the frequent radical innovations achieved by AI firms, large investments by large and small AI firms and private equity and venture capital investors, and the diversity of business models in the sector as companies experiment. Barriers to entry in AI continue to drop, with many AI developers now using large, open-source (or open-weights) foundation models as the basis for developing more specialised services – avoiding the huge costs of building an entirely new model. Similarly, the shift away from large-scale training towards more fine-tuning and inference might provide more scope for smaller cloud computing companies, which do not have the large data centres of the hyper-scalers, to serve AI developers. While overall promising, these developments also come with some risks to competition – such as growing dependence of some AI developers on large foundation models, and on particular datasets – which may warrant regulatory scrutiny and, potentially, targeted interventions.
- 'Static' measures of competition in the general cloud computing sector, such as market shares and profit margins, have concerned European competition authorities. However, **the provision of computing power to AI firms (known as 'accelerated compute') requires its own analysis**. The hyperscalers have very strong positions, and collectively they appear to be winning an even higher proportion of business from AI developers than from general cloud computing customers. This is a potential concern. However, **dynamic measures of competition – such as** levels of investment and innovation, and the ability of smaller players to attract funding – suggest that there are opportunities for smaller firms to grow in the market. **Competition authorities should scrutinise the sector over time to see how these opportunities play out**.
- To date, **the influence of hyperscalers has been, on balance, positive for competition in AI:** significantly lowering barriers to entry for AI developers and providing support to smaller AI developers, with limited evidence that those AI developers' choices have been unduly constrained.
- The concentrated nature of some parts of the AI value chain, and the strong position of the hyperscalers across multiple parts of the value chain, create the **possibility of future harm to competition** – in particular, if concerns about potential lock-in turn out to be substantial and are not addressed, and if currently relatively open platforms, models, and services become more closed over time. However, many firms in the sector are generally adopting relatively open approaches and it is not yet clear whether (or when) they will have the incentive or the ability to adopt more closed business models. If they do, authorities should examine carefully how these changes impact competition: practices which take advantage of vertical integration, while they have potential to limit competition in some contexts, **can also be markers of strong dynamic competition – for example by helping create efficiencies, lower prices, and diffuse, disseminate and encourage take-up of innovative AI services**. This calls

for a fact-specific, case-by-case analysis rather than a presumption than 'openness' is always the most pro-competitive outcome.

Our findings engage the broader debate among European policy makers about the role of competition policy in high-technology sectors, and in particular the way in which innovation and investment should be taken into account in competition assessments. Letta and Draghi recommend that competition authorities give more consideration to innovation and growth when they enforce competition law.[1] Commission president Ursula von der Leyen has also directed the European Commissioner responsible for competition policy, Teresa Ribera, to "modernise" the bloc's competition policy.[2] In mature markets – where most customers are already served, competition tends to focus on price and quality, and innovations tend to be incremental – an approach to competition policy which pays less attention to factors like price carries some risks.[3] **High-growth, high-potential and disruptive sectors like AI offer a more promising context in which an innovation-focused competition policy can support Europe's economic and technological goals.**

This creates a question about how pre-emptive policy makers ought to be when intervening in the AI sector, particularly given the increasing emphasis on the use of regulation to shape digital markets rather than traditional competition law. In some more mature digital markets, competition authorities have concluded that they waited too long to intervene and that those markets have now tipped irrevocably towards one or two players. That remains a risk in AI because its future trajectory is so uncertain.

Nevertheless, several factors – such as the decreasing costs of developing AI, the growing interdependence of AI models, and the prominence of relatively open AI models – suggest that **competition authorities cannot assume market consolidation is inevitable or that 'winner takes all' outcomes will occur in either the provision of AI foundation models or the provision of computing resources to AI developers, even if there are grounds for intervention elsewhere in the tech stack**. That suggests that policy makers ought to take a balanced approach – considering both the benefits and the risks of intervention – and ensure regulatory intervention is targeted at specific and well-evidenced problems.

[1] See Draghi and Letta, above n 4.
[2] Letter from European Commission president Ursula von der Leyen to the Executive Vice President Designate for a Clean, Just and Competitive Transition, 17 September 2024.
[3] Zach Meyers, 'Competition policy must reflect Europe's reality, not its aspirations', Centre for European Reform, 23 October 2024.

Table of Contents

About CERRE

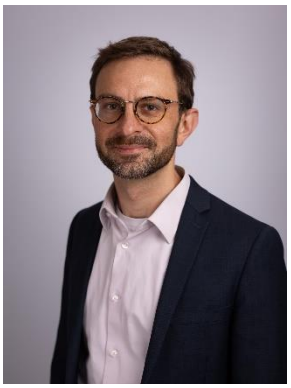
Providing high quality studies and dissemination activities, the Centre on Regulation in Europe (CERRE) is a not-for-profit think tank. It promotes robust and consistent regulation in Europe's network, digital industry, and service sectors. CERRE's members are regulatory authorities and companies operating in these sectors, as well as universities.

CERRE's added value is based on:

- its original, multidisciplinary and cross-sector approach covering a variety of markets, e.g., energy, mobility, sustainability, tech, media, telecom, etc.;
- the widely acknowledged academic credentials and policy experience of its research team and associated staff members;
- its scientific independence and impartiality; and,
- the direct relevance and timeliness of its contributions to the policy and regulatory development process impacting network industry players and the markets for their goods and services.

CERRE's activities include contributions to the development of norms, standards, and policy recommendations related to the regulation of service providers, to the specification of market rules and to improvements in the management of infrastructure in a changing political, economic, technological, and social environment. CERRE's work also aims to clarify the respective roles of market operators, governments, and regulatory authorities, as well as contribute to the enhancement of those organisations' expertise in addressing regulatory issues of relevance to their activities.

About the Authors



Zach Meyers is Director of Research at CERRE, where he has a wide remit, including managing cross-sectoral programmes and projects.

Previously the assistant director of the Centre on European Reform, Zach Meyers has a recognised expertise in economic regulation and network industries such as telecoms, energy, payments, financial services and airports. In addition to advising in the private sector, with more than ten years' experience as a competition and regulatory lawyer, he has consulted to several governments, regulators and multilateral institutions on competition reforms in regulated sectors. He is also a regular contributor to media.

Zach holds a BA, LLB (with First Class Honours) and a Master of Public & International Law from the University of Melbourne.

Marc Bourreau is a CERRE Academic Co-Director and Professor of Economics at Telecom Paris, Institut Polytechnique de Paris, France, where he acts as the Director of the Chair for Innovation & Regulation.

He holds a master's degree in engineering from Telecom Paris and a Ph.D. in Economics from University Paris 2 Panthéon-Assas. He has published in leading journals in Economics. His current research interests concern the economics of digital platforms, the impact of competition and regulation on entry and investment in network industries, and licensing and trading of standard essential patents.

1. Introduction

Two major reports by former Italian prime ministers, Enrico Letta and Mario Draghi, recently painted a dire picture of the bloc's economic performance – blaming much of it on Europe's inability to take advantage of new technologies.[4] **European policy-makers have identified artificial intelligence (AI) as a technology which could help boost Europe's sluggish economic growth**.[5] While the EU has been able to adopt new technologies to boost the efficiency of its manufacturing sector, productivity growth in the services sector has been languid for many years. AI offers an opportunity to change this by automating many tasks in services sectors. Despite the ICT revolution, European services firms have often struggled to use technology to boost their output and efficiency: the US saw a huge productivity boom in services from the ICT revolution of the 1990s but Europe missed out.[6] Since services represent 70% of Europe's economy, AI has significant potential to help Europe's economic growth catch up.[7]

There are two ways in which this technology could boost growth:

- First, as a general-purpose technology – which helps lower the cost of making predictions, a task important to many businesses – **AI can be adopted by companies across many sectors of the economy**. It can help new firms enter established markets, existing firms become more efficient, and transform markets by providing completely new products and services. As competition authorities have acknowledged, widespread use of AI can bolster innovation and competition across the economy – creating stronger competitive pressure that forces all firms to use technology to become more innovative and efficient.
- Second, **AI could offer new opportunities for Europe's technology sector**. For example, AI could allow European tech firms a new foothold to compete using a technology where (unlike in some other parts of the technology sector) market leadership is up for grabs and market structures are far from settled. This seems important given Europe's growing concern about excessive reliance on foreign tech services and growing opposition from the US administration about EU regulation which aims to ensure that foreign services abide by European values when they do business in the bloc.

To maximise the economic benefits of AI and its opportunities to boost European competitiveness, **policy makers will need to encourage and enable the growth of the AI sector and, equally, ensure that firms in the sector face competitive pressure to continue to invest and innovate**. In terms of growth, only about 13.48% of European firms say they are actively using AI.[8] Although that number seems unrealistically low (given that AI is already embedded in much everyday software and that many employees will be using publicly available AI tools on their own initiative), markets for using AI – in particular specialised AI services designed specifically for individual firms – remain relatively nascent, with huge potential to grow. Work to reduce barriers to investment in AI, to encourage take-up, and

[4] Mario Draghi, 'The future of European competitiveness', September 2024; Enrico Letta, 'More than a Market', April 2024.
[5] European Commission, 'AI Continent Action Plan', 9 April 2025.
[6] Robert Gordon and Hassan Sayed, 'Transatlantic technologies: The role of ICT in the evolution of US and European productivity growth', National Bureau of Economic Research, 2020.
[7] Zach Meyers and John Springford, 'How Europe can make the most of AI', Centre for European Reform, 14 September 2023.
[8] Eurostat, 'Use of artificial intelligence in enterprises', January 2025.

to ensure that customers of AI are not locked in and can take advantage of new AI services will be important.

This issue paper provides an overview of how effective competition between providers of AI foundation models is functioning. We consider where relevant upstream inputs, in particular the provision of computing power and data for AI foundation models. This report does not consider in detail downstream uses of AI – such as competition between applications and devices that deploy AI, or future markets that might develop using AI, such as markets for AI agents.

