

Future of Life Institute's Feedback to the European Commission's consultation on competition in Generative AI

March 10th, 2024

1) What are the main components (i.e., inputs) necessary to build, train, deploy and distribute generative AI systems? Please explain the importance of these components

Generative AI (GAI) systems are the user-facing applications built on top of general purpose AI (GPAI) models. These models undergo training and inference using cloud computing, typically infrastructure as a service (IaaS), and advanced semiconductors.¹ This requires access to limited and expensive chips, cloud capacity, based on an extensive volume of servers, vast amounts of (high quality) data, and sought-after skills needed to develop competitive GPAI models, including discovering innovative algorithms that advance the state-of-the-art.

Currently, buying API access to these models, or downloading open source alternatives, to adapt them into generative AI applications is much cheaper and faster than building these models in-house from scratch.² This means that only a handful of the most well-resourced corporations in history can afford to bankroll the development of the models that structure and underpin the applications upon which they are built. The generality of these models in competently performing a wide range of distinct tasks means that they could quickly become the digital infrastructure that forms the bedrock of the entire economy.³

GPT-4 boasts an end-user base of 100 million weekly active users and a business user base of over two million developers using it as a platform, including 92% of Fortune 500 companies.⁴ OpenAI's GPT store allows users to develop and monetise their own GPTs, illustrating the base layer infrastructural nature of their GPAI model.⁵ These corporations are also preeminent in other markets, allowing them to disseminate GPAI models across cloud, search, social media, operating systems, app stores, and productivity software.⁶ Thus, the implications of market concentration are much starker than for other technologies. The combination of concentrated development resources with ubiquitous adoption and distribution throughout adjacent markets risks a winner-take-all scenario, as explored in this feedback.

¹ Narechania and Sitaraman. "[An Antimonopoly Approach to Governing AI](#)".

² UK CMA. "[AI Foundation Models Initial Report](#)".

³ Vipra and Korinek. "[Market concentration implications of foundation models: The invisible hand of ChatGPT](#)".

⁴ Malik. "[OpenAI's ChatGPT now has 100 million weekly active users](#)".

⁵ Stringer, Wiggers, and Corral. "[ChatGPT: Everything you need to know about the AI-powered chatbot](#)".

⁶ Lynn, von Thun, and Montoya. "[AI in the Public Interest: Confronting the Monopoly Threat](#)".

2) What are the main barriers to entry and expansion for the provision, distribution, or integration of generative AI systems and/or components, including AI models? Please indicate to which components they relate.

Increasing the number of model parameters enhances a model's capabilities by improving its capacity to learn from data. However, this requires more computing power, in chips and cloud capacity, and data, which makes it cost prohibitive for many SMEs and startups.⁷ The market for chips, the first layer in the AI value chain, is highly concentrated, a phenomenon which is exacerbated by shortages stemming from significant demand-supply imbalances in components. General purpose AI is fuelled by the parallel computation processing capabilities of NVIDIA-designed graphic processing units (GPUs), which capture 90% of the market⁸, and are manufactured by Taiwan Semiconductor Manufacturing Company (TSMC), which, in turn, captures the largest share in the global chips foundry market at 56%⁹. Many developers train AI models in CUDA, NVIDIA's proprietary software development platform, but they must use NVIDIA's GPUs.¹⁰ Even the well-capitalised challengers in this market observe competition issues, as OpenAI's CEO, Sam Altman, has sought to raise \$5-7 trillion to create his own chip-building capacity, highlighting the difficulties of competing on chips.¹¹

While the hardware market in semiconductors is almost a monopoly, the infrastructure market is more like an oligopoly¹², which should still concern the Commission from a competition perspective.¹³ Amazon's AWS (31%), Microsoft's Azure (24%) and Google Cloud (11%) collectively cover two thirds of the cloud computing market.¹⁴ This collective dominance arises from the significant investment required to establish data centres, server farms, and the network infrastructure to interconnect them.¹⁵ If OpenAI, Anthropic or DeepMind were to create their own in-house cloud infrastructure, independent of the Big Tech companies that have partnered, merged, or acquired them, it would require considerable investments in land, energy, and datacentre equipment (cabling servers, server racks, coolers, etc.).¹⁶ While the Data Act may abolish egress charges, excluding those related to parallel data storage, there remain additional non-financial hurdles hindering GPAI model developers from establishing their own in-house cloud hosting infrastructure. Namely, risks of services downtime for their customers, including generative AI developers and end-users.¹⁷

Hyperscalers (large cloud service providers that can provide computing and storage services at enterprise scale) enjoy privileged access to limited hardware resources, enabling them to offer

⁷ Carugati. [“The Generative AI Challenges for Competition Authorities”](#).

⁸ Techo Veda. [“NVIDIA has 90% of the AI GPU Market Share; 1.5 to 2 million AI GPUs to be sold by NVIDIA in 2024”](#).

⁹ Statista. [“Semiconductor foundries revenue share worldwide from 2019 to 2023, by quarter”](#).

¹⁰ Whittaker, Widder, and West. [“Open \(For Business\): Big Tech, Concentrated Power, and the Political Economy of Open AI”](#).

¹¹ Field. [“OpenAI CEO Sam Altman seeks as much as \\$7 trillion for new AI chip project: Report”](#).

¹² Informed by discussion with Friso Bostoën, Assistant Professor of Competition Law and Digital Regulation at Tilburg University.

¹³ AI Now Institute. [“Computational Power and AI”](#).

¹⁴ Statista. [“Amazon Maintains Cloud Lead as Microsoft Edges Closer”](#).

¹⁵ Narechania and Sitaraman. [“An Antimonopoly Approach to Governing AI”](#).

¹⁶ Belfield and Hua. [“Compute and Antitrust”](#).

¹⁷ Narechania and Sitaraman. [“An Antimonopoly Approach to Governing AI”](#).

exclusive access to GPAI models developed internally, in collaboration with partners, or through investments, thereby creating serious barriers to entry.¹⁸ Amazon not only provides cloud infrastructure to Anthropic, Stability AI, and AI21, but also competes with them by offering its own GPAI models on its Amazon Bedrock platform.¹⁹ Cloud hosts have unparalleled power to monitor, detect, and stifle competitors emerging on their cloud platform²⁰, while resources generated using upstream dominance allows them to influence research and development downstream.²¹

Big Tech-backed research papers are frequently cited in research, indicating a notable uptake of their ideas among the wider scientific community.²² Their ownership and operation of software development frameworks – standardised processes for developing AI and other software, including vast data repositories, training methods, and evaluation tools – shapes AI development and deployment by ensuring that engineers adhere to development practices that are interoperable with Big Tech products.²³ Although PyTorch functions as a research foundation within the Linux Foundation, it is bankrolled by Meta. Google’s TensorFlow programming is specifically designed for Google’s Tensor Processing Units (TPUs), Google’s inhouse AI semiconductors, available on Google Cloud Platform, facilitating Google’s vertical integration from development practices to compute resources.

