# The Generative AI challenges for competition authorities

Dr Christophe Carugati

## Executive summary

Generative Artificial intelligence (GenAI) stimulates market and regulatory developments. It has spurred competition and innovation and created new emerging markets and technologies ranging from chips, machine learning models, cloud computing, and software to answer engines.

Market developments pose competition issues. Developers need access to three key components in the value chain: computing resources, machine learning models, and data. Despite various players in the field, certain sectors face competition concerns due to market features and potential issues, such as tying.

Regulatory developments impact GenAI. In particular, GenAI raises unsolved regulatory challenges arising from the use of copyrighted data (intellectual property rights), personal data (data protection), AI risks (AI governance) and competition.

Competition authorities closely monitor GenAI developments by doing market studies. These initiatives will inform how competition in GenAI works.

Challenges for competition authorities stem from emerging markets and technologies, along with regulatory instabilities. New products and services are shaping new and existing markets, like answer engines and advertising. Regulatory uncertainties influence competition in GenAI.

At this current development stage, competition authorities should focus on understanding market and regulatory developments by cooperating among themselves and with relevant competent authorities. They should exercise formal enforcement powers and potentially update competition tools only when necessary and justified, guided by the insights gained from these market studies.

# About

## Digital Competition

Digital Competition (digital-competition.com) is an independent advisory firm that aims to develop open digital and competition policies for better innovation. We leverage expertise and dialogue with stakeholders while maintaining our commitment to providing open access, full transparency, and impartial advice to tech firms, law firms, consulting firms, competition authorities and international organisations. This working paper did not receive any funding.

This paper is part of our GenAI and Competition Hub (https://www.digital-competition.com/genaiandcompetitionhub), which strives for responsible GenAI development, ensuring favourable market conditions that benefit all. Our Hub helps stakeholders and decisionmakers navigate complex and rapid GenAI market and regulatory development. We also nurture the discussion in designing competition policies that deliver favourable market conditions in the context of intense monitoring of GenAI by competition authorities worldwide and the forthcoming 2024 G7 Italian presidency.

We provide research and market studies and invite stakeholders to contribute with relevant input. We also offer private consultations, training, and conferences on GenAI and competition. Contact us to join the Hub and/or for consultation/press inquiries.

We will present this working paper at the Concurrences' event on Generative AI and Antitrust with Linsey Mccallum (European Commission), Paulo Rocha Abecasis (Copenhagen Economics) and Tone Oeyen (Freshfields Bruckhaus Deringer). Registration is free of charge (https://www.concurrences.com/en/evenement/generative-ai-antitrust)

## Dr. Christophe Carugati

Dr. Christophe Carugati (christophe.carugati@digital-competition.com) is the founder of Digital Competition. He is a renowned and passionate expert on digital and competition issues with a strong reputation for doing impartial, high-quality research. After his PhD in law and economics on Big Data and Competition Law, he is an ex-affiliate fellow at the economic think-tank Bruegel and a lecturer in competition law and economics at Lille University.

# 1 Introduction

Generative Artificial Intelligence (GenAI) has been a catalyst for market and regulatory developments. Since the launch of the chatbot ChatGPT in November 2022, the tech industry has frequently released new products and services. These range from chips, machine learning models, cloud computing, and software to answer engines, showcasing GenAI's impact in fostering competition, innovation, and the creation of emerging markets and technologies.

Nevertheless, running and deploying GenAI requires developers to access three main components in the value chain: computing resources, machine learning models, and data. Despite various players in the field, certain sectors face competition concerns due to market features and potential issues, such as tying.

Computing resources, including graphic cards and cloud computing resources, are pivotal in the development and deployment of machine learning models. However, access to these resources is primarily provided by a limited number of tech firms. To secure access, some model developers enter into partnerships with large cloud computing providers, exchanging access to their models for access to cloud computing resources. Although these collaborations might foster competition, they may also give rise to problematic practices such as tying (Carugati, 2023c).

Machine learning models generate output based on input data, such as text, images, videos, or music. Despite claims that models are concentrated in the hands of a few large online platforms[1], there is no substantive evidence supporting this assertion. On the contrary, thousands of closed-source and open-source models compete on various parameters like task requirements, language specifications, and model size.