The German, French, Dutch and UK competition authorities have all begun studying competition and the AI sector.[9] The CMA, FTC, DOJ and European Commission issued a Joint Statement on Competition in Generative AI Foundation Models and AI Products, highlighting the competition risks of AI and indicating that they sought cooperation and to share knowledge.[10]

Some authorities' initial worries about competition to develop AI have receded to some degree.[11] As explained below, several potential barriers to entry for developers of AI foundation models or 'FMs' (referred to in this paper as 'AI developers') have proven less significant than authorities initially feared. The result is a thriving ecosystem of diverse FMs. However, the future shape of the AI sector is highly uncertain.[12]

In assessing how the AI sector could develop, **the European Commission and national competition authorities are particularly focused on the multi-faceted relationship between the largest cloud computing platforms (provided by Microsoft, Amazon and Google, collectively commonly called the 'hyperscalers') and other AI developers** such as Mistral, OpenAI, Hugging Face and Anthropic. The relationship between these AI developers and the cloud computing platforms is complex, and there is some concern about the influence that hyperscalers may have over AI providers and over the direction of the AI sector generally and its disruptive potential.

Our findings are focused purely on ensuring competition for AI, regardless of the identity of the market participants. In parallel, policy makers are considering ways to support the presence of EU-based cloud computing providers and AI firms.[13] We note these initiatives where relevant to our analysis below. However, a desire to ensure a place for European AI and cloud computing firms is best addressed through ensuring that competition is effective and disruptive firms have fair commercial opportunities, rather than by adopting an approach to competition policy that assumes foreign firms

[9] Bundeskartellamt and Autorite de la Concurrence, 'Working Paper - Algorithms and Competition', 6 November 2019; Autorite de la Concurrence, 'Generative artificial intelligence: the Autorité issues its opinion on the competitive functioning of the sector', 28 June 2024; Autoriteit Consument & Markt, 'Onderzoek naar toezicht op algoritmische toepassingen', 10 December 2020; Competition and Markets Authority, 'AI Foundation Models: Update paper', April 2024.

[10] European Commission, Competition and Markets Authority, US Department of Justice and US Federal Trade Commission, 'Joint Statement on Competition in Generative AI Foundation Models and AI Products'.

[11] See, eg, Competition and Markets Authority, 'AI Foundation Models: Update paper', April 2024 cf Competition and Markets Authority, 'AI Foundation Models: Initial report', September 2023.

[12] See Cade Metz, Karen Weise and Tripp Mickle, 'A.I. Start-Ups Face a Rough Financial Reality Check', New York Times, 29 April 2024.

[13] For example, at time of writing, the Commission is consulting on a proposed Cloud and AI Development Act: https://digital-strategy.ec.europa.eu/en/consultations/have-your-say-future-cloud-and-ai-policies-eu.

(or those which already have a presence in other digital markets) have a negative influence on competition.

This paper is structured as follows:

- Section 2 sets out the current AI value chain and explains the current levels of competition and relationship between AI developers and cloud computing services.
- Section 3 examines how competitive dynamics may change in future.
- Section 4 draws some tentative conclusions about how authorities should ensure continued innovation and growth in the sector.

We will follow up this issue paper with a report which will build on this analysis, assess the various regulatory initiatives which European authorities have proposed for the AI and cloud sectors, and provide conclusions about how authorities should decide whether to make competition and regulatory interventions and about the shape that any such interventions should take.

2. Competition in the AI Value Chain Today

In this section, we describe how competition works in key parts of the AI 'value chain': starting with inputs into AI models, competition between AI models, and then examining downstream activities. Our conclusion is that, in general, competition appears to be thriving. This means that, in assessing whether there are competition problems that might emerge in future, and the risk of intervening too late to effectively correct these problems, authorities will need to take account of the risks of being overly interventionist – and should be careful to ensure interventions are targeted and informed by evidence.

In assessing competition, we note below that both AI and the provision of accelerated compute are characterised by very high levels of investment by a variety of different players, including some of the largest tech firms and also those many times smaller, in addition to other investors such as large private equity and venture capital firms.[14] Innovation in this sector is occurring rapidly, with high levels of uncertainty about which business models will succeed, and the markets are growing in size with room for many different types of players to increase their customer bases. In markets where most customers are already served, competition tends to focus on price and quality, and innovations tend to be incremental; but that is not the case where firms are innovating and experimenting to unlock latent customer demand.

It is widely accepted that competition authorities have often failed to adequately account for this type of innovation in the past.[15] In these contexts, assessing competition with 'traditional' metrics such as market share would be inadequate, since the 'markets' themselves are still being determined, and since players in the AI sector are competing to position themselves against potential future competitors, and face uncertainty about which AI businesses will succeed and whether the sector will create new focal points for competition. Below, we draw from the emerging body of work on assessing 'dynamic competition'[16] – which emphasise that consumers benefit more when firms are incentivised to innovate, particularly in radical ways, rather than when competition only drives decreases in price or marginal increases in quality.

2.1 Inputs for AI developers

AI developers generally need access to a number of inputs to develop foundation models, such as access to datasets, skilled talent, and computing power. In the past, competition authorities have

[14] https://www.ey.com/en_ie/newsroom/2024/12/venture-capital-investment-in-generative-ai-almost-doubles-globally-in-2024-as-momentum-accelerates-in-transformative-sector; https://www.spglobal.com/market-intelligence/en/news-insights/articles/2024/3/private-equity-backed-investment-surge-in-generative-ai-defies-2023-deal-slump-80625128.

[15] Wolfgang Kerber & Simonetta Vezzoso, 'Competition and innovation: Incorporating a more dynamic perspective into enforcement', 3 January 2025.

[16] See, eg, David Teece, 'Understanding Dynamic Competition: New Perspectives On Potential Competition, "Monopoly," And Market Power', 2024.

raised concerns that these were potential 'bottlenecks' that could constrain competition in AI.[17] In practice, however, many of these inputs have turned out to be widely accessible.

Datasets, for example, were feared to be a particular bottleneck, particularly since several large technology firms have huge private datasets which could be used to train FMs. In **practice, however, many FMs are being trained on depositories of public data (such as those provided by the free Common Crawl, Institutional Data Initiative or The Pile archives) some of which are available to anyone**.[18] Policy makers are increasingly concerned by a different question – namely, tackling what rights AI developers should have to train models on data covered by intellectual property ('IP') laws without the consent of the rights holders.[19] In addressing this issue, policy makers must be conscious of the potential competitive impacts of constraining AI developers, since larger AI developers will be more likely to have the resources to negotiate access to content protected by IP compared to smaller firms.

To build FMs, AI developers have begun shifting away from training models on ever-more data,[20] and have instead started to rely more on alternative ways to improve their models.[21] These include:

- Relying more heavily on specific or highly curated datasets necessary to "fine-tune" AI models to work in particular use cases and/or for specific industries. According to a Gartner prediction, by 2027, more than 50% of generative AI models will be specific to either an industry or business function, up from approximately 1% in 2023.[22] In these cases, the data essential to fine-tune the model will depend on the intended use case – for example, a business customer wanting to use an AI model to optimise its business practices may want to fine-tune the model on the business's own data. In such contexts, data will not necessarily pose a bottleneck.
- FM developers are relying more heavily on alternatives to more data, for example by instead improving the quality and structure of that data, so that AI models can identify the chain-of-thought that links a particular request or question to an answer, and can replicate that chain-of-thought to produce its own answers.[23] Such data can often be produced through manual categorisation of data or from other AI models (with the effect of increasing AI developers' reliance on the continued openness of other AI models – a point we explore further below).

For some use cases, AI developers will nevertheless require access to specific datasets for which no alternative is available. **There remain concerns about some of these datasets being withheld (or controlled) in ways that might limit competition for using AI in particular use cases – in particular**

[17] Competition and Markets Authority, 'AI Foundation Models: Initial report', September 2023.
[18] Geoffrey Manne and Dirk Auer, 'From Data Myths to Data Reality: What Generative AI Can Tell Us About Competition Policy (And Vice Versa)', CPI Antitrust Chronicle, February 2024.
[19] Ana Rački Marinković, 'Liability for AI-related IP infringements in the European Union', Journal of Intellectual Property Law & Practice, 19(10), October 2024.
[20] Bertin Martens, 'How DeepSeek has changed artificial intelligence and what it means for Europe', Bruegel, Policy Brief 12/25, March 2025.
[21] The use of synthetic data produced by AI has also been mooted as a way to avoid data bottlenecks, but in practice this approach has not proved as promising as hoped, due to concerns about the quality of synthetic data and its close association with the initial data on which it was based.
[22] See Gartner, *3 Bold and Actionable Predictions for the Future of GenAI* (Apr. 12, 2024): https://www.gartner.com/en/articles/3-bold-and-actionable-predictions-for-the-future-of-genai
[23] Maarten Grootendorst, 'A Visual Guide to Reasoning LLMs', available at https://newsletter.maartengrootendorst.com/p/a-visual-guide-to-reasoning-llms.

where they compete with or pose a commercial challenge to the data holder.[24] For example, many AI services use a technique called 'reinforcement learning from human feedback' or 'RLHF', which relies on gathering usage data about how users interact with a service, including how satisfied they are with the proposed outputs. RLHF may help the most popular services increase in quality, creating a 'winner takes all' dynamic. Further research on these dynamics would be helpful, as at the moment the importance of such 'feedback loops' is not very certain, but their impact on competitive dynamics could prove very significant.[25]

In other cases, information about a particular customer's usage of a service can help the service provider build up a 'profile' to help deliver the most useful and relevant outputs to that user – making it harder for a user to switch services without suffering a loss in service quality. In principle this may justify measures to help ensure users can 'port' their usage history from one AI service to another, to help minimise the extent of such 'lock in'.

AI services may also rely on proprietary datasets which only a single third party owns or controls. In those cases, there may be a case for targeted interventions to preclude a potential competitor from entering a market, or from exerting competitive pressure which would not otherwise exist.

However, in each of these cases, it is important that regulatory interventions do not undermine incentives for investment and innovation in the market – for example by unfairly depriving firms of the opportunity to exploit data they have collected through their own investments or through undermining incentives to collect data in the first place. That may require regulatory interventions to be targeted only at genuine 'bottlenecks' and to set terms of access that allow a fair return on the data holder's investments and risk-taking.

Access to AI talent – such as skilled AI coders – was also thought to be a particular problem: many commentators worried that the deep pockets of large tech firms meant they could monopolise access to the skilled workers necessary to build AI models. However, while some hyperscalers have engaged in so-called 'acquihires' – recruiting important employees or whole teams from other AI developers, such as Microsoft's hiring of key staff from AI firm Inflection[26] – there is not much evidence that these have had anti-competitive effects given the large number of successful AI developers which remain independent.[27] In practice, **many skilled developers still choose to work in smaller and more agile start-ups. The emergence of more open AI models (some of which are open source) has also significantly lowered the amount of skilled talent which AI developers need access to**: there is an increasing ability of AI models to learn from and interact with each other, reducing the need for both intensive datasets and for each model to have large numbers of skilled experts helping with its development.