Developers of the most advanced GPAI models currently on the market have a first-mover advantage. It is more straightforward for OpenAI to maintain and attract business users, because some of those clients may be hesitant to switch to GPAI competitor due to data security concerns or the cost and complexity of moving their data²⁴, as previously witnessed in the cloud space²⁵. Having made the large initial investment, OpenAI have a head start, learning from and building upon their already advanced model, while recovering that initial investment from the monetisation of GPT-4, as others seek to get their models off the ground.

Early entry allows these providers to purchase access to compute and data at rates that will be lower than for new entrants if increased demand pushes prices up.²⁶ It affords them greater time to improve their models’ performance through finetuning, build brand visibility, and a strong consumer base. They have a head start in harvesting user data to feed into future training runs and developing higher performance models. This reinforces their prominence, whereby greater performance attracts more users, builds better trust in the model at the expense of new and unknown alternatives, and gives them capital to continue crowding out the market.

¹⁸ Bornstein, Appenzeller and Casado. [“Who Owns the Generative AI Platform?”](#)

¹⁹ UK CMA. [“AI Foundation Models Initial Report”](#).

²⁰ Lynn, von Thun, and Montoya. [“AI in the Public Interest: Confronting the Monopoly Threat”](#).

²¹ Kak, Myers West, and Whittaker. [“Make no mistake – AI is owned by Big Tech”](#).

²² Giziński et al. [“Big Tech influence over AI research revisited: memetic analysis of attribution of ideas to affiliation.”](#)

²³ Whittaker, Widder, and West. [“Open \(For Business\): Big Tech, Concentrated Power, and the Political Economy of Open AI”](#).

²⁴ Economist. [“Could OpenAI be the next tech giant?”](#)

²⁵ Savanta. [“European cloud customers affected by restrictive licensing terms for existing on-premise software, new research finds”](#).

²⁶ UK CMA. [“AI Foundation Models Initial Report”](#).

The best models currently belong to Big Tech, not universities – in 2022, industry produced 32 leading models, while academia produced three.²⁷ This reduces academic access to cutting-edge models for evaluations of systemic risks and the development of effective mitigation measures. Compared to nonprofits and universities, the private sector has the most resources to recruit the best talent, use large amounts of compute, and access data, both in quantity and quality, all of which is required to build state-of-the-art GPAI. This limits the amount of high skilled workers, needed to build the most competitive AI to industry, hindering academia in training the next generation of advanced AI developers.²⁸ As a result, supply is not meeting demand, not least because there is a race to find better engineers, who can discover algorithmic innovations that reduce the amount of compute or data – and costs – required for training.²⁹ SMEs and startups must try to attract talent away from more resourceful incumbents, who can offer bigger employee remunerations.

GPAI models, and generative AI systems, involve fixed costs in development, such as pretraining and fine-tuning compute resources, data collation, and inhouse and outsourced labour, and relatively low marginal costs in deployment.³⁰ These economies of scale are a significant barrier to entry for startups, as they would need to develop and deploy models and systems at scale from the outset in order to compete.³¹ It is usually more realistic for smaller European providers to fine-tune American models into customised models or domain-specific systems that require less compute, data, and labour.³² But this still renders downstream developers and deployers dependent on larger upstream model providers.

The general purpose nature of these AI models and systems allows for versatile and flexible deployment settings, which will increase their uptake throughout diverse industries. For providers, this allows them to spread substantial initial investment spending across these use cases, while their downstream customers will save money by deploying the same capability across different tasks.³³ These economies of scope are a challenging barrier to entry for Big Tech challengers, as they would need to be servicing the same range of sectors in order to compete.³⁴

3) What are the main drivers of competition (i.e., the elements that make a company a successful player) for the provision, distribution or integration of generative AI systems and/or components, including AI models?

The leading GPAI models and generative systems are more performant because they have access to the most or best data, computational resources, and skilled developers. These factors allow them to attract more users; amass more data and capital to purchase more chips; access more cloud infrastructure; develop better models and applications; and, in turn, attract more and

²⁷ Standford University HAI. [“AI Index Report 2023”](#).

²⁸ UK CMA. [“AI Foundation Models Initial Report”](#).

²⁹ Vipra and Korinek. [“Market concentration implications of foundation models: The invisible hand of ChatGPT”](#).

³⁰ Ada Lovelace Institute. [“Foundation models in the public sector”](#).

³¹ Vipra and Korinek. [“Market concentration implications of foundation models: The invisible hand of ChatGPT”](#).

³² Leicht. [“The Economic Case for Foundation Model Regulation”](#).

³³ Vipra and Korinek. [“Market concentration implications of foundation models: The invisible hand of ChatGPT”](#).

³⁴ UK CMA. [“AI Foundation Models Initial Report”](#).

better developers.³⁵ OpenAI engineers can make up to \$800,000 per year, salaries that no SME or startup, especially in Europe, could afford.³⁶ Importantly, as their models become increasingly capable, doors open up for the leading GPAI providers to monetise and advertise their models, as well as enter into commercial relationships with downstream system developers, which not only provides even greater user-facing visibility, but can also offer access to specialised domain or task specific data that is held by particular downstream parties.³⁷ If they are unable to obtain such unique data from these partnerships, then their increased revenues can be used to purchase it elsewhere.