Data, another essential component, raises concerns about a potential competitive advantage for models provided by large online platforms with access to vast proprietary datasets[2]. However, like models, there is no substantive evidence supporting this concern. Model

[1] Diane Coyle, Preempting a Generative AI Monopoly, *Project Syndicate*, 2 February 2023 (accessed 29 January 2024). Available at: https://www.project-syndicate.org/commentary/preventing-tech-giants-from-monopolizing-artificial-intelligence-chatbots-by-diane-coyle-2023-02

[2] Georg Riekeles and Max von Thun, AI Won't be Safe Until We Rein in Big Tech, *European Policy Centre*, 22 November 2023 (accessed 29 January 2024). Available at: https://www.epc.eu/en/publications/AI-wont-be-safe-until-we-rein-in-Big-Tech~55e63c

developers can leverage thousands of public and proprietary datasets, differentiated based on task requirements, language specifications, and domain specificity.

Furthermore, GenAI developments grapple with unsolved regulatory challenges related to the use of copyrighted data (intellectual property rights), personal data (data protection), AI risks (AI governance), and competition.

Competition authorities around the globe, including those in the United Kingdom (Competition and Markets Authority, 2023), Portugal (Autoridade da Concorrência, 2023), India[3], Hungary[4], Europe[5], the United States[6], and soon France[7], are closely monitoring GenAI developments through forward-looking analyses.

In this context, GenAI poses challenges for competition authorities due to emerging markets and technologies coupled with regulatory instabilities. New products and services are shaping new and existing markets, like answer engines and advertising. Regulatory uncertainties are influencing competition in GenAI.

At this current developmental stage, competition authorities should focus on understanding market and regulatory developments through cooperation among themselves and relevant competent authorities. They should exercise formal enforcement powers and potentially update competition tools only when necessary and justified, guided by the insights gained from these market studies.

[3] CCI Conducting Market Study on Artificial Intelligence to Assess Competition Landscape, *Business Outlook India*, 8 December 2023 (accessed 16 January 2024). Available at: https://business.outlookindia.com/news/cci-conducting-market-study-on-artificial-intelligence-to-assess-competition-landscape

[4] GVH Launches Market Analysis on the Impact of Artificial Intelligence, *Hungarian Competition Authority*, 4 January 2024 (accessed 16 January 2024). Available at: https://www.gvh.hu/en/press_room/press_releases/press-releases-2024/gvh-launches-market-analysis-on-the-impact-of-artificial-intelligence

[5] Commission Launches Calls for Contributions on Competition in Virtual Worlds and Generative AI, *European Commission*, 9 January 2024 (accessed 16 January 2024). Available at: https://ec.europa.eu/commission/presscorner/detail/en/IP_24_85

[6] FTC to Host Virtual Summit on Artificial Intelligence, *Federal Trade Commission*,5 January 2024 (accessed 16 January 2024). Available at: https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-host-virtual-summit-artificial-intelligence

[7] Benoît Cœuré Delivers his 2024 New Year's Message, *Autorité de la concurrence*, 18 Janvier 2024 (accesssed 25 January 2024). Available at : https://www.autoritedelaconcurrence.fr/en/article/benoit-coeure-delivers-his-2024-new-years-message

# 2 Emerging Markets and Technologies

GenAI requires developers to have access to three main components. This value chain includes computing resources, machine learning models, and data (Carugati, 2023b). Then, application developers integrate GenAI into their products and services.

## 2.1 Computing Resources

Computing resources enable the development and deployment of machine learning models. The two main components include graphic cards for computation and AI workloads and cloud computing for running and deploying models at scale over the internet.

The graphic card sector is the main driver of running models. However, the sector faces card shortages due to high demand and low supply of components (Carugati, 2023c). In this sector, Nvidia is the leading supplier, especially Graphics Processing Units (GPUs), which perform several computations simultaneously. Competition authorities in France, Europe, the United States and China are currently investigating Nvidia's business practices[8].

However, new players enter the graphic card market, challenging Nvidia's position. Advanced Micro Devices (AMD) and Intel have announced graphic cards dedicated to AI[9]. Meta, Amazon, and Alphabet also develop in-house chips to improve AI workloads[10]. However, to what extent chips provided by new players and in-house chips exert competition pressure on Nvidia deserves in-depth scrutiny in the graphics card sector, and it is thus out of the scope of this paper.