[24] For example, Google limited access to its search data and Microsoft reportedly threatened to cut access to its search index to customers using the data to build AI tools: see Leah Nylen and Dina Bass, 'Microsoft threatens data restrictions in rival AI search', Bloomberg, 25 March 2023. See also Autorite de la Concurrence, 'Opinion 24-A-05 on the competitive functioning of the generative artificial intelligence sector', 28 June 2024, para 261.

[25] Klaus Kowalski, Cristina Volpin, and Zsolt Zombori, 'Competition in Generative AI and Virtual Worlds', Competition Policy Brief, September 2024

[26] This case was cleared by the Competition and Markets Authority: see *Microsoft/Inflection* decision, 4 September 2024.

[27] Contra Autorite de la Concurrence, 'Opinion 24-A-05 on the competitive functioning of the generative artificial intelligence sector', 28 June 2024, p 8.

2.2 Accelerated Compute

AI developers' access to compute remains one of the most significant inputs over which competition authorities have had concerns. AI developers generally require access to computing power which can perform many different operations in parallel (known as 'accelerated compute'); this functionality can only be provided by specialised 'accelerator' processing chips, for which there is currently enormous global demand. AI developers require accelerated compute for several purposes:

- to train their models, which requires processing vast amounts of data;
- for intermediary steps like fine-tuning a model with more curated datasets, so it can be adapted to deliver better results at particular types of tasks or uses. The datasets may be curated manually, using AI, or with a combination of the two; and
- for responding to an input provided to the model by a user (known as 'inference').

AI developers have a number of options to access compute including:

- **Self-provision**. For example, an AI developer may acquire its own accelerator chips and provide its own dedicated compute; or it might outsource that function to a third-party to operate on-premises dedicated computing equipment. At present, only a small number of AI developers (such as Samsung and Meta) appear to have the scale and resources needed for this option to be economically viable.[28] However, industry sources indicate that large enterprises are expected to increasingly rely on their own infrastructure, particularly for inference.[29]
- **Using a range of general cloud computing providers, including but not limited to the 'hyperscalers'**. These include AWS, Microsoft, Google, Oracle, IBM, Alibaba Cloud, OVHCloud and Scaleway, all of which supply AI developers with accelerated compute via their cloud platforms (along with standard compute). In particular, in January 2025, OpenAI announced a partnership with Oracle called the "Stargate Project", intending to invest $500 billion over four years in AI infrastructure across the US – illustrating that, among general cloud computing providers, it is not just the hyperscalers who can provide accelerated compute.[30]
- **Smaller cloud computing providers which specialise in providing accelerated compute to AI developers**, such as Lambda Labs, Denvr, TensorWave, CoreWeave, Vultr, Nebius, GenesisCloud and San Francisco Compute. Sophisticated AI developers may prefer to use these smaller and more specialised accelerated compute providers – since these can offer better value for money, give AI developers more control, and help the AI developer avoid paying a general cloud computing provider to manage their compute workloads where the AI

[28] Autorite de la Concurrence, 'Opinion 24-A-05 on the competitive functioning of the generative artificial intelligence sector', 28 June 2024, p 4.

[29] https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://go.techinsights.com/l/1043171/2024-10-23/8lvv8x&ved=2ahUKEwjG5IiA2fGMAxUOSPEDHbcJCcYQFnoECBcQAQ&usg=AOvVaw022juzX3HqcXMsYl8hxBpe

[30] OpenAI, 'Announcing The Stargate Project', 21 January 2025, available at https://openai.com/index/announcing-the-stargate-project/.

developer can do that itself. For example, even hyperscalers like Microsoft have relied on specialised providers like CoreWeave.[31] In addition to general cloud computing providers, for compute, AI developers can also choose from traditional hardware makers such as HP, Dell and Nvidia.

- **Colocated facilities**. Colocation can offer a middle ground for AI developers, between the efficiencies of using general purpose cloud computing facilities and the greater control offered by operating (or outsourcing) dedicated computing facilities. Under this model, computing power is outsourced to another provider's data centre, which may serve a number of different customers. This model combines benefits of control with some of the scalability and efficiency of using cloud computing providers.
- **Publicly funded supercomputers or those provided by educational institutions**. For example, the European Commission has announced a €200 billion plan to invest in AI, including for 'AI gigafactories' which would provide compute to European AI developers,[32] and French president Emmanual Macron announced €109 billion in private investments. A number of research institutions in Europe also host supercomputers which may be capable of providing accelerated compute to AI developers, often for free in return for contributing to published scientific research.[33] However, access to public supercomputers is often for a time-limited period so it can be useful in training models but not for performing inference.[34] To be successful at improving choices for AI developers, announced funding for future AI gigafactories will need to provide ongoing access to scalable accelerated compute for large FMs.

AI developers will choose between these options by evaluating their needs against the different commercial and technical options available to them. They may also choose between different solutions and providers. They may also 'mix and match' depending on the needs of particular computing workloads.

Cloud computing may have advantages for AI developers over self-provision, because it allows the high up-front cost of accelerator chips to be shared across multiple AI developers. This can be particularly economical for training models, since AI developers do not need everyday access to compute for training.[35] **The growth of cloud computing for AI developers has therefore had a significant positive impact on competition in AI**, lowering barriers to entry for AI developers by allowing AI firms to rent computing power, and increasing their access to such power on an 'as needs' basis, rather than forcing them to make large up-front investments in computing power.

[31] See, eg, CoreWeave S-1, 3 March 2025, available at https://www.sec.gov/Archives/edgar/data/1769628/000119312525044231/d899798ds1.htm?ref=runtime.news.

[32] European Commission, 'EU launches InvestAI initiative to mobilise €200 billion of investment in artificial intelligence', press release, 11 February 2025.

[33] Autorite de la Concurrence, 'Opinion 24-A-05 on the competitive functioning of the generative artificial intelligence sector', 28 June 2024, p 5.

[34] Autorite de la Concurrence, 'Opinion 24-A-05 on the competitive functioning of the generative artificial intelligence sector', 28 June 2024, para 172.

[35] Brian Albrecht & Geoffrey Manne, 'ICLE Response to the FTC's Cloud Computing RFI', 21 June 2023.

A static analysis suggest that the sector is relatively concentrated and potentially becoming more concentrated. This is of potential concern. The three largest cloud computing providers account for about 75% of the overall public cloud sector, but account for almost all new public cloud AI projects.[36]

However, a more dynamic analysis suggests that the largest three firms might be under more competitive pressure than market share figures would assume. Microsoft and Google appear to be performing relatively well in providing cloud computing services for AI projects, for example, suggesting that they are imposing significant pressure on the overall public cloud computing leader, Amazon Web Services. This – together with the extraordinarily high level of investments being made by the hyperscalers, with cloud computing providers investing about $250 billion in global investment for AI in 2025 alone[37] – suggests, that **competition between the largest cloud computing companies is intensifying**, with Microsoft and Google's smaller share of overall cloud computing not necessarily holding them back from winning a significant share of AI projects.

Several other factors also suggest that other players have some potential to grow and may be exerting growing competitive constraint on the hyperscalers:

- Some smaller cloud providers focus on helping AI developers conduct fine-tuning and inference, and have been unable to provide training due to the high costs and investments involved. In recent months – and especially as AI models have started to train on each other's outputs, which partly allowed the Chinese model DeepSeek to produce a high performance LLM at low cost – the costs of AI development have shifted away from initial training and towards fine-tuning and inference.[38] For example, models are now produced by being trained on higher-quality data (often curated by AI), rather than 'more' data, and the models are instead using more compute power to assess multiple potential chains-of-thought to produce the best possible result in response to a user's query.[39] **This development has the potential to significantly increase the ability of smaller cloud providers to compete with the hyperscalers for AI developer customers. It should also increase the viability of many AI developers increasing the amount of processing they undertake 'on premises' or using collocated data centres, rather than in the cloud.**
- Second, as Table 1 below shows, **when AI developers do use cloud providers, they will often choose multiple providers**. Where an AI developer has chosen only one of the accelerated compute providers, it is not necessarily one of the hyperscalers. This illustrates that many AI developers are keeping their options open and are trying to avoid 'lock in' to any one cloud provider. This should impose competitive tension on cloud providers, at least so long as customers have the option of shifting at least some of their compute demand between providers (an issue we address below).

[36] Source: https://iot-analytics.com/who-is-winning-the-cloud-ai-race/

[37] Brooke Dane and Ty York, 'Technology in 2025: The Cycle Rolls On', Goldman Sachs, 3 February 2025, available at https://am.gs.com/en-gb/institutions/insights/article/2025/technology-in-2025-the-cycle-rolls-on.

[38] Tim Bradshaw, 'How 'inference' is driving competition to Nvidia's AI chip dominance', Financial Times, 11 March 2025.

[39] The ability to use data produced from another AI model (the "teacher" model) can significantly reduce the costs of producing another model (the "student" model).

- Third, often an AI developers' choice of accelerated compute provider is influenced by the existence of a formal partnership with that provider (described in more detail below). These often involve the cloud computing provider taking a financial stake in the AI developer in return for concessions about the AI developer's access to accelerated compute, along with an agreement to host the AI model on the hyperscaler's platforms, or to integrate the AI developer's services into the hyperscaler's existing services, and sometimes facilitate the accelerated compute provider obtaining data, IP or information about the AI developer's business model.[40] These partnerships differ greatly, and are not always transparent, but in at least some of these deals the AI developer has agreed to use one of the large cloud computing providers as their 'primary' or 'preferred' provider (and to have their model hosted on the hyperscalers' platforms – see section 2.5 below). While authorities have scrutinised these partnerships, they have often – and particularly now that the number of AI developers continues to grow – not found these partnerships to pose competitive problems that justified intervention.[41] Currently, the commercial incentive for these partnerships appears to be because the hyperscalers are themselves unsure about the future shape of the market, or which AI providers will succeed long term, and therefore fear 'losing out' if their own AI efforts fail; whereas AI providers benefit from more certain access to accelerated compute, in some cases tailored to the needs of the particular AI developer. Furthermore, few partnerships seem to involve any exclusive arrangements;[42] and even supposedly exclusive deals do not preclude an AI developer exploring other options: despite having a much-publicised quasi-exclusive partnership with Microsoft, OpenAI recently negotiated a new partnership with Microsoft's cloud rival Oracle and agreed to a $11.9 billion deal with CoreWeave, a smaller and more specialised provider of accelerated compute.[43] **This suggests that many hyperscaler/AI partnerships are driven by hyperscalers' uncertainty and that independent AI developers can often benefit from these partnerships while retaining significant commercial freedom, including to continue to build applications which compete with the hyperscalers' own services.**

[40] Federal Trade Commission, 'Partnerships Between Cloud Service Providers and AI Developers FTC Staff Report on AI Partnerships & Investments 6(b) Study', January 2025, section 4.4.