These network effects are accelerated by data feedback effects, whereby general purpose AI developers leverage data generated from the conversations between the system and its users to advance capabilities.³⁸ While data generated during user interactions is not automatically used to train the model, since developers need to vet feedback for quality and safety, this may change if innovations lead to safe automatic continuous learning post-deployment.³⁹ OpenAI recently announced that ChatGPT will be able to memorise conversations in order to better tailor its responses to user preferences.⁴⁰ The more GPAI developers can refine a model toward their customers, the more useful it will be for customers, who will be less inclined to try out another challenger platform.⁴¹

Even if they mainly use feedback data in aggregate to understand wider trends, this is still a considerable competitive advantage for the most widely used models and systems that can collect the most amount of user data, providing more enriched aggregate analysis. Companies like OpenAI are at a particular advantage because they are present at both the model and system level, allowing them to use system level feedback to improve their model. European GPAI system developers, who will be largely reliant on building their systems upon American GPAI models, would be unable to benefit from this feedback loop, because they would be unable to use the data generated from their system to improve the underlying model. Superior access to resources – capital, computing power, data, and expertise – enables the creation of superior models. These models attract more consumers, resulting in increased revenue. This revenue, in turn, provides access to even better resources, thus perpetuating the cycle of developing high-quality models, asserting market dominance, and the capacity to foreclose competition from challengers.

4) Which competition issues will likely emerge for the provision, distribution, or integration of generative AI systems and/or components, including AI models? Please indicate to which components they relate.

While user feedback may not necessarily be leveraged for marketing services or constructing advertising profiles, enhanced capabilities can improve downstream GPAI services. This enables more precise customisation to consumer preferences, thereby driving adoption rates and

³⁵ Hausfeld. [“ChatGPT, Bard & Co.: an introduction to AI for competition and regulatory lawyers”](#).

³⁶ Constantz. [“OpenAI Engineers Earning \\$800,000 a Year Turn Rare Skillset Into Leverage”](#).

³⁷ Schrepel and Pentland. [“Competition between AI Foundation Models: Dynamics and Policy Recommendations”](#).

³⁸ OpenAI. [“How your data is used to improve model performance”](#).

³⁹ UK CMA. [“AI Foundational Models Initial Report”](#).

⁴⁰ OpenAI. [“Memory and new controls for ChatGPT”](#).

⁴¹ UK CMA. [“AI Foundational Models Initial Report”](#).

establishing user loyalty.⁴² End users and business users will be locked in unless it is sufficiently practical to port data when switching to an alternative. Even with adequate interoperability, they may be discouraged from trying alternatives due to the convenience of accessing all their GPAI and related tools, services, and plug-ins via the one established platform. Such lock-in creates a positive feedback loop for the GPAI provider, positioning the model for further monetisation, as it continues to progressively build a more robust and holistic picture of the user, thereby empowering it to offer more tailored targeting of products, including the provider's other downstream GPAI services in adjacent markets, such as search, social media, app stores and productivity software.

This grants the provider the power to engage in unfair and abusive practices. Dominance in both the GPAI model and system market coupled with dominance in these adjacent markets allows large incumbents to buttress their dominance in each by bundling their GPAI service with their other services in search, online platforms, or productivity software. Beyond the convenience of a one-stop shop interface, users may be unable to switch if doing so means they would lose access to another tied service. The first-mover advantage of the currently leading GPAI models – GPT-4, Gemini, Llama 2 – allows them to enjoy network effects, and with customer lock-in, switching costs will also act as a barrier to entry for SMEs and startups.

5) How will generative AI systems and/or components, including AI models likely be monetised, and which components will likely capture most of this monetization?

As recognised by Stanford researchers⁴³, when GPAI model providers grant sufficient access to their models to downstream system developers, through an application programming interface (API), they are operating AI as a platform, similar to platform as a service (PaaS) for software, allowing them to access models to adapt to specific user facing GPAI and AI systems, like an app store for app developers. Beyond this, OpenAI, for example, also allows plugin integrations that connect third-party apps to the paid version of ChatGPT (based on GPT-4, not GPT-3.5, as in the free version). This increases capabilities by allowing ChatGPT to retrieve real-time information, proprietary information, and action real-world user instructions.⁴⁴ Plugins empower ChatGPT to act as a platform by enabling it to select options among different providers or present different options to the user.⁴⁵

More recently, OpenAI launched GPT Store⁴⁶, so its non-expert paying users can find and build fine-tuned versions of the ChatGPT GPAI system.⁴⁷ All of this attracts third-party app and plugin developers to OpenAI's ecosystem, rendering more applications compatible with its main GPAI system, while providing OpenAI with oversight on developments that threaten their offerings.⁴⁸ Smaller plugin providers, in particular, may come to rely on platforms like ChatGPT,

⁴² Ibid.

⁴³ Stanford University HAI. "[AI Accountability Policy Request for Comment](#)".

⁴⁴ OpenAI. "[Chat Plugins](#)".

⁴⁵ Vipra and Korinek. "[Market concentration implications of foundation models: The invisible hand of ChatGPT](#)".

⁴⁶ OpenAI. "[Introducing the GPT Store](#)".

⁴⁷ Sentance. "[The GPT Store isn't ChatGPT's 'app store' – but it's still significant for marketers](#)".

⁴⁸ OpenAI. "[ChatGPT plugins](#)".

the fastest growing consumer application in history⁴⁹, for their user base⁵⁰, but OpenAI may use this position to provide their own competing applications downstream.⁵¹ As OpenAI expands its plug-in offerings, their platform becomes more appealing for plug-in developers, allowing OpenAI to draw in more plug-ins, which increases the amount of consumers, motivates more developers, and makes their platform ever-more appealing.⁵²

6) Do open-source generative AI systems and/or components, including AI models compete effectively with proprietary AI generative systems and/or components? Please elaborate on your answer.

The considerable costs required to develop general purpose AI models from square one and then deploy them at scale, apply equally to closed and open models. While open source licenses offer new entrants more accessibility at the model level (parameters, data, training support), open source models do not address compute concentration in the markets for semiconductors and cloud infrastructures.⁵³ All take frontier open source models rely on Big Tech compute⁵⁴: Meta's Llama 2 runs on Microsoft Azure; UAE-based Technology Innovation Institute's Falcon 180B model runs on AWS⁵⁵; and Mistral's Mixtral models runs on Google Cloud⁵⁶. EleutherAI's GPT-NeoX-20B runs on NVIDIA-backed, AI-focused CoreWeave Cloud⁵⁷, who rent out GPUs at an hourly rate⁵⁸, allowing them to scale from 3 to 14 data centres in 2023⁵⁹, but remains well below Meta and Microsoft, who are NVIDIA's top GPU customers⁶⁰. Microsoft have committed to billions of dollars in investment in CoreWeave in the coming years to secure access to NVIDIA's GPUs⁶¹ ahead of their real rivals, AWS and Google Cloud⁶².