[8] Form 10-Q Nvidia Corporation, 21 November 2023 (accessed 19 January 2024). Available at: https://investor.nvidia.com/financial-info/sec-filings/sec-filings-details/default.aspx?FilingId=17074143

[9] *See* for Intel, Intel Core Ultra Ushers in the Age of the AI PC, *Intel*, 14 December 2023 (accessed 19 January 2024). Available at: https://www.intel.com/content/www/us/en/newsroom/news/core-ultra-client-computing-news-1.html *See* for AMD, AMD Showcases Growing Momentum for AMD Powered AI Solutions from the Data Center to PCs, *AMD*, 6 December 2023 (accessed 3 January 2024). Available at: https://www.amd.com/en/newsroom/press-releases/2023-12-6-amd-showcases-growing-momentum-for-amd-powered-ai-.html

[10] *See* for Meta, Santosh Janardhan, Reimagining Our Infrastructure for the AI Age, *Meta*, 18 May 2023 (accessed 19 January 2024). Available at: https://about.fb.com/news/2023/05/metas-infrastructure-for-ai/

*See* for Amazon, AWS Trainium, *Amazon AWS* (accessed 19 January 2024). Available at: https://aws.amazon.com/machine-learning/trainium/

*See* for Alphabet, Introduction to Cloud TPU, *Google* (accessed 19 January 2024). Available at: https://cloud.google.com/tpu/docs/intro-to-tpu

The cloud computing sector is an essential infrastructure for deploying models (Carugati, 2023c). Cloud computing providers and model developers nurture a close, interdependent relationship. Model developers need cloud computing providers to run and deploy their models at scale without investing in the infrastructure. In turn, cloud providers see the model developer as a business driver. Accordingly, some cloud providers have done partnerships with model developers[11]. Partnerships take various forms. Some are exclusive, like the partnership between Microsoft and OpenAI, while others are non-exclusive, like the partnership between Amazon and Anthropic[12]. The partnership generally enables the cloud computing provider to invest in cloud infrastructure. Some providers even develop an infrastructure dedicated to the partners' needs, as Microsoft did by developing a specific computer to run OpenAI models[13]. In exchange, the cloud provider can host, exclusively or non-exclusively, the partner on its cloud service and use the models in related services. For instance, Microsoft exclusively hosts OpenAI models on its Microsoft Cloud Azure and uses OpenAI models on its services, including Office 365, Edge, Bing, and Windows.

The cloud sector is competitive with several global, regional, and national players, including Microsoft, Amazon, Google, OVH, Orange, and Scaleway. However, the sector is under intense scrutiny by competition authorities worldwide, including South Korea, the Netherlands, Japan, France, the United Kingdom, the United States, and Spain[14]. They are concerned with a trend towards concentration in the hands of a few global hyperscalers due to their scale and investment capabilities, which include Amazon, Microsoft, and Google. Those hyperscalers are also the main partners with model developers. As the demand for GenAI increases, the partnership might intensify the trend towards concertation. They are also concerned with potential competition issues arising from barriers to switching, like data transfer fees, software licensing practices, and interoperability, that make it more difficult for a customer to change a cloud provider. Partnerships with hyperscalers might raise additional competition issues from vertical integration, like tying and self-preferencing.

[11] Microsoft and OpenAI Extend Partnership, *Microsoft Blog*, 23 January 2023 (accessed 22 January 2024). Available at: https://blogs.microsoft.com/blog/2023/01/23/microsoftandopenaiextendpartnership/

[12] Expanding Access to Safer AI With Amazon, *Anthropic*, 25 September 2023 (accessed 22 January 2024). Available at https://www.anthropic.com/news/anthropic-amazon

[13] Jennifer Langston, Microsoft Announces New Supercomputer, Lays Out Vision for Future AI Work, *Microsoft Blog*, 19 May 2020 (accessed 22 January 2024). Available at: https://news.microsoft.com/source/features/ai/openai-azure-supercomputer/

[14] Christophe Carugati, Competition Authorities Are Studying Similar Digital Markets, *Digital Competition*, 10 January 2024 (accessed 19 January 2024). Available at: https://www.digital-competition.com/infographics/competition-authorities-are-studying-similar-digital-markets

## 2.2 Machine Learning Models

Machine learning models derive output from input data, such as text, image, video, or music. Models are either closed or open-source models. Developers of closed-source models might license the use of their models to third parties, allowing them to develop commercial applications. By contrast, developers of open-source models make them publicly available for free for research and/or commercial use. They might release various model elements, including the model and training data. This enables third-party developers to modify the model. While some developers might improve the model, others might revise the model for malicious use (OECD, 2023).