[41] The Microsoft/OpenAI partnership was examined by the European Commission, which concluded the partnership was not a deal which qualified for review under the EU's merger control regime.; The CMA launched investigations into the Amazon/Anthropic and Google/Anthropic partnerships. The Google/Anthropic deal did not meet the criteria for merger review.; The CMA also investigated Microsoft's partnership with Inflection AI, but found that it did not substantially reduce competition, and Amazon's partnership with Anthropic.

[42] See Cristophe Carugati, 'Competition and cooperation in AI: How co-opetition makes AI available to all', Digital Competition Working Paper 3/2024, 11 March 2024.

[43] https://www.nextplatform.com/2025/03/11/what-a-tangled-openai-web-we-coreweave/

Table 1. Use of Cloud Computing Companies by AI Developers[44]

	AWS	Microsoft	Google	Oracle	CoreWeave
AI21 Labs	Yes		Yes		
Adept AI		Yes		Yes	
Anthropic	Yes		Yes		
Character AI			Yes	Yes	
Cohere			Yes	Yes	
EleutherAI					Yes
Meta	Yes	Yes			
Midjourney			Yes		
Mistral		Yes			
Mosaic ML				Yes	
OpenAI		Yes		Yes	Yes
Runway ML	Yes		Yes		
Stability AI	Yes				

- **Some AI providers are deliberately choosing to avoid hyperscalers.** They may either prefer cloud providers such as Oracle, which do not compete in the provision of AI FMs.[45] Others may prefer options like CoreWeave because, unlike the hyperscalers, these cloud computing firms have optimised their resources specifically for AI training, fine-tuning and inference (for example with faster connections between accelerator chips and hardware designed specifically for AI-related processing). Customers in sectors like financial services may rely more on on-premises or private cloud services where they can maintain more direct control over the data used with their AI models.
- Several big tech players – such as Apple and Meta – do not have large-scale cloud computing platforms of their own. **These large and influential players have the resources and incentives to avoid the provision of accelerated compute becoming too concentrated.** Meta, for example, is focused on developing AI models which are more open than many alternatives and can operate across different cloud environments. Apple is working on AI models which can – to a large extent – run inference on a user's own device, limiting the need to rely on cloud computing.[46]
- **There is growing pressure in Europe to support and encourage more "sovereign" providers of cloud computing**, and to enable European firms to have more of a role in the AI value chain. For example, the Commission's AI Continent Action Plan proposes a range of investments to boost Europe's AI industry, including AI factories to train and finetune AI models. The Commission has also mooted an EU Cloud and AI Development Act, which might set minimum standards for cloud computing services in Europe, potentially providing advantages to local

[44] Source: CMA; authors' updates (OpenAI). Table 1 may understate the degree of freedom AI developers have, since it does not include the use of alternatives such as public supercomputers or on-premises IT (used by some large firms like Meta).
[45] Jai Vipra and Sarah Myers West, 'Computational Power and AI', AI Now Institute, 2023.
[46] See, eg, Apple, 'Introducing Apple's On-Device and Server Foundation Models', 10 June 2024.

European firms.[47] Regardless of whether or not initiatives to support this ambition are necessary, such initiatives may well support the entry and growth of new local players in the sector.

- While some competition authorities have found that hyperscalers can negotiate preferential agreements to access accelerated chips,[48] there is also some evidence that **NVIDIA – which as explained below is the most significant provider of accelerated AI chips today – has an active strategy of supporting alternatives to the hyperscalers**, for example by giving smaller cloud computing services early access to NVIDIA's latest accelerator chips and financially investing in alternative providers of accelerated compute such as CoreWeave. This business strategy ensures that NVIDIA maintains a diverse customer base, but also helps to ensure that AI providers have a good range of access to different external providers of accelerated compute. Notably, Nvidia appears to be deepening partnerships with some large 'challenger' cloud providers while not partnering to the same extent with AWS. [49]

Many of these competitive threats are nascent. However, the mere threat of competitive entry can impose important discipline on firms. Competition authorities will, however, need to keep a close eye on the sector to see how these opportunities and dynamics play out. Authorities will also need to examine the sector as the shape of economic markets becomes clearer. For example, even if – as seems likely – more cloud computing providers will be able to help AI developers in operating AI models, that might not necessarily increase their choices for training AI models. Furthermore, a number of concerns about hyperscalers' business practices in their provision of accelerated compute deserve consideration.

First, many cloud computing providers offer 'credits' to AI developers to attract them as customers. These practices are also prevalent for AI developers,[50] including in AI developer/hyperscaler partnerships.[51] Discounts are generally pro-competitive: they can illustrate strong competition between cloud computing companies to win customers, and can boost competition in the AI sector by lowering AI developers' costs. **Anti-competitive effects from discounts seem unlikely in a market which is highly competitive.** In general, since offering discounts simply requires a willingness to accept lower returns for a period, competing cloud computing companies should be able to offer their own discounts – Oracle, a large cloud computing provider which is still significantly smaller than the hyperscalers, offers significant credits to AI providers,[52] for example – and 'free trials' or similar schemes are common in many highly competitive parts of the economy. Competition authorities may need to examine discounts if they have problematic characteristics, such as (contractually or in practice) imposing a high degree of exclusivity and/or limiting AI developers' choices, and are

[47] European Commission, 'A Competitiveness Compass for the EU', 29 January 2025.
[48] Autorite de la Concurrence, 'Opinion 24-A-05 on the competitive functioning of the generative artificial intelligence sector', 28 June 2024, p 6.
[49] Nvidia reportedly gave Azure, GCP, and smaller players such as Coreweave, Equinix, and Lambda earlier access to its H100s than AWS; one reason may be that Amazon refuses to use Nvidia's product NVLink: Lisa Sparks, 'Will Nvidia's AI dominance shake up the public cloud 'big three'?', ITPro, 26 October 2023.
[50] Autorite de la Concurrence, 'Opinion 24-A-05 on the competitive functioning of the generative artificial intelligence sector', 28 June 2024, fns 146-147.
[51] Federal Trade Commission, 'Partnerships Between Cloud Service Providers and AI Developers FTC Staff Report on AI Partnerships & Investments 6(b) Study', January 2025, p 3.
[52] Aaron Holmes, 'AI Startups Find an Unlikely Friend: Oracle', The Information, 22 February 2023, https://theinformation.com/articles/ai-startups-find-an-unlikely-friend-oracle.

therefore likely to be a way to attract developers and then lock them in. For example, if discounts are conditional on an AI developer achieving very high levels of usage of a service, then they may have the effect of preventing that developer using any other provider of accelerated compute, or being forced to repay the discount unless they recommit to the same provider. The commercial justification for high minimum spends may need to be carefully monitored.

Barriers to switching have also raised concerns. There are a variety of potential commercial and technical barriers to AI developers switching between accelerated compute platforms. Commercial barriers may be one problem. As noted above, discounts, particularly when coupled with certain characteristics such as high minimum spend commitments, may result in barriers to switching. Authorities have been investigating the impact of commercial practices like charging egress fees for moving data in connection with switching providers, which some smaller firms argue can pose an unreasonable barrier to switching.[53] **While some cloud computing providers argue there is a commercial rationale for egress fees connected with switching, other providers do not charge such fees. In any event, the EU's Data Act requires cloud computing providers to remove switching and egress fees related to switching, and to limit other egress fees,** by 12 January 2027.[54]

Technical barriers pose greater challenges because they are often associated with new features and service differentiation. These features and differences can benefit customers even if they consequently make switching providers more challenging. For example, accelerated compute providers provide specialised and proprietary tools and software to train and fine-tune models, including Machine Learning Operations services ('MLOps') such as Google Vertex AI, Azure Machine Learning, and Amazon SageMaker, all of which streamline the AI lifecycle. Switching accelerated compute provider could result in significant costs and difficulties as employees learn how to use a different set of tools. However, it is difficult to conclude that the development of these tools is inherently anti-competitive. In the general cloud sector, regulatory interventions to enable easier switching and interoperability have been difficult to implement successfully because cloud computing operators' service offerings are not identical and services are not standardised. This poses a dilemma between the competitive benefits of enabling easier switching, on the one hand, with the potentially negative effects on innovation of imposing greater standardisation between competing services.

As noted above, an important issue for switching is the ability of AI developers to train a model in one cloud environment and then deploy it in another when these tools are used. This is because, as explained above, the barriers to providing accelerated compute for the purposes of inference (i.e. operating a model) may be significantly less than for training a model. An AI developer may have significantly more choice for the former than the latter. The three hyperscalers do tend to allow integration of third-party solutions, including open-source ones, such as TensorFlow and PyTorch. This can allow AI developers to fine-tune open-source models in one cloud environment and then deploy the model using a different cloud provider.[55] However, when proprietary tools are used to fine-tune a model in one cloud provider's environment, AI developers may not always have full access to the fine-

[53] Competition and Markets Authority, 'Cloud Services Market Investigation: Summary of provisional decision, 28 January 2025.
[54] Regulation 2023/2854 (Data Act) art 29.
[55] Felix Theisinger, 'Multi-Cloud: Minimizing lock-in risks for AI and Generative AI', Cognizant, December 2024, https://www.cognizant.com/de/de/insights/blog/articles/multi-cloud-ai-lock-in-risks.

tuned model, but only access to its outputs.[56] This risks effectively concentrating the market for accelerated compute only to those firms which are able to provide sufficient computing power to train a model. Whether that poses a serious and sustainable barrier for smaller cloud computing providers will depend on how significantly future training costs for AI models decrease.