At first glance, Meta's Llama 2 meets the definition of a free and open source license in the recently agreed AI Act, considering that Meta publishes the model parameters, including weights, and information on model architecture and model usage. However, Meta does not publish information on the model training data – precisely why providers of such models are required to do so under the AI Act, regardless of whether they present systemic risks or not. Nevertheless, Meta's Llama 2 licence⁶³ is not open source⁶⁴, as widely recognised⁶⁵, particularly

⁴⁹ Reuters. [“ChatGPT sets record for fastest-growing user base - analyst note”](#).

⁵⁰ UK CMA. [“AI Foundational Models Initial Report”](#).

⁵¹ Narechania and Sitaraman. [“An Antimonopoly Approach to Governing AI”](#).

⁵² UK CMA. [“AI Foundational Models Initial Report”](#).

⁵³ Narechania and Sitaraman. [“An Antimonopoly Approach to Governing AI”](#).

⁵⁴ Whittaker, Widder, and West. [“Open \(For Business\): Big Tech, Concentrated Power, and the Political Economy of Open AI”](#).

⁵⁵ Jackson. [“TII trains state-of-the-art LLM, Falcon 40B, on AWS model”](#).

⁵⁶ Reuters. [“Google Cloud partners with Mistral AI on generative language models”](#).

⁵⁷ Hjelm. [“Looking Ahead to 2023: How CoreWeave Is Using NVIDIA GPUs to Advance the New Era of AI and Machine Learning”](#).

⁵⁸ Krazit. [“How CoreWeave went all-in on Nvidia to take on Big Cloud”](#).

⁵⁹ Economist. [“Data centres improved greatly in energy efficiency as they grew massively larger”](#).

⁶⁰ Elder. [“Sell Nvidia”](#).

⁶¹ Novet. [“Microsoft signs deal for A.I. computing power with Nvidia-backed CoreWeave that could be worth billions”](#).

⁶² Haranas. [“Microsoft's CoreWeave Deal 'Adds AI Pressure' To AWS, Google”](#).

⁶³ Meta. [“Request access to Llama”](#).

⁶⁴ OpenUK. [“State of Open: The UK in 2024 Phase One AI and Open Innovation”](#).

⁶⁵ Tarkowski. [“The Mirage of Open-source AI: Analysing Meta's LLaMa 2 release strategy”](#).

by the Open Source Initiative⁶⁶, whose open source definition⁶⁷ is the global community benchmark. Meta does not allow developers to use Llama 2 or its outputs to improve any other large language model (LLM), and app developers with more than 700 million monthly active users must request a license from Meta, which Meta is not obliged to grant, presumably if it feels competitively challenged.⁶⁸ By permitting commercial use of Llama 2, on a small and non-threatening scale, Meta leverages unpaid labour to enhance the model's architecture, enabling it to monetise such improvements, as endorsed by their CEO.⁶⁹

European SMEs and startups will still be highly dependent on the largest chips developers (largely NVIDIA) and cloud providers (Amazon, Microsoft, and Google), as well as the leading AI development frameworks (Meta and Google). This dependence affirms and asserts gatekeepers' market monitoring powers that can anticipate and foreclose competition from innovative new entrants through self-preferencing or copying.⁷⁰ Even with leveraging open source GPAI models, EU players will still need major funding to train and deploy their GPAI models, if they are to be competitive, which will need to come from EU governments and venture capital firms, if they are not to be bought up by Big Tech. Otherwise, EU GPAI developers will be limited to fine-tuning existing models, open or closed, which does not empower downstream parties to fundamentally alter data and design choices that were shaped upstream.⁷¹

According to Mistral, their latest Mixtral 8x7B model matches or exceeds Meta's Llama 2 70B and OpenAI's GPT-3.5 on many performance metrics and is better on maths, code, and multilingual tasks, while using fewer parameters during inference.⁷² By utilising only a portion of the overall parameters per token, it effectively manages costs and reduces latency. It is open source (though this is reasonably disputed)⁷³, released under the Apache 2.0 license, and free for academic and commercial use. Until recently, the European developer's capacity to develop competitive GPAI models was supported by €385 million, among other funding, including from American venture capital firms, such as Andreessen Horowitz and Lightspeed.⁷⁴ Building on their successes, and seeking to secure their long-term financing and operational viability, they have reached a deal with Microsoft, who will invest €15 million. This allows Mistral to use Microsoft supercomputers to train their GPAI models on Azure and access Microsoft customers for greater distribution of their products, while it allows Microsoft to offer Mistral models as premium features for its customers.⁷⁵ The partnership positions Microsoft with a leading presence in both the open source model market (through Mistral) and closed proprietary model

⁶⁶ Open Source Initiative. [“Meta’s LLaMa 2 license is not Open Source”](#).

⁶⁷ Open Source Initiative. [“The Open Source Definition”](#).

⁶⁸ OpenSource Connections. [“Is Llama 2 open source? No – and perhaps we need a new definition of open...”](#).

⁶⁹ Whittaker, Widder, and West. [“Open \(For Business\): Big Tech, Concentrated Power, and the Political Economy of Open AI”](#).

⁷⁰ Lynn, von Thun, and Montoya. [“AI in the Public Interest: Confronting the Monopoly Threat”](#).

⁷¹ Whittaker, Widder, and West. [“Open \(For Business\): Big Tech, Concentrated Power, and the Political Economy of Open AI”](#).

⁷² Jiang et al. [“Mixtral of Experts”](#).

⁷³ Robertson. [“France’s Mistral takes a victory lap”](#).

⁷⁴ Volpicelli. [“Microsoft’s AI deal with France’s Mistral faces EU scrutiny”](#).

⁷⁵ Volpicelli. [“European lawmakers question Commission on Microsoft-Mistral AI deal”](#).

market (through OpenAI).⁷⁶ While Microsoft's investment in Mistral currently doesn't confer ownership stake, it could convert to equity in Mistral's subsequent funding round.⁷⁷

This episode vividly illustrates that when an open source alternative appears to threaten the most well-funded proprietary models, such as GPT-4, those funding the challenged model quickly move in to stake their financial interests in the upstart new entrant, foreclosing competition. Microsoft is hedging its bets in case Mistral's models should come to outperform those of their other investment, OpenAI, in case open source AI becomes the dominant business model or ecosystem that generates the greatest economic value. While open source holds promise for greater transparency and accessibility, this development underscores that it is incredibly difficult for open source AI models to get off the ground without the financial backing of Big Tech.