In addition, model developers compete on various factors, including task requirements, language specifications, and model size.

First, models differentiate on the intended generated output. A non-exclusive list of models includes text-to-text models (e.g., OpenAI GPT, Google PaLM, Anthropic Claude), text-to-image models (e.g., OpenAI DALL-E, Google Imagen, Adobe Firefly, Midjourney), text-to video models (e.g., Runway Gen-2, Meta Make-A-Video), and text-to-music models (e.g., Google MusicLM, Meta MusicGen, Stability AI Stable Audio). As of January 2024, the community website Hugging Face counts more than 477 thousand open-source models on its model repository[15].

Second, the language of the training dataset is an important quality parameter of the output. If the training dataset does not contain a given language, it might provide a poor output due to the difficulty deriving output from little or no input data. To address this issue, model developers develop multilingual models (e.g., OpenAI GPT) by training them on input data containing various languages, including English, French, German, and Spanish. Others develop monolingual models for a specific language (e.g., Meta CamemBERT for French). Monolingual models sometimes perform better than multilingual ones (OECD, 2023).

Third, models have different sizes. The size refers to the number of parameters required to adjust the model to provide the appropriate output from input data during the training session. The parameters thus encode the knowledge of the model. (Competition and Markets Authority, 2023). The performance of the model and its cost depend on the number of parameters (Rae et al., 2022). The more parameters, the more the models can learn from datasets. The downside is that more parameters require more data and computing power, thus increasing the model's cost. Large models with a high number of parameters are, therefore,

[15] Models, *Hugging Face* (accessed 22 January 2024). Available at: https://huggingface.co/models

called large language models (LLMs). LLMs can perform various tasks even if the dataset contains general domain data, as they are zero-shot reasoners. They can thus generate output without having specific input on the prompt (Kojima et al., 2023). Some models are also fine-tuned on specific datasets to achieve specific tasks better than LLMs with general domain data, like Meta Code Llama, which generates code. Researchers and developers are already proposing smaller models with fewer parameters to reduce financial and environmental costs; these are thus called small language models (SLMs)[16]. Some SLMs perform similarly to LLMs (Schick and Schütze, 2021). Finally, some models, like Google Gemini Nano, can even run on a device and are thus called edge language models or on-device models. These models can perform on-device tasks offline and do not require cloud computing resources, thus reducing financial costs while ensuring greater privacy as data do not leave devices. (Alizadeh et al., 2024).

However, how competition between models of similar characteristics works deserves in-depth scrutiny with quantitative and qualitative data, especially on model performance and user preference, and is thus out of the scope of this paper.

## 2.3 Data

Finally, model developers require data to run and deploy their models. Data is thus the indispensable input to derive output. As in traditional data-driven markets, the volume (scale), variety (scope), velocity (freshness) and quality of the dataset determine the quality of the generated output (Stucke and Grunes, 2016) (Carugati, 2023b).

Model developers train their models on publicly available data from the internet or open-source repository. As of January 2024, the community website Hugging Face counts more than 99 thousand open-source datasets on its dataset repository[17]. Developers also use proprietary datasets from their own first-party or third-party services, such as data brokers, data marketplaces, and publishers.

In addition, models compete on various dataset factors, including task requirements, language specifications, and domain specificity.

[16] For instance, Meta open-source Llma2 is available in three model sizes, 7 billion, 13 billion, and 70 billion parameters. Introducing Llama 2, *Meta* (accessed 22 January 2024). Available at: https://ai.meta.com/llama/

[17] Datasets, *Hugging Face* (accessed 22 January 2024). Available at: https://huggingface.co/datasets

First, the dataset determines the generated output. There are thus various datasets for various task requirements. A non-exhaustive list of datasets includes text-to-image, text-to-video, and text-to-music datasets.

Second, the language of the training dataset determines the output of multilingual or monolingual models. A non-exhaustive language dataset includes majority languages, such as English, French, German, and Spanish, and minority languages, such as Italian, Greek, and Dutch.