2.3 Inputs into the Provision of Accelerated Compute

Given the growth in the sector, a large potential barrier to thriving competition in AI might be the capacity for the accelerated compute sector to grow. Accelerated compute requires its own inputs – such as large-scale data centres and access to accelerator chips. Some of these seem likely to be potential constraints on the development of a competitive market for accelerated compute:

- **In terms of accelerator chips, a potential concern is NVIDIA's strong position in the provision of accelerated AI chips, and in particular the increasing price for cutting-edge AI accelerator chips.[57]** Competition authorities in France, Europe, the United States and China are currently investigating Nvidia's business practices.[58] However, there are emerging signs that its lead might be becoming eroded. For example, AMD and Intel have both announced new chipsets dedicated to AI.[59] Cloud computing and AI companies are now increasingly designing their own custom chips optimised for their specific AI models or computing workloads. As examples, Google is designing processing chips which are used by the AI developers AI21, Anthropic and Midjourney. Amazon has similarly designed its own Trainium and Inferentia accelerated chips, and Microsoft has announced a Maia 100 chip. Google's and Amazon's specialist chips have been used by leading AI providers Cohere and Anthropic.[60] OpenAI is also now designing its own chips.[61] A number of smaller firms such as SambaNova, Cerebras and Groq have also designed AI chips which have some advantages – such as on cost and power consumption – over NVIDIA's.[62] Large tech firms are also providing or supporting software development kits like Amazon's Neuron, which can support switching between Nvidia and third-party chips. Furthermore, the recent shift in emphasis away from compute-intensive initial training of models towards spending more compute resources on the inference stage has opened up new opportunities for firms to design chips which could be better than

[56] Autorite de la Concurrence, 'Opinion 24-A-05 on the competitive functioning of the generative artificial intelligence sector', 28 June 2024, para 251.
[57] NVIDIA's H100 accelerator chip can cost $40,000 whereas a chip from two generations before (V100) cost only $10,000: TechInsights, AI Outlook Report 2025.
[58] See, eg, Meaghan Tobin et al, 'China Opens Investigation Into Nvidia Over Potential Antitrust Violations', New York Times, 9 December 2024.
[59] Ian King, 'Why Nvidia is the King of AI Chips, and Can It Last?', Bloomberg, 25 February 2025.
[60] Cohere, 'Cohere and Google Cloud Announce Multi-Year Technology Partnership', press release, 17 November 2021.
[61] Anna Tong, Max Cherney and Krystal Hu, 'OpenAI set to finalize first custom chip design this year', Reuters, 10 February 2025.
[62] TechInsights, AI Hardware Summit 2024.

NVIDIA's at inference.[63] Barclays predicts Nvidia will supply almost all chips used for frontier AI training, but only about half the market for chips used for inference.[64]

- Provision of accelerated compute requires sufficient access to data centres to host the computing equipment. However, the planning process for building new data centres, and the need for electricity grid investments, often means the infrastructure to support provision of more accelerated compute is lacking. In 2024, about $465 billion was spent on data centre investments and that figure is expected to grow significantly.[65] In practice, **regulatory or other barriers to constructing data centres could have the effect of both hindering the growth of AI and also ossifying the existing market structure, by advantaging firms which have already made investments in data centres over new entrants**.
- CUDA, the proprietary coding language owned by NVIDIA, is the industry standard for running AI software on NVIDIA's accelerator chips and is the only coding language that is fully compatible with those chips. **This may give NVIDIA the ability to influence the development of AI.** However, companies like DeepSeek appear to have developed efficient and high-performance models by directly engaging with the machine code rather than relying on options provided by CUDA. There are also open-source languages available, including the Triton language developed by OpenAI, and OpenCL, SYCL and OneAPI, which aim to allow developers to write code once and then deploy it seamlessly across different vendors' chips. These tools to date have not achieved their full potential, in part because they have been implemented in different ways by different chip providers.[66] Tools like TensorFlow have also been developed which are chip-agnostic and avoid AI developers needing to engage with CUDA directly.

2.4 Competition Among AI Developers

The AI foundation model sector itself appears to be thriving, with thousands of AI models on the market today, many of which have only emerged in recent months.[67] These are provided by both large technology firms which have existing businesses and many independent AI developers, many of which are attracting significant investment. AI developers have a number of choices to reach consumers, including developing their own FM from scratch to maximise their ability to customise and control their model; fine-tuning third-party FMs; or using a third-party FM through an application programming interface or a plug-in or extension.

Several metrics indicate that the sector as a whole enjoys high levels of business dynamism, with incoming investment which is spread across many different players, and a fast pace of innovation:

- **Thousands of different foundational models exist**. Many models offered by AI developer start-ups have an edge on models from the largest firms in certain respects: when start-up

[63] https://www.ft.com/content/d5c638ad-8d34-4884-a08c-a551588a9a28
[64] https://www.ft.com/content/d5c638ad-8d34-4884-a08c-a551588a9a28
[65] The Economist, 'The data-centre investment spree shows no signs of stopping', 5 February 2025.
[66] CortextFlow, 'Making AI Compute Accessible to All', Medium, 30 March 2025.
[67] Competition and Markets Authority, 'AI Foundation Models: Technical update report', 16 April 2024.

Anthropic launched its Claude 3 FM in March 2024, for example, it could beat Google's Gemini FM on both undergraduate-level expert knowledge and graduate-level expert reasoning.[68]

- **AI providers are continuously leapfrogging each other in innovation and performance**.[69] Many AI providers are also innovating by exploring different means of differentiation, for example by increasing their performance at certain specific tasks. This suggests that AI providers are still exploring how to fulfil different customers' needs. Innovations do not only come from the largest and most established or highest-performing models: DeepSeek demonstrates that quite radical innovations can come from unknown new entrants.
- **Prices for using AI models are low (or non-existent) compared to the costs of providing such models** – in fact AI providers generally appear to be operating their services at a loss. That is particularly significant now that the marginal costs of performing inference to answer queries appears to be increasing.[70] This can indicate high levels of dynamism because companies tend to use below-cost or zero pricing (sometimes supported by cross-subsidising profits from other services) in order to 'create' a market and encourage customers to try out an innovative service.
- **A spectrum of business models exist** – at one end, closed-source proprietary models, some of which are only available to selected customers, often with usage limitations; more open models such as Meta's Llama model, where information such as the weights and design of the model are public and the model can be widely used, but the full training model and code is not always available; and on the other end of the spectrum fully open models (like Hugging Face's Bloom model) which can be freely used, adapted, and deployed to provide training to new models, and where the source code and training data are transparently disclosed. In addition, Hugging Face provides a platform that provides access to many open-source models, as well as information about how the models were designed and trained. Even among the largest AI developers, including the hyperscalers, many of their smaller AI models are published openly with only the most advanced models being fully closed source.[71]

This illustrates that there are high levels of dynamism, with firms fighting to identify and fulfil latent customer demand for services and to create and establish the boundaries of economic markets. This implies high levels of experimenting and risk-taking.

One notable development in the sector is that many smaller AI models now rely heavily on larger models: for example to produce high-quality training data or "teach" a smaller AI model chains of reasoning, or to specialise in fine-tuning larger model, or designing value-added services to work on top of existing models.[72] The ability for AI developers to piggy-back off larger models appears to explain how some smaller models, including China's DeepSeek model, have been able to achieve very

[68] Angela Yang, 'Move over, ChatGPT: AI startup Anthropic unveils new models that challenge Big Tech', NBC, 4 March 2024.
[69] Competition and Markets Authority, 'Cloud Infrastructure Services: Provisional decision report', 28 January 2025, para 3.479.
[70] Bertin Martins, 'How DeepSeek has changed artificial intelligence and what it means for Europe', Bruegel policy brief, 12/25, March 2025.
[71] Autorite de la Concurrence, 'Opinion 24-A-05 on the competitive functioning of the generative artificial intelligence sector', 28 June 2024, paras 183-4.
[72] Autorite de la Concurrence, 'Opinion 24-A-05 on the competitive functioning of the generative artificial intelligence sector', 28 June 2024, p 43.

high performance outcomes with very low costs. In turn, other models have now adopted many of the techniques used by DeepSeek. While this development has lowered potential barriers to entry, it also suggests a growing degree of interdependence between different AI models and services. As the sector is evolving, it is too early to tell whether this will eventually result in:

- a small number of "master" foundation models which will become essential inputs to many smaller or fine-tuned models and value-added services provided by independent AI developers; or
- a more general level of interdependency and feedback between different models and services provided by AI developers generally.

In part, the current market structure appears to have arisen because the largest AI developers do not seem able to prevent their models being used to develop other, competing models (and/or do not currently have incentives to do so, perhaps in order to maximise their influence and importance in the AI supply chain). We explore below whether this may change over time as patient investors in AI start to demand a return on investment: while some models are open today, the model developers could introduce technical or licensing use restrictions to prevent their use for creating new AI models. If this occurs, it may pose challenges to some AI providers, potentially reshaping the market. However, if AI providers have incentives to keep their models open (or are unable to effectively restrict how their models are used) then openness may prove to be an enduring characteristic of the sector.

This risk of large FMs becoming more closed over time appears to be encouraging many AI providers to rely more on genuinely open-source AI models. In particular, much of the AI sector is characterised by remarkably high levels of public sharing of resources such as know-how, data repositories, and coding, many shared through open-source facilities such as those of HuggingFace. The existence of tools like HuggingFace provides a degree of assurance that the fate of many smaller AI developers need not rely solely on a few of the largest AI developers including the hyperscalers. It will be important, in this context, that AI regulation does not impose significant burdens on small and particularly open-source AI providers.

2.5 Channels to Market

In addition to providing accelerated compute providers as an input to AI developers, **some of the accelerate compute providers have roles in the downstream value chain** – by operating platforms by which AI developers can reach business users via the use of APIs, and by being major customers of AI developers. For example:

- Google's cloud computing platform offers Model Garden, which hosts over 130 FMs;
- Amazon Bedrock allows developers to access numerous FMs from providers such as Meta, Anthropic, AI21, Cohere, and Stability AI; and
- Microsoft Azure AI Model Catalogue hosts over 1,700 FMs for business customers.

The development of such 'platforms' from the hyperscalers is positive for competition and innovation, since they enable widespread and affordable accessibility of different foundation models to businesses. More open models like Meta's Llama are available across AWS, Amazon and

Microsoft's platforms, and Llama has achieved 1.2 billion downloads.[73] In addition to these proprietary platforms, others like HuggingFace offer an ecosystem of open-source models and tools which firms can access and run either on their own hardware or on any cloud computing provider.

One concern, however, is that the operators of some of these platforms, such as Amazon which operates the Bedrock platform, have also developed their own FMs. This has caused some competition authorities to question whether these platforms may either:

- **deploy an 'open first, closed later' strategy** – where they develop an open platform with many foundation models to attract as many users as possible to that platform, and then later actively drive customers towards their own foundation models, in ways that squeeze out foundation models of third parties; or
- actively or otherwise, **influence the direction of AI development in ways that complement rather than challenge the hyperscalers**.