It highlights that the AI Act threshold for classifying models as systemic – those models trained on compute using 10^{25} or more FLOPS – should not be raised, as desired by industry. During trilogue discussions, and even now, the French government argue that the threshold should be 10^{26} , in part due to concerns that their national champion, Mistral, would reach the threshold within a year. The deal between Microsoft and Mistral makes it clear that reaching that threshold, which depends on vast resources in cloud computing capacity, requires funding from those with entrenched positions in digital markets.

The partnership has undermined the self-proclaimed⁷⁸ European independence of Mistral.⁷⁹ For EU policymaking, naturally there is concern about conflicts of interest during the AI Act negotiations, as highlighted by Green MEPs in a letter to the Commission⁸⁰, especially considering that this deal was likely also under negotiation over the same period. While this may not reach a threshold of a competition breach or market abuse, the Commission should be concerned when European AI startups, that are able to achieve a certain scale, can only survive through gatekeeper funding. This renders the European AI startup vulnerable to being co-opted as a useful voice or vehicle for Big Tech lobbying that seeks to minimise their compliance burden at the expense of safety for European consumers. For highly capable or impactful open model GPAI, risks are amplified by the inability of the original provider to effectively remediate or recall a dangerous open model after it has been released and downloaded innumerable times. While their inherent transparency may have benefits for accountability, it can also provide malicious actors with access to the model weights, enabling the discovery and exploitation of vulnerabilities or the circumvention of guardrails to generate harmful illegal outputs, including the development of lethal weapons, cyberattacks against critical infrastructure, and electoral manipulation.

⁷⁶ Murgia. [“Microsoft strikes deal with Mistral in push beyond OpenAI”](#)

⁷⁷ Coulter and Yun Chee. [“Microsoft's deal with Mistral AI faces EU scrutiny”](#).

⁷⁸ Mensch. [X \(Twitter\) post](#).

⁷⁹ Zenner. [“Microsoft-Mistral partnership and the EU AI Act”](#)

⁸⁰ Volpicelli. [“European lawmakers question Commission on Microsoft-Mistral AI deal”](#).

7) What is the role of data and what are its relevant characteristics for the provision of generative AI systems and/or components, including AI models?

While publicly available data is still broadly accessible to new entrants, public data can be low quality, leading to less capable and even dangerous models.⁸¹ Stanford researchers found that one of the leading public datasets, LAION-5B, includes thousands of images of child sexual abuse material.⁸² Lensa, an image generator built on top of the GPAI model Stable Diffusion, which is trained on LAION-5B, was found to create realistic sexualised and nude avatars of women, particularly from traditionally marginalised communities, with less propensity to do the same in male renditions when prompted.⁸³

Proprietary datasets can offer more specialised and unique data that will give a model a deeper understanding of the world.⁸⁴ This not only makes a model more capable, but also allows it to be more easily aligned with our interests since it can understand us better in theory. This mitigates biases and inaccuracies within models, generating trust and encouraging adoption, thereby facilitating positive feedback loops for those with the best data. Big Tech's accumulated data banks – both personal data from their B2C markets and non-personal data from their B2B/B2G markets – gives them an edge, as they have access to the public datasets that new entrants would, as well as their own enormous proprietary datasets which are closed off to new entrants.⁸⁵ High quality proprietary data is often held in downstream companies that specialise in a certain domain and have gathered data on that domain's customers.⁸⁶ Google's \$2.1 billion acquisition of Fitbit gives them millions of users' health data, which has been tracked for over a decade, as well as access to Fitbit's insurance partners.⁸⁷ This allows Google to leverage this wealth of proprietary data if they seek to fit their GPAI models with health features tailored to their users, giving them a competitive edge over models without this capability. Such an acquisition is beyond the reach of European startups.

The innovation gap is widened further by Big Tech's greater experience in developing models, housing the best expertise, and scraping, labelling, and analysing data. Moreover, search engine providers, namely Google and Microsoft⁸⁸, can leverage public data more effectively by using web indexes to filter out meaningless or useless information, leaving behind the high quality public data, which may be a more efficient process since the web data is more vast than proprietary datasets.⁸⁹ One way European SMEs and startups could catch up is through algorithmic innovations that can do more with less data, but this requires access to the best talent, which is another increase in costs. The current competitive frontier goes even further in

⁸¹ UK CMA. [“AI Foundation Models Initial Report”](#).

⁸² Ropek. [“An Influential AI Dataset Contains Thousands of Suspected Child Sexual Abuse Images”](#).

⁸³ Heikkilä. [“The viral AI avatar app Lensa undressed me without my consent.”](#)

⁸⁴ UK CMA. [“AI Foundation Models Initial Report”](#).

⁸⁵ Lynn, von Thun, and Montoya. [“AI in the Public Interest: Confronting the Monopoly Threat”](#).

⁸⁶ Vipra and Korinek. [“Market concentration implications of foundation models: The invisible hand of ChatGPT”](#).

⁸⁷ Austin. [“The Real Reason Google Is Buying Fitbit”](#).

⁸⁸ Hausfeld. [“ChatGPT, Bard & Co.: an introduction to AI for competition and regulatory lawyers”](#).

⁸⁹ UK CMA. [“AI Foundation Models Initial Report”](#).

that ChatGPT and Gemini will compete on how many other systems they are connected to, providing them with continuous real-time up-to-date data.

Successes in commercial GPAI and its vast potential across innumerable use cases have also led to data providers seeking to cash in on the AI gold rush by monetising their data for AI training.⁹⁰ When data was free and easy to access, the current GPAI model leaders got in early.⁹¹ As content creators, or the platforms on which such content is hosted, restrict access, or seek remuneration, new entrants may face barriers to entry with prohibitively costly data on top of exorbitant compute costs. If legislation and judicial rulings reassert rightsholders' intellectual property rights, public data could become increasingly scarce, pushing the price up further.⁹² European SMEs and startups could turn to synthetic data as an alternative to proprietary data, but more resources on compute are needed to generate such artificial information.⁹³ Saving on data pushes costs up elsewhere. Using AI models to generate data for future models can transfer errors and bugs from the old model to the new one.⁹⁴

8) What is the role of interoperability in the provision of generative AI systems and/or components, including AI models? Is the lack of interoperability between components a risk to effective competition?