Third, the domain specificity of the dataset is an important quality factor in specifying the intended task. Thus, datasets contain specific data for various intended tasks, such as code, legal, finance and art.

The training dataset has a time limitation, as it only contains data up to a certain date. Model developers can retrain the dataset on updated data. However, retraining the dataset is costly. Model developers can deploy the model on real-time data to deal with this issue. For example, Microsoft Bing generates real-time output from the internet by deploying GPT models on Microsoft Search and Index data[18].

In traditional data-driven markets, data is an important factor of market power (Cabral et al., 2021). Data is a competitive advantage that benefits large online platforms like Alphabet and Meta (Competition and Markets Authority, 2020). Users also benefit from data due to data-driven network effects when the user utility changes with improved learning from data, creating value for users (Gregory et al., 2021). Some models, such as GPT, improve by learning from user dialogue data, suggesting that the more the model has users, the higher the model quality[19].

However, the validity of these claims in model markets deserves in-depth scrutiny. Indeed, no evidence suggests that data is a source of market power or that models developed by large online platforms benefit from data advantages. Indeed, available research suggests that some SLMs, like the Koala model trained on high-quality open-source datasets, perform similarly to

[18] Jordi Ribas, Building the New Bing, *Microsoft Blog*, 21 February 2023 (accessed 19 January 2023). Available at: https://blogs.bing.com/search-quality-insights/february-2023/Building-the-New-Bing

[19] Data Controls FAQ, *OpenAI* (accessed 24 January 2024). Available at: https://help.openai.com/en/articles/7730893-data-controls-faq

LLMs trained on much larger volumes of proprietary datasets[20]. Besides, there is not yet evidence that models developed by large online platforms, such as Google PaLM, Google Gemini, or Meta Llama, outperform models from newcomers, such as OpenAI GPT, Anthropic Claude, Mistral AI MISTRAL 7B. Finally, there is no widely available empirical research on the importance of data-driven network effects on model performance.

## 2.4 Applications

Models enable the development of applications for intended tasks, such as generating text. Model developers either develop their own first-party AI-powered applications (e.g., OpenAI ChatGPT) or enable third-party ones (e.g., Hervey AI).

Then, some applications enable first-party and third-party add-ins that complement the app. For instance, OpenAI ChatGPT allows the development of tailored GPTs dedicated to a specific domain. These complementors are then available on an app store[21].

Applications raise several competition issues at both downstream and upstream levels that deserve in-depth scrutiny.

At the upstream level, model and application developers use a cloud provider that hosts the model. They then become customers of the cloud provider. The latter might develop an infrastructure dedicated to the hosted model. The cloud provider might thus have the ability and an incentive to impose technical and commercial conditions to recover the investment cost. For instance, a condition might make it more difficult to move the model and/or application from one cloud provider to another by limiting interoperability between cloud providers. This condition might negatively impact competition in the cloud sector and reinforce the position of the cloud provider hosting the model.

At the downstream level, some model and application developers might provide other services in several markets. They might have the ability and incentive to integrate their own AI-powered applications with other services. For instance, Google is integrating its AI-powered solutions

[20] Xinyang Geng et al, Koala: A Dialogue Model for Academic Research, *Barkley Artificial Intelligence Research*, 3 April 2023 (accessed 24 January 2024). Available at: https://bair.berkeley.edu/blog/2023/04/03/koala/

[21] Introducing the GPT Store, *OpenAI*, 10 January 2024 (accessed 24 January 2024). Available at: https://openai.com/blog/introducing-the-gpt-store

into Google Search[22], Google Chrome[23], and Google Workplace[24]. Microsoft is doing the same with its AI-powered Copilot applications in its search engine Bing, browser Edge, productivity software Office, and operating system Windows. These vertical integrations pose potential competition issues related to tying, bundling, and self-preferencing, as the firm has the incentive to promote its own services over third-party ones. For instance, Google promotes its AI-powered Search Generative Experience (SGE) in Google Search to generate an answer in direct competition with third-party chatbots, like OpenAI ChatGPT. Vertical integration also poses issues related to refusal to deal. It will be the case if a dominant firm prevents third parties from offering competing services in the dominant market. For instance, a hypothetical problematic scenario would be if Microsoft refuses a third party from providing a competing version of Copilot in Microsoft Windows and Microsoft Office.