There remains an open question about whether these platforms may in the future have incentives to become closed: for example, when the market is more mature and the race for new customers ends, and firms may choose to focus on keeping existing users locked in rather than being attractive for new users. However, there does not currently appear to be evidence of actively squeezing out or unduly influencing AI developers. On the contrary, the largest cloud computing platforms seem to be actively building their portfolios of third-party foundation models.[74] They see this as a way to improve their services and grow their market share in cloud computing, which is (unlike AI) currently a highly profitable business.[75] This makes sense given the degree of competition between AI platforms: no competition authority has suggested that any of the hyperscalers' AI platforms is approaching a position of dominance, and one plausible outcome is that many different AI models will be successful, since different models will be more efficient at different use cases. Where there are formal partnerships between the hyperscalers and AI developers, these occasionally exclude or restrict how the AI model can be offered on other cloud platforms,[76] however as we describe above, most are non-exclusive. Some – like the Microsoft/OpenAI partnership – have become less exclusive over time. As we have noted above, in many cases it appears that AI developers retain significant freedom and autonomy even when they have entered into partnerships with hyperscalers.

A further concern arises where a platform does not facilitate customer switching. For example, when an AI developer fine-tunes one model using another larger model as the 'base', AI developers do not necessarily have continued direct access to the fine-tuned model other than through the hyperscaler's platform. This could result in AI developers being unable to move their models to another accelerated compute provider for the purposes of inference. This could pose a potential constraint to some smaller cloud providers which specialise in supporting inference but not training and/or fine-tuning.

[73] Meta, 'The Llama Ecosystem: Past, Present, and Future', 27 September 2023, available at https://ai.meta.com/blog/llama-2-updates-connect-2023/.
[74] Competition and Markets Authority, 'Cloud Infrastructure Services: Provisional decision report', 28 January 2025, para 3.454.
[75] See James Bessen, 'The New Goliaths: How Corporations Use Software to Dominate Industries, Kill Innovation, and Undermine Regulation', 2022.
[76] Competition and Markets Authority, 'Amazon.com Inc.'s partnership with Anthropic PBC: Decision on relevant merger situation', 27 September 2024.

2.6 Users of AI

The final stage in the AI value chain is the use of models for customer-facing applications. In some cases, AI developers such as OpenAI have produced their own applications such as, in OpenAI's case, ChatGPT. However, AI developers may also embed AI features in their existing products and services. Each of the largest cloud computing companies, many of which also host AI platforms, also provides a range of existing digital services which integrate AI features – such as Microsoft and Google's search engines and Amazon's online marketplace.

In many cases, the hyperscaler solely integrates its own AI models into its existing services. However, this is not always the case as, for example, Microsoft relies heavily on OpenAI services, and in other cases existing tech platforms – such as Apple and Google's operating systems – users can choose to use third-party AI services, even if they might not always have the full features of a fully integrated AI service. Nevertheless, **integration with an incumbent's services can give an AI developer an important 'anchor tenant', guaranteeing revenue and use of their AI model**.

In some markets, the digital services provided by a hyperscaler and in which AI is being integrated is a challenger, such as Microsoft's Bing. In other markets, the hyperscalers' existing digital services have a strong position in their respective markets.

Given that, as noted above, there is strong competition between hyperscalers in the provision of accelerated compute and in providing channels to market for AI developers, in our view, **the business practices of hyperscalers are most likely to be of concern where they involve leveraging or tying of an AI-related service with one of the hyperscaler's potentially dominant digital services in a more mature market.** In these contexts, however, competition authorities will need to examine the impacts of the practice in the particular case. Competition authorities should be cautious about intervention, since the integration of new AI model functionality may represent an innovation and benefit consumers. In some cases, integration may even be essential to help the hyperscaler innovate – for example, by building a strong enough customer base to justify making an otherwise too-risky investment. Although integration provides hyperscalers with an advantage which independent AI developers do not necessarily enjoy, the advantage does not so far seem to have been strong enough to dissuade significant investment in the AI sector (it may, however, be a factor influencing the direction of investment). Furthermore, it will not always be clear whether there is even (commercially or technically) potential for third-party competition to provide an AI feature deeply embedded with a large firm's existing service: arguably such features are simply an extension of the existing service (as when smartphones began integrating cameras) rather than a standalone market (as when smartphones began allowing third-party apps). There remain many other ways for AI developers to be successful.

2.7 Conclusions

As some competition authorities have already observed, the **hyperscalers along with NVIDIA enjoy some advantages over other accelerate compute providers and AI developers** – for example from having access to expensive inputs (such as GPUs) and existing extremely popular services which can serve as "anchor tenants" for AI models, provide data to train that model, ensuring that their own

accelerate compute enjoys significant demand.[77] Not all of these advantages can be replicated by smaller companies. **However, there is not yet compelling evidence that these advantages are insurmountable in the provision of FMs** – indeed, the level of investment, dynamism and innovation in the development of AI models suggests that investors believe many of these advantages could at least potentially be overcome, and that some alternative providers of accelerated compute have advantages over the hyperscalers, for example because their computing architecture is optimised for AI development. The biggest concerns about the future of the sector arise from potential changes in how the sector develops in future, and are likely to arise in how AI models are deployed downstream.

[77] Autorite de la Concurrence, 'Opinion 24-A-05 on the competitive functioning of the generative artificial intelligence sector', 28 June 2024.

3. Potential Future Developments

Both AI and the delivery of accelerated compute to the AI sector are growing at a fast pace, most of the addressable market is so far under-served, and levels of investment and innovation are very high. However, **the future shape of the sector – and way in which relationships between independent AI developers and the hyperscalers might develop – is uncertain**. This poses questions about the future characteristics of the sector, in particular around several interrelated questions identified in the previous section:

- Given that few AI developers are profitable, to what extent is consolidation likely and what will be the impacts on competitive dynamics?
- What does the increasing dependence of some AI models on other, larger models mean for the sustainability of competition? For example, when the sector is more mature, may AI developers find themselves locked into particular providers of accelerated compute or AI model platforms?
- Could the advantages and important positions in the AI value chain which hyperscalers currently enjoy translate to negative impacts on competition in the future?

3.1 Profitability and Sector Consolidation

Today, few if any AI developers are profitable. Investors do not tolerate loss-making business ideas indefinitely. There are several possible outcomes for AI developers (not mutually exclusive) including:

- leaving the market;
- consolidation into another firm with synergies, for example where the service remains loss-making on its own but adds value to the acquirer's overall ecosystem and is therefore cross-subsidised on an ongoing basis;
- turning a profit by raising prices or reducing costs, which will probably rely on one of, or a combination of, other firms leaving the market, developing a specialist feature which at least some users are prepared to pay for, or building a significant customer base to enable economies of scale; or
- developing new business models such as integrating online advertising into end-user services, as a number of AI providers are currently preparing to do.

The likelihood that a diverse range of AI providers can survive independently in the long run will depend on a number of factors. These largely relate not to factors that competition authorities have much control over, but rather to technological developments and the inherent economics of the sector, such as:

- **the extent to which the costs of accelerated compute for training and/or running AI models continues to decrease**. However, as many AI models are currently being offered for free or at very marginal prices, lowering costs by itself seems unlikely to allow all current AI developers to become profitable while remaining independent;

- **how the need for accelerated compute changes**. For example the low-cost, high-performance model DeepSeek was apparently able to develop its model despite US export controls which constrained its access to high-end AI accelerator chips, and using far less compute for fine-tuning than older AI models. DeepSeek prompted the emergence of many other small and cheaper AI models. AI developers such as Mistral, Microsoft and Apple have also successfully created smaller AI models using much less computing power.[78] More generally, LLM training is shifting away from data and compute-intense pre-training towards more targeted fine tuning, which demands less computing power.[79] Open-source AI models have also fundamentally changed the competitive environment. There are a number of well-performing open-source AI models on the market today, which AI developers can freely modify and fine-tune with far less need for accelerated compute than if they needed to train the models themselves.[80] Finally, the more AI models are able to recognise abstract principles and chains-of-thought, rather than simply relying on vast amounts of training data to identify the most statistically-likely appropriate outputs to a question, the less they might require ever-growing amounts of accelerated compute;
- **the extent to which customers are willing to pay for use of (or accept advertising in) AI models**, including where customer demand for AI settles in terms of balancing quality and accuracy, on the one hand, with cost on the other. Another compromise may lie in allowing AI models more time to infer the best answer to a given question, by allowing them to explore different reasoning processes, rather than relying solely on vast amounts of data and a 'brute force' approach to determining the most statistically relevant response;[81]
- **the extent to which there remains scope for AI developers to differentiate** – for example, will customers choose different AI providers and pay for specialised value-added services based on the use case, or since AI models can quickly learn from each other, will all models and related services essentially become "commoditised" with price becoming the primary choice driver for customers rather than functionality, price, and innovation?;
- **how plausible it is for the best performing AI models to prevent smaller models from "free riding" for example by using the best performing model to prepare customised training data** (for example to train the smaller model on chains-of-thought). Such restrictions could be developed either through technical means, enforceable contractual limitations, or new pricing models (for example higher prices for more extensive queries and outputs). It may be difficult to argue such restrictions are anti-competitive if the market for AI remains dynamic and the limitation is genuinely necessary to protect the larger model provider's incentive to continue investing and innovating; and

[78] See *Mistral NeMo: our new best small model* (18 July 2024); Microsoft Research Blog, *Phi-2: The surprising power of small language models* (12 December, 2023); Ars Technica, *Apple releases eight small AI language models aimed at on-device use* (25 April 2024).

[79] Bertin Martens, 'How DeepSeek has changed artificial intelligence and what it means for Europe', Bruegel, Policy Brief 12/25, March 2025.

[80] Meta says that "Tens of thousands of startups are using or evaluating Llama 2 including Anyscale, Replicate, Snowflake, LangSmith, Scale AI, and so many others": see https://ai.meta.com/blog/llama-2-updates-connect-2023/

[81] Bertin Martens, 'How DeepSeek has changed artificial intelligence and what it means for Europe', Bruegel, Policy Brief 12/25, March 2025, p 5.

- **whether market characteristics such as network effects and economies of scale prove significant**, such that only one or a small number of model providers provide higher quality models, and achieve an unassailable advantage, encouraging other AI developers to 'give up'.