The concentrated cloud market, upon which GPAI is developed and deployed, combined with the lack of interoperability between AWS, Azure, and Google Cloud Platform, provides single points of failure that could be disruptive and destabilising across sectors, given the market share of these three providers.⁹⁵ As single failure points, they are an attractive target for cyberattacks by malicious actors.⁹⁶ If such an attack were successful, it would cut off not only the cloud infrastructure platform, but also the general purpose AI model and the many downstream generative AI systems deriving from it that run on the same cloud. The lack of interoperability means critical applications, such as those in defence, health, or finance, cannot be easily migrated to another cloud provider in order to get them up and running again.⁹⁷ In a scenario where a hostile nation or well-funded terrorist group penetrates a single point of failure to cripple critical services, a full blown assault on both private and public databases could not only cause widespread disruption, but it may also be difficult to detect, making it all the more challenging to restore organisational data to a safe and reliable state.⁹⁸ Concentrations at the model level can also produce similar security risks given that any vulnerability in a model upon which user-facing applications are built could produce systemic hazards, exacerbated by

⁹⁰ Narechania and Sitaraman. [“An Antimonopoly Approach to Governing AI”](#).

⁹¹ Lynn, von Thun, and Montoya. [“AI in the Public Interest: Confronting the Monopoly Threat”](#).

⁹² UK CMA. [“AI Foundation Models Initial Report”](#).

⁹³ Vipra and Korinek. [“Market concentration implications of foundation models: The invisible hand of ChatGPT”](#).

⁹⁴ UK CMA. [“AI Foundation Models Initial Report”](#).

⁹⁵ Narechania and Sitaraman. [“An Antimonopoly Approach to Governing AI”](#).

⁹⁶ Lutkevich. [“Foundation models explained: Everything you need to know”](#).

⁹⁷ Narechania and Sitaraman. [“An Antimonopoly Approach to Governing AI”](#).

⁹⁸ World Economic Forum. [“Understanding Systemic Cyber Risk”](#).

emergent capabilities that can develop unpredictably with further fine-tuning at the system level.⁹⁹

9) Do the vertically integrated companies, which provide several components along the value chain of generative AI systems (including user facing applications and plug-ins), enjoy an advantage compared to other companies? Please elaborate on your answer.

General purpose AI models and generative AI systems that are plugged into third-party applications can operate as platforms, or they can be vertically integrated into another platform.¹⁰⁰ To catch up with Google, Microsoft integrated GPT-4 as Bing Chat into its platforms, Bing search and Edge¹⁰¹, and as Copilot into its 365 productivity suite.¹⁰² As a response, Google is testing its integration of generative AI into Google Search in the US¹⁰³ – Search Generative Experience (SGE)¹⁰⁴ – which allows Google to leverage an adjacent market, bolstering its advertising revenues and strengthening its grip on online website traffic.¹⁰⁵ This illustrates a transition from search engines to answer engines.¹⁰⁶ Users may be less inclined to visit third-party websites, provided as citations or footnotes, since the answer is already in the chatbot interface, to which advertisers turn their attention at the expense of third-party websites.¹⁰⁷ This could allow Google's generative search engine to benefit from the intellectual property of others, whose data is fed into the generative interface, not only without compensation, but also without the usual user traffic to their site.

For users, and society at large, reliance on generative search engines risks reducing the accuracy of information, as it is difficult to distinguish between outputs derived from the training data and those from the search results, and the hidden hallucinations therein.¹⁰⁸ Stanford found that users' perception of utility is inversely correlated with the precision of citations purporting to support claims made by the generative search/answer engine.¹⁰⁹ While Bing Chat achieves the highest citation precision rate, users find it the least useful, whereas YouChat has the lowest citation precision rate, but users deem it the most useful. Given that upstream GPAI models are likely to significantly augment the user experience or perceived utility, if not accuracy, of downstream search or answer engines, users will be increasingly drawn to these platforms¹¹⁰, which will be a barrier for entry for GPAI model providers that don't compete on both nodes of the value chain by only offering GPAI models, but not GPAI-infused search engines.¹¹¹

⁹⁹ Vipra and Korinek. [“Market concentration implications of foundation models: The invisible hand of ChatGPT”](#).

¹⁰⁰ Ibid.

¹⁰¹ Techsyn. [“Microsoft Integrates OpenAI’s GPT-4 Model Into Bing For A Powerful Search Experience”](#).

¹⁰² Sullivan. [“Inside Microsoft’s sprint to integrate OpenAI’s GPT-4 into its 365 app suite”](#).

¹⁰³ Google. [“Supercharging Search with generative AI”](#).

¹⁰⁴ Google. [“Search Labs”](#).

¹⁰⁵ Lynn, von Thun, and Montoya. [“AI in the Public Interest: Confronting the Monopoly Threat”](#).

¹⁰⁶ Carugati. [“Competition in Generative AI Foundation Models”](#).

¹⁰⁷ Carugati. [“The Generative AI Challenges for Competition Authorities”](#).

¹⁰⁸ Miller. [“Generative Search Engines: Beware the Facade of Trustworthiness”](#).

¹⁰⁹ Liu, Zhang, and Liang. [“Evaluating Verifiability in Generative Search Engines”](#).

¹¹⁰ Vipra and Korinek. [“Market concentration implications of foundation models: The invisible hand of ChatGPT”](#).

¹¹¹ UK CMA. [“AI Foundation Models Initial Report”](#).

Google is present throughout the entire AI value chain¹¹²: it produces its own semiconductors (TPUs), hosts its own cloud infrastructure (Google Cloud), develops its own GPAI models (PaLM-2 and Gemini), provides GPAI systems (Gemini, so users can interact directly with the model)¹¹³ and integrates those systems into its platforms (Search Generative Experience). From these platforms, it also gathers data that can be used to improve future iterations of Gemini, increasing the model's capabilities and utility. The revenues can also be used to purchase more compute, data, and talent. Vertically integrated corporations will have easier access to unique data, such as conversations between users on their platforms.¹¹⁴

Vertical integration risks institutionalising unbreakable tech oligopolies, hindering innovative efforts of European SMEs and startups, weakening consumer choice, and inflating the cost of gatekeeper services beyond their value, either in subscription charges or data extraction. While NVIDIA is currently leading on GPUs, Microsoft, Google and Meta are all seeking to compete by developing their own chips for AI.¹¹⁵ If Microsoft or Google were to overtake NVIDIA, their vertical integration, either legally or in practice, from semiconductors (Microsoft's ; Google's TPUs) to cloud (AWS; GCP) to models (GPT-4; Gemini) to systems (Bing, ChatGPT, Copilot; Gemini) could ensure that AI development becomes a two-horse race, as it would be incredibly difficult, if not impossible, for challengers to compete at each level of that value chain. In this scenario, Microsoft or Google could then engage in unfair and abusive practices, such as limiting the access of GPAI model and system competitors to key ingredients like chips and cloud infrastructure.