Moreover, these new AI-powered applications impact how competition works in several sectors, including advertising and cloud computing, to name a few. For instance, search engines are moving from providing search results with links that redirect to a publisher's website to answer engines that generate answers with citations. When the search engine offers search and answer results, the generated answer might substitute or complement the publisher's content. Whether it leads to substitutability or complementary requires in-depth scrutiny and implies different considerations for publishers and advertisers. In case of substitution, the user will most likely not click on the publisher's website or will click significantly less. In turn, publishers might lose traffic and advertising revenue[25]. Advertisers might then dedicate their spend their advertising on the answer engine, as users will remain on it (Carugati, 2023a).

[22] Elizabeth Reid, Supercharging Search with Generative AI, *Google Blog*, 10 May 2023 (accessed 24 January 2024). Available at: https://blog.google/products/search/generative-ai-search/

[23] Parisa Tabriz, Chrome Is Getting 3 New Generative AI Features, *Google Blog*, 23 January 2024 (accessed 24 January 2024). Available at: https://blog.google/products/chrome/google-chrome-generative-ai-features-january-2024/

[24] Johanna Voolich Wright, A New Era for AI and Google Workspace, *Google Blog*, 14 March 2023 (accessed 24 January 2024). Available at: https://workspace.google.com/blog/product-announcements/generative-ai?hl=en

[25] Keach Hagey, Miles Kruppa, and Alexandra Bruell, News Publishers See Google's AI Search Tool as a Traffic-Destroying Nightmare, *The Wall Street Journal*, 14 December 2023 (accessed 25 January 2024). Available at: https://www.wsj.com/tech/ai/news-publishers-see-googles-ai-search-tool-as-a-traffic-destroying-nightmare-52154074

# 3 Regulatory Instabilities

GenAI raises several regulatory concerns that impact how competition works. The four main concerns are intellectual proprietary rights, data protection, AI governance, and competition[26].

## 3.1 Intellectual Proprietary Rights

Model developers train and deploy their models on public and proprietary datasets, including copyrighted data. Copyright protection requires the data owner's consent to use data, which might decide to license its dataset against a fee.

Around the globe, there have been concerns that some model developers use datasets without consent, leading to litigations in several countries, like the ongoing lawsuit by the New York Times in the US against OpenAI and Microsoft[27]. In this case, OpenAI declined any wrongdoing because it considers that it can fairly use the publisher's content without consent under US copyright law[28]. To address this issue, some model developers, including OpenAI and Google, propose opt-out mechanisms, allowing publishers to proactively block model developers from collecting their content to train their models[29]. Some model developers also conclude partnerships with publishers for the use of their content, like the partnership between OpenAI and Alex Springer[30].

In response, legislators and regulators have proposed several legislative and regulatory initiatives to address copyright concerns, including an obligation to publish summaries of

[26] Christophe Carugati, The Age of Competition in Generative Artificial Intelligence Has Begun, *Bruegel*, 11 May 2023 (accessed 25 January 2024). Available at: https://www.bruegel.org/first-glance/age-competition-generative-artificial-intelligence-has-begun

[27] Michael M. Grynbaum and Ryan Mac, The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work, *The New York Times*, 27 December 2023 (accessed 25 January 2024). Available at: https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html

[28] OpenAI and journalism, *OpenAI*, 8 January 2024 (accessed 25 January 2024). Available at: https://openai.com/blog/openai-and-journalism

[29] GPTBot, *OpenAI* (accessed 25 January 2024). Available at: https://platform.openai.com/docs/gptbot

[30] Axel Springer and OpenAI Partner to Deepen Beneficial Use of AI in Journalism, *Alex Springer*, 13 December 2023 (accessed 25 January 2024). Available at: https://www.axelspringer.com/en/ax-press-release/axel-springer-and-openai-partner-to-deepen-beneficial-use-of-ai-in-journalism

copyrighted data used for training models in the forthcoming European AI Act[31], a code of practice on copyright and AI in the UK[32], workshops with content creators in the US[33], and a proposal to amend the EU copyright directive by some French politicians[34].

These regulatory developments directly impact GenAI, as data access is primordial to the development of models. OpenAI even argued that prohibiting using copyrighted data to train models would prevent GenAI development[35].