It is beyond the scope of this issue paper to fully assess the likelihood and potential impacts of these changes, other than to note that they all imply potential changes to the cost and competitive structure of the AI sector. It is possible that market consolidation will prove inevitable due to relatively immutable characteristics of the sector – such as if all models tend towards homogeneity and customers are unwilling to pay significant amounts for accessing AI models.

However, some of the competitive dynamics which created 'winner takes all' outcomes in other digital markets are not – or at least not yet – as pronounced in the provision of accelerated compute and in AI. For example:

- many digital services are characterised by 'network effects', where a service becomes increasingly attractive as it acquires new users, creating a 'suction effect' that draws in new users and makes it more difficult for newer entrants to compete. **Network effects do not appear to be significant in the provision of accelerated compute.** They may become a more important characteristic in the AI sector if models start to significantly increase in quality based on human feedback – meaning that models with more users can improve faster than those with fewer users. However, the evidence of this happening is so far equivocal and **the increasing interdependence and mutual learning taking places between AI models currently suggests the impact of network effects could be relatively subdued**;
- economies of scale seem to play a role in the provision of accelerated compute, because this requires making expensive up-front investments in data centres and accelerator chips, the costs of which accelerated compute providers seek to recover from as many customers as possible. However, since accelerated compute is infrastructure-heavy, **the role of economies of scale seems somewhat smaller in cloud and AI than in some other digital markets**. Economies of scale have not prevented the growth of new accelerated compute providers like CoreWeave. In AI, economies of scale seem to be decreasing in significance, as the cost structure shifts away from high up-front costs (rewarding firms with the most customers, since they can spread the up-front costs among a larger user base) and towards higher ongoing costs per query. While some models may learn from how users respond, creating a positive feedback loop, many of the most sophisticated models now only have incremental (or in some cases dubious) performance improvements over previous models with less feedback – suggesting that the scope for improvements from additional data or feedback slows over time;[82]
- **providers of accelerated compute and AI do seem to enjoy economies of scope**. In particular, investments in data centres can be used to support general cloud computing, the provision of accelerated compute, and computing resources for hyperscalers' own digital services – since the requirements for land, electricity, and water are all common to both types of data centres,

[82] The Commission has said that "According to respondents and interviewees, some uncertainty remains around how powerful data feedback loops and network effects will be": Klaus Kowalski, Cristina Volpin, and Zsolt Zombori, 'Competition in Generative AI and Virtual Worlds', Competition Policy Brief, September 2024.

and a data centre requires little modification to switch uses.[83] Given that cloud computing is still a hugely growing sector, this allows hyperscalers to "de-risk" their investments in data centres in a way which AI providers or dedicated providers of accelerated compute cannot. However, given demand for both general cloud computing and accelerated compute are growing at a rapid rate, it seems likely that new entrants should also be equally able to enjoy economies of scope. Relatedly, vertical integration provides hyperscalers important advantages, by guaranteeing access to compute, data, channels to market and customers, which independent AI developers must negotiate commercially. However, the ability and willingness of hyperscalers to enter into (often non-exclusive) partnerships with AI developers suggests many of these advantages are currently replicable; and

- **there are powerful players in the AI value chain who have an interest in maintaining effective competition for the provision of accelerated compute**, or at least in limiting the overall economic power of the hyperscalers. For example, Apple is working on maximising the AI functionality that can be performed on a user's device without resort to the cloud; and NVIDIA has been sponsoring the growth of some of its smaller cloud computing customers to diversify its customer base.

Authorities will need to be careful when assessing consolidation: for example, allowing AI firms to consolidate and continue to provide value as part of a firm's broader digital ecosystem may be a better option than leaving them with no option but to close shop. **However, it is too early to assume that significant market consolidation – for example the case for accepting only one or two providers of FMs – would be inevitable**. Authorities need to ensure that "tipping" does not occur, but given this is not a particularly certain outcome even if market forces are allowed to play out, authorities should weigh carefully the risks and benefits of interventions.

3.2 Market Characteristics Which Might Constrain Smaller AI Developers

A more difficult balancing exercise may arise from active business decisions by firms which pose difficulties for smaller AI developers, or market characteristics which may make it harder for smaller providers to remain in the market. These include:

- **larger vertically integrated firms limiting access to essential inputs for AI developers** – such as compute, or access to their large models for fine-tuning or development of more specialised model;[84]
- **decisions or market characteristics which may constrain AI developers from switching** – for example, hyperscalers may be in a position to 'lock in' some AI developers to their cloud computing services or AI model platforms. Other lock-in effects may occur without any active decisions to make switching more difficult. For example, consumers and businesses may use

[83] The Economist, 'The data-centre investment spree shows no signs of stopping', 5 February 2025.
[84] This is already a possibility: for example, Meta's 'open' LLaMA 2 model requires a licence if the usage exceeds 700 million users: The Economist, 'The data-centre investment spree shows no signs of stopping', 5 February 2025, fn 170.

particular AI services, which over time may learn from that consumer or business's usage of the service, making it difficult to switch without losing that usage history; and

- **taking advantage of vertical integration**, such as through self-preferencing their own models on their AI model platforms for business users, tying services, or integrating and bundling their AI services with other 'must have' services, or giving only certain AI developers access to the most cutting-edge models.

Competition authorities will need to be cautious about overreacting to the potential for larger firms' decisions to limit smaller developers for two reasons.

First, these are mere hypotheses of future conduct: **currently, there is not much evidence that hyperscalers have incentives to harm AI developers. On the contrary, hyperscalers seem to have strong incentives to see the AI sector grow and thrive in order to maximise their own customer base** – and some firms like Microsoft have made public commitments about maintaining openness.[85] For example, the willingness of hyperscalers to enter into partnerships with AI developers – and to tolerate those AI developers entering into partnerships with other cloud computing providers, and to continue to develop services which compete with the hyperscalers' own offerings – suggests that AI developers are able to negotiate beneficial commercial deals.

Competition authorities will need to understand the dynamics and business incentives behind this positive market outcome, and be in a position to understand if and when these incentives may shift. This is particularly important when relying on metrics of dynamic competition to justify non-intervention. For example, uncertainty about the future of the sector – and which firms and types of services will succeed – seems to be one reason why the hyperscalers are supporting smaller AI developers. Hyperscalers may, for example, fear ceding the AI space to their competitors, who might then make their AI products the new focal point of digital ecosystems. For example, if consumers start using chatbots or new AI-powered devices as their main entry point to the digital world, then the influence of providers of existing operating systems, browsers, search engines and similar platform services could be radically diminished – in the same way that the internet (as an open and interoperable way of communicating across different types of devices, operating systems and browsers) allowed smartphones to disrupt the focal position of desktop-based operating systems. The digital sector has not enjoyed innovations which have truly disrupted incumbents in this way for many years. As long as this possibility exists (which we also discuss below), hyperscalers may prioritise preventing their competitors from securing an unassailable lead in that space, rather than being able to dominate it themselves. The tech firms will retain strong incentives to build open platforms in which no single player can dominate. Authorities should be conscious that this may change once the future direction of the AI sector becomes clearer, but should also try to maximise the investment and benefits for consumers that the current level of uncertainty is producing.

Second, **even if damage to smaller AI developers occurs due to decisions by larger AI developers, there may be sensible economic justifications and the impacts might not be anti-competitive**. For example, it could be justifiable in future for a hyperscaler to limit access to their models to produce competing models, if that is necessary to protect the hyperscaler's incentives to invest and innovate

[85] Brad Smith, 'Microsoft's AI Access Principles: Our commitments to promote innovation and competition in the new AI economy', 26 February 2024, available at https://blogs.microsoft.com/on-the-issues/2024/02/26/microsoft-ai-access-principles-responsible-mobile-world-congress/.

in its own services. Similarly, integration of AI into existing services is being pursued by small and large tech firms as a way of delivering new product improvements, and can be an important way to persuade customers to try a new feature, and to create efficiencies, lower prices and better-quality services. Competition authorities will need to assess such justifications and their rationale extremely carefully, but it is commonly accepted that practices like self-preferencing should not be assumed to be anti-competitive. While business decisions may harm some independent AI developers, they will not necessarily harm levels of competition, provided there remains a thriving ecosystem of AI firms. The presence and success of genuinely open-source models suggests that independent AI firms will continue to have wide access to large FMs into the future.

Nevertheless, authorities ought to be alert to future changes in the market which may constrain the choices of AI developers, business users and consumers. As noted above, some lock-in practices may arise without any active anti-competitive conduct and already seem to be occurring: for example, as business users and consumers use an AI service, they will build up a history of queries and outputs, allowing the AI model to provide answers more tailored to the user's needs. Ensuring users can transfer this history from one AI model to another may be key to ensuring ongoing freedom of choice.[86] Authorities should therefore consider encouraging the development of portability and interoperability tools for innovative AI services, even if their implementation may prove complex. In other cases, such as if larger firms start closing down access to previously open models, platforms or computing power, authorities will need to review the effects of that conduct on their merits and anticipatory intervention would not seem appropriate.

3.3 Influence Over the Direction of Independent AI Developers

Finally, there is a broader and more overarching question of **whether a market structure where hyperscalers have strong positions in the AI value chain (and are a major provider of funding to independent AI developers) will negatively influence the direction of AI development**, persuading independent AI developers to deploy models in ways which complement rather than challenge the hyperscalers – in other words, creating an implicit 'kill zone'.

The idea that AI could become a new focal point for digital ecosystems has led to **suggestions from some competition authorities that independent AI developers should be 'protected' from the influence of large technology firms** – and concern that independent AI developers may be forced "to cooperate with big tech firms to get access to computing infrastructure and end users".[87] Presumably, this is so that AI developers can feel freer to focus on radical and disruptive innovation rather than 'sustaining' innovation which complements rather than upends hyperscalers' existing market positions in tech markets. The CMA, for example, has set out 'principles' to guide how the AI sector

[86] Chris Riley, 'The future of AI hinges on data portability and APIs', Data Transfer Initiative, 11 February 2025, available at https://dtinit.org/blog/2025/02/11/future-of-AI-portability; Chris Riley, 'Digging in on personal AI portability', Data Transfer Initiative, 4 June 2024, available at https://dtinit.org/blog/2024/06/04/digging-in-personal-AI.

[87] Bertin Martins, 'Why artificial intelligence is creating fundamental challenges for competition policy', Bruegel, policy brief, 18 July 2024.

should look – emphasising high levels of interoperability and that hyperscalers should not exercise "undue influence" over independent AI firms.[88]

While authorities' instinct to protect possibilities for disruptive innovation is sound, it is not clear at this stage that this requires special protection of independent AI developers. First, **there is no compelling evidence of independent AI developers shaping their business plans or investments due to the influence of hyperscalers** – on the contrary, many AI products and services are being developed by independent AI developers that directly compete with the hyperscalers, and which could pose radical challenges to the existing ecosystems of incumbent tech firms.