Microsoft's CEO, Satya Nadella, claims ¹¹⁶ that his company's partnership with OpenAI challenges vertically integrated companies like Google.¹¹⁷ Yet, concerns are mounting that the partnership may amount to a decisive influence under Article 3 of the Mergers Control Regulation, given their exclusivity arrangements, as well as the successful pressure Microsoft put on OpenAI's board to reinstate their fired CEO, Sam Altman, including offering him and other OpenAI staff employment.¹¹⁸ This raises questions about OpenAI's ability to operate independently and be considered a separate company that is not vertically integrated with Microsoft "in spirit", if not in narrow legal terms. The grey-area manoeuvres of Altman's firing and rehiring illuminate that Microsoft can control OpenAI, without acquiring it or breaking antitrust or merger laws, by leveraging its exclusive access to their leading GPAI models and the scaleups access to gatekeeper's compute – an arrangement that prohibits OpenAI from migrating their models to other cloud providers .¹¹⁹

¹¹² Narechania and Sitaraman. [“An Antimonopoly Approach to Governing AI”](#).

¹¹³ Google. [“Bard becomes Gemini: Try Ultra 1.0 and a new mobile app today”](#).

¹¹⁴ UK CMA. [“AI Foundation Models Initial Report”](#).

¹¹⁵ David. [“Chip race: Microsoft, Meta, Google, and Nvidia battle it out for AI chip supremacy”](#).v

¹¹⁶ Hartmann. [“Microsoft CEO defends OpenAI's ‘partnership’ amid EU, UK regulators’ scrutiny”](#).

¹¹⁷ Smith. [“Microsoft's AI Access Principles: Our commitments to promote innovation and competition in the new AI economy”](#).

¹¹⁸ Irish Council for Civil Liberties et al. [“Submission to European Commission on Microsoft-OpenAI “partnership” merger inquiry”](#).

¹¹⁹ Callaci. [“The Antitrust Lessons of the OpenAI Saga”](#).

10) What is the rationale of the investments and/or acquisitions of large companies in small providers of generative AI systems and/or components, including AI models? How will they affect competition?

Cloud providers typically prefer to create partnerships with established GPAI providers, affording the latter preferential access to scarce computational resources and investment opportunities.¹²⁰ This is cheaper for the GPAI developer than paying access via on-demand rates or via upfront or subscription charges, let alone building their own data centre. OpenAI must use Azure, while Microsoft can integrate OpenAI products across all its offerings¹²¹, with priority access.¹²²

11) Do you expect the emergence of generative AI systems and/or components, including AI models to trigger the need to adapt EU legal antitrust concepts?

While the Digital Markets Act (DMA) is not strictly an antitrust instrument, it does seek to ensure open digital markets and to provide an additional lever in the Commission toolkit for lengthening antitrust investigations. Although the DMA does not explicitly cover AI, generative AI should be in-scope when it is integrated into a core platform service.¹²³

At the infrastructural level in GPAI model and system development and deployment, cloud computing is already listed as a core platform service. However, none have been designated at the time of writing, primarily due to hyperscalers not meeting the quantitative thresholds given that they don't technically have end-users, according to the DMA definition.¹²⁴ There is recognition that business users may also be counted as end users when they use cloud computing services for their own purposes (Recital 14 of the DMA). This should be included when counting active end users of cloud computing services, given that AI labs such as OpenAI and Anthropic (and the many other businesses fine-tuning their GPAI models via an API that is run on cloud services) might all be considered end-users¹²⁵ of Azure, Google Cloud Platform and AWS, rather than business users¹²⁶, based on DMA definitions. This could mean that the cloud services of hyperscalers would be designated as core platform services, and would thereby ensure that the oligopolist cloud market is tackled by EU ex-ante regulation, rather than complaints brought by cloud service challengers that would struggle to afford lengthy legal proceedings.

¹²⁰ UK CMA. [“AI Foundation Models Initial Report”](#).

¹²¹ Lynn, von Thun, and Montoya. [“AI in the Public Interest: Confronting the Monopoly Threat”](#).

¹²² Irish Council for Civil Liberties et al. [“Submission to European Commission on Microsoft-OpenAI “partnership” merger inquiry”](#).

¹²³ Informed by discussion with Friso Bostoën, Assistant Professor of Competition Law and Digital Regulation at Tilburg University.

¹²⁴ Abecasis et al. [“6 reflections on the recent designation of gatekeepers under the DMA”](#).

¹²⁵ Digital Markets Act definition of active end users for cloud computing: “Number of unique end users who engaged with any cloud computing services from the relevant provider of cloud computing services at least once in the month, in return for any type of remuneration, regardless of whether this remuneration occurs in the same month.”

¹²⁶ Digital Markets Act definition of active business users for cloud computing: “Number of unique business users who provided any cloud computing services hosted in the cloud infrastructure of the relevant provider of cloud computing services during the year.”

As in the Commission’s initial DMA impact assessment¹²⁷, the end user and business user definitions should equally cover infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) in order to avoid loopholes in the DMA’s application. If this is not already the case, the Commission should amend and update the methodology and list of indicators in the Annex of the DMA through delegated acts (Article 3(7)) to ensure the DMA can mitigate further concentrations in the cloud market, which underpins GPAI development and deployment. To ensure the DMA remains fit for purpose given the rapid advances in AI, as well as API as an intermediation platform, the Commission should consider whether they have the legal basis to update the list of core platform services to accommodate GPAI models and systems.¹²⁸

According to reports, Microsoft has cautioned competing search engines that it will terminate licenses providing access to its Bing search index if they continue to use it for generative AI chat development.¹²⁹ Article 6(2) of the DMA prohibits gatekeepers from using non-publicly available data, generated by business users or their customers using the core platform service, to compete with those business users. This could help to ensure that GPAI model providers are hindered in preventing GPAI system developers, dependent on the model provider for API access, from competing using data generated through their use of the cloud service.¹³⁰ Although Bing has not been designated, it may reach the threshold if its integration with OpenAI models makes it more attractive to end users and business users.