Second, **there is no obvious reason why the influence of hyperscalers should be assumed to always be negative**. If there is fierce competition between hyperscalers, they will have the incentives to pursue radical innovations in order to disrupt their competitors – as we saw when Microsoft's integrated OpenAI services into its search engine Bing,[89] which in turn prompted Google to quickly release its own AI models. The hyperscalers also have the ability to pursue radical innovations (and help radical innovations by independent AI developers succeed) due to the hyperscalers' size, access to capital, and their ability to leverage their existing customer bases to encourage a critical mass of consumers to try an experimental new service. Chatbots like ChatGPT have attracted significant consumer interest, but they have not yet posed existential challenges to existing large digital platforms; it is more likely that the benefits of AI will be enjoyed by greater numbers of customers, and disruptions will happen, when they are well integrated into existing services. That does not necessarily preclude a role for European digital industrial policy to support domestic technologies and firms – but it does mean that such initiatives should be justified on grounds other rather than competition policy.

Competition policy should ensure that competition between hyperscalers, and between hyperscalers and newer firms, remains fierce rather than trying to minimise the influence of hyperscalers per se.

[88] Competition and Markets Authority, 'CMA AI strategic update', 29 April 2024.
[89] Statista, 'Global search engine traffic market share of Bing from January 2018 to January 2025', available from https://www.statista.com/statistics/1219326/market-share-held-by-bing-worldwide/.

4. Conclusion

Competition authorities have been scrutinising the cloud and AI sectors, vigilant to ensure that do not intervene too late, after a market has already 'tipped'. Past experience in digital markets shows that radical changes to market structure can happen relatively quickly. However, the dynamics of the AI and accelerate compute sectors have some different characteristics, and the future direction of the sectors is highly uncertain, which means authorities ought to consider carefully both the benefits and the potential risks of intervening.

Analysing competition in these sectors – which are nascent and still growing quickly – requires a different approach to competition policy. Authorities are rightly keen to ensure that AI and the provision of accelerated compute are not foreclosed – but they also need to assess levels of current and future competition based on metrics that acknowledge the high levels of innovation, investment and growth in the sector. For example, if many firms are investing, that provides a strong indication that investors see potential space to compete, and that regulatory interventions to promote competition might prove unnecessary or even counterproductive. The table below indicates markers of weak and strong competition in a dynamic market, and illustrates why we have concluded that competition between AI developers is strong and that the largest providers of accelerated compute are subject to growing pressure from actual and potential competitors, and from broader technological and commercial developments.

Table 2. Markers of Dynamic Competition

	Markers of lower levels of dynamic competition	Markers of high levels of dynamic competition
Nature of investments	Investment tends to be for assets which complement, rather than challenge, existing services.	Investment tends to be for 'radical big bets', suggesting that disruptive innovation remains a strong possibility. There is strong evidence of this type of investment in AI where players like DeepSeek have produced (or capitalised on) groundbreaking innovations. The provision of accelerated compute is also characterised by significant investment for example in new chips designed specifically for AI training.
Source of investments	Only a few big players can attract significant capital for investments.	A range of providers are able to attract funding, indicating that non-incumbents believe they still enjoy scope for growth. This is true both for AI developers, many of which are start-ups, and for emerging providers of accelerated compute such as CoreWeave.

Innovation	Innovations are slow to appear and tend to be incremental to existing services; 'leapfrog' innovations are rare.	Innovations emerging from the market are frequent, heterogenous and sometimes radical – indicating a lack of certainty about which solutions will succeed, and about the boundaries of the market. Market players repeatedly 'leapfrog' each other in innovation. This seems true in AI (where the most sophisticated models are leapfrogging each other, and other models such as DeepSeek are making significant leaps in efficiency) and accelerated compute (where players are investing in innovative new chipsets and computing architectures).
Pricing	Prices are relatively static and reflect cost, with low levels of risky investment.	Prices may be initially below-cost as a way of encouraging customers to try innovative products and 'create' markets, as in many of today's AI markets where services are free or very low cost. The pricing structure for different providers of accelerate compute varies greatly, suggesting high levels of risk-taking and experimentation.
Market saturation	Customer needs are relatively well understood.	Customers are still identifying their needs and firms/customers are both experimenting with the right product fit. This is clearly true in AI, whose uses and applications are still being fully explored, and in the accelerated compute sector where firms are offering a range of different solutions.

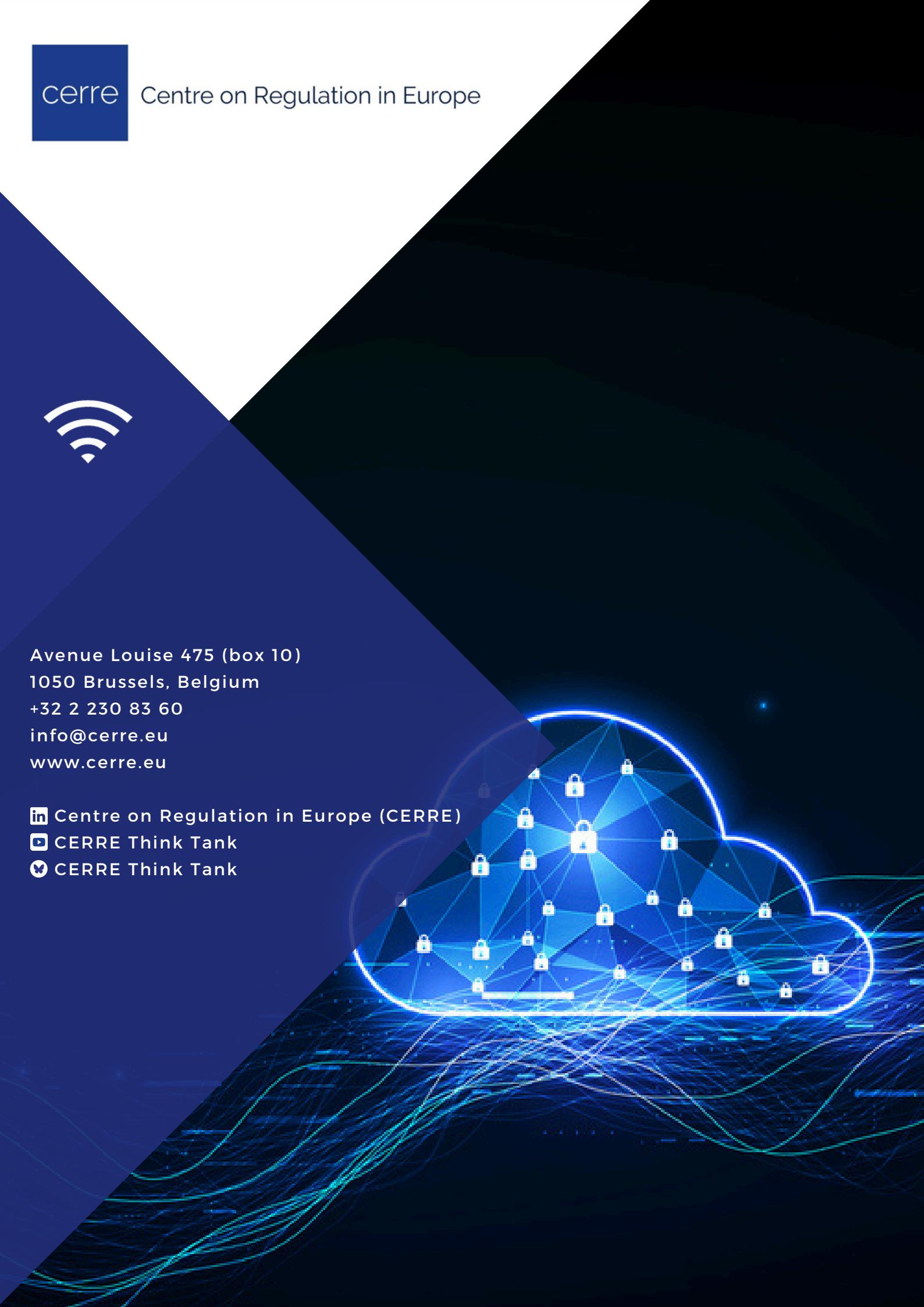
The three hyperscalers seem likely to enjoy a strong position in providing accelerated compute to AI developers, along with providing other important inputs to AI developers such as access to "parent" models and channels to market. Currently, there is insufficient evidence to conclude that this concentration is likely to persist, or that the hyperscalers will be in a position and have the incentive to foreclose competition in the AI or accelerate compute sectors. **However, since relying on metrics of dynamic competition requires making assessment of firms' capabilities and future market dynamics, competition authorities will need to maintain close scrutiny of changes in the sector, in particular any changes in the business practices of hyperscalers.** Competition authorities should understand why hyperscalers currently support open markets and ensure they are in a position to intervene quickly if the sectors' dynamics shift away from openness. That may require competition authorities to ensure that AI developers still enjoy incentives to pursue disruptive and radical innovation – rather than focusing on reducing the role of hyperscalers as a goal in itself.

The EU will have to be careful to apply competition policy impartially and objectively. Geopolitical rivalry with China means that Europe may too eagerly disregard the positive competitive pressure that Chinese firms add to the market. But from a competition analysis, Chinese AI developers – particularly

those releasing open-source models, which can be publicly scrutinised, used and adapted – provide an important competitive spur to AI developers worldwide. Similarly, the EU is under increasing pressure to decrease its reliance on US AI developers. But in the meantime, the EU is severely underperforming in commercialising AI.[90] Growing transatlantic tensions may raise questions about the shape of an effective EU digital industrial policy. Nevertheless, competition authorities should continue to prioritise ensuring effective competition, including by supporting competitive dynamics that support innovation and investment, rather than assuming the impact of foreign firms is necessarily negative for competition.

[90] In 2024, European AI startups raised the equivalent of $12.5 billion in venture capital funding, far less than the $81.4 billion raised by US AI startups: WorldFund, 'Green Computing in the AI Era', white paper, 1 April 2025.

cerre Centre on Regulation in Europe



Avenue Louise 475 (box 10)
1050 Brussels, Belgium
+32 2 230 83 60
info@cerre.eu
www.cerre.eu

Centre on Regulation in Europe (CERRE)
CERRE Think Tank
CERRE Think Tank