Given their foundational and systemic function across the economy and society, large cloud computing and GPAI model providers should be regulated like public utilities, adhering to similar rules on non-discrimination, equitable treatment of all customers, and assurance of safety and security.¹³¹ Since antitrust law primarily seeks to address monopolies, public utility framing is critical, as the oligopoly in the cloud market may make the AI market more concentrated in the coming years.¹³²

The Commission could also consider the feasibility of structural separation to prohibit companies from owning both GPAI models and other technologies and platforms that enable them to engage in unfair and abusive practices.¹³³ While this could be achieved through antitrust law, it typically requires a lengthy investigation process, which means that regulation may be

¹²⁷ European Commission. [“Impact assessment of the Digital Markets Act 1/2”](#): “Cloud services . . . provide infrastructure to support and enable functionality in services offered by others and at the same time offer a range of products and services across multiple sectors, and mediate many areas of society. . . They benefit from strong economies of scale (associated to a high fixed cost and minimal marginal costs) and high switching costs (associated to the integration of business users in the cloud). The vertical integration of the large cloud services providers and the business model they deploy has contributed to further concentration on the market, where it is very difficult for other less-integrated players, or market actors operating in just one market segment to compete. Consequently, these startups are likely to be completely reliant on large online platform companies.”

¹²⁸ von Thun. [“EU does not need to wait for the AI Act to act”](#).

¹²⁹ Dixit. [“Microsoft reportedly threatens to cut-off Bing search data access to rival AI chat products”](#).

¹³⁰ Yasar et al. [“AI and the EU Digital Markets Act: Addressing the Risks of Bigness in Generative AI”](#).

¹³¹ Lynn, von Thun, and Montoya. [“AI in the Public Interest: Confronting the Monopoly Threat”](#).

¹³² Informed by discussion with Friso Bostoan, Assistant Professor of Competition Law and Digital Regulation at Tilburg University.

¹³³ von Thun. [“After Years of Leading the Charge Against Big Tech Dominance, is the EU Falling Behind?”](#)

more viable. As in the AI Act, at present, the amount of compute used during training is one of the best ways of quantifying a GPAI model's impact and capabilities. Based on the current state of the art, the Commission could use compute as a proxy for determining the largest market players in order to apply structural remedies that would mitigate market failures.¹³⁴

12) Do you expect the emergence of generative AI systems to trigger the need to adapt EU antitrust investigation tools and practices?

Notwithstanding the already increased scrutiny of competition authorities towards issues related to the digital economy in recent years¹³⁵, detecting and assessing potential competition law infringements will become increasingly complex. Such complexity is particularly pronounced when facing companies with business models that deploy network effects or benefit from ecosystems¹³⁶, which generate and collect data to enhance value. This data allows companies to refine algorithms and services, which subsequently increases their value on markets. Their business models often use GPAI models and systems, blockchain, IoT, robotics, algorithms, and machine learning¹³⁷ to offer services, such as providing search results (Google), recommending products (Amazon) or accommodation (Airbnb).¹³⁸ These digital platforms centered around data are changing competitive dynamics rapidly, posing considerable challenges for competition authorities.

As a result, the current competition law enforcement framework and tools will be under pressure. It might be necessary to account for increasingly more diverse parameters, beyond the traditional focus on output and prices. For example, in fast-moving and dynamic markets powered by AI, competition authorities will be required to capture and understand data more quickly. In addition, in the context of devising a market definition, which has become more complex for digital platforms, the traditional SSNIP test may no longer suffice. Similarly, while the EU Merger Regulation can be somewhat adapted, it doesn't adequately capture collaborations where companies like Microsoft partner with, and provide minority investments, in other parties (such as OpenAI) gaining influence and control without outright ownership.¹³⁹ If it is not possible to tackle these kinds of relationships, of vertical outsourcing rather than vertical integration, then reassessment and revision of the Merger Regulation is needed.

GPAI also enables companies to engage in new kinds of anticompetitive behaviour (see also the answer to question 4). For example, algorithms enable companies to automatically monitor the prices of competitors in real time and then re-price (algorithmic collusion). Companies with substantial market power in a certain market, may use GPAI to reinforce their market power in another market or to exclude competitors (as seen in the Google Shopping Case¹⁴⁰).

¹³⁴ Belfield and Hua. [“Compute and Antitrust”](#).

¹³⁵ [Google Android decision](#); [Apple Pay Investigation](#); [Apple App Store investigation](#); [Amazon's use of marketplace seller data Investigation](#).

¹³⁶ Lianos: Hellenic Competition Commission and BRICS Competition Law and Policy Centre. [“Computational Competition Law and Economics: An Inception Report”](#).

¹³⁷ Schrepel. [“Collusion by Blockchain and Smart Contracts”](#).

¹³⁸ Jansiti and Lakhani. [“From Disruption to Collision: The New Competitive Dynamics”](#).

¹³⁹ Lynn, von Thun, and Montoya. [“AI in the Public Interest: Confronting the Monopoly Threat”](#).

¹⁴⁰ T-612/17 - Google and Alphabet v Commission (Google Shopping)

In view of the transformations and advancements stemming from the emergence and deployment of GPAI, there is a significant risk that competition authorities may struggle to grasp companies' behaviour and market dynamics in a timely manner in order to prevent anti-competitive effects from taking place. Considering that the European Commission directly enforces EU competition rules and possesses an extensive toolkit for antitrust investigations, it is imperative to bolster enforcement tools and reevaluate how competition is analyzed to ensure EU competition policy remains future proof. By fostering a competitive GPAI market and value chain, other regulations – such as the AI Act, the Product Liability Directive, the forthcoming AI Liability Directive, the Data Act, the GDPR etc. – become more enforceable. Monopolists and oligopolists should not become too big to regulate, treating fines for non-compliance with these EU laws as operating expenses.¹⁴¹ Better compliance improves AI safety, fostering trust, and accelerating adoption of beneficial AI across the EU, while levelling the playing field for innovative European AI startups to offer competitive alternatives.

¹⁴¹ Lynn, von Thun, and Montoya. [“AI in the Public Interest: Confronting the Monopoly Threat”](#).