## 3.2 Data Protection

Training datasets contain personal and non-personal data. The collection and use of personal data raises data protection and privacy concerns, especially regarding the user's consent.

Data protection authorities worldwide are increasingly looking at how models use personal data and their implications in terms of regulatory requirements, like the ongoing consultation on generative AI and data protection in the UK[36]. Some regulators even took enforcement actions by preventing the use of AI-powered applications due to alleged data protection infringements, like the temporary ban of ChatGPT in Italy in April 2023[37].

[31] EU AI Act: First Regulation on Artificial Intelligence, *European Parliament*, 19 December 2023 (accessed 25 January 2024). Available at: https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

[32] The Government's Code of Practice on Copyright and AI, *UK Government*, 29 June 2023 (accessed 25 January 2024). Available at: https://www.gov.uk/guidance/the-governments-code-of-practice-on-copyright-and-ai

[33] FTC Staff Report Details Key Takeaways from AI and Creative Fields Panel Discussion, *Federal Trade Commission*, 18 December 2023 (accessed 25 January 2024). Available at: https://www.ftc.gov/news-events/news/press-releases/2023/12/ftc-staff-report-details-key-takeaways-ai-creative-fields-panel-discussion

[34] Théophane Hartmann, French MPs Want to Amend EU's Copyright Rules to Cover Generative AI, *Euractiv*, 20 January 2024 (accessed 25 January 2024). Available at: https://www.euractiv.com/section/artificial-intelligence/news/french-mps-want-to-amend-eus-copyright-rules-to-cover-generative-ai/

[35] James Titcomb and James Warrington, OpenAI Warns Copyright Crackdown Could Doom ChatGPT, *The Telegraph*, 7 January 2024 (accessed 25 January 2024). Available at: https://www.telegraph.co.uk/business/2024/01/07/openai-warns-copyright-crackdown-could-doom-chatgpt/

[36] ICO Consultation Series on Generative AI and Data Protection, *Information Commissioner's Office*, 15 January 2024 (accessed 25 January 2024). Available at: https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-consultation-series-on-generative-ai-and-data-protection/

[37] Luca Bertuzzi, Italian Data Protection Authority Bans ChatGPT Citing Privacy Violations, *Euractiv*, 13 April 2023 (accessed 25 January 2024). Available at: https://www.euractiv.com/section/artificial-intelligence/news/italian-data-protection-authority-bans-chatgpt-citing-privacy-violations/

These regulatory developments also impact GenAI, as data protection requirements have important implications for the lawfulness of the training datasets. Non-compliance could potentially result in bans and fines.

## 3.3 AI Governance

AI models give rise to new governance issues, such as addressing risks associated with their use, including manipulation and biometric identification. In this context, AI governance is the priority of legislators around the globe. As of January 2024, the OECD counts 646 legislative initiatives worldwide about AI governance only[38]. The forthcoming European AI Act, which is still under legislative process after a political agreement in December 2023, is just one of them[39]. However, the AI Act will likely have far-reaching implications in other jurisdictions worldwide, as Europe is often the global rule-setter (Bradford, 2020). The proposed text includes specific provisions concerning GenAI and high-impact General Purpose AI (GPAI) models, identified as posing systematic risks. According to the latest officially available information, developers of such models must disclose to users that the content is AI-generated, design the model to prevent the generation of illegal content, and publish summaries of copyrighted data used for training. High-risk GPAI, which might only apply to OpenAI GPT-4, is required to conduct impact assessments of risks and report them to the European Commission[40]. The latter announced in January 2024 a dedicated AI office within the Commission, tasked with coordinating AI policy at the EU level and overseeing the AI Act[41].

These regulatory developments significantly impact competition. Firstly, certain models classified as high-risk GPAI face more extensive regulatory compliance requirements than models that potentially pose similar risks. The question of whether this regulatory burden will place the former at a competitive disadvantage compared to the latter merits closer scrutiny. As of January 2024, European legislators have not disclosed any impact assessment of the provisions affecting high-risk GPAI models on competition. This is particularly concerning as legislative debates indicated a desire to promote European models such as the French Mistal

*See* also, ChatGPT: Italian DPA Notifies Breaches of Privacy Law to OpenAI, *Garante per la protezione dei dati personali*, 29 January 2024 (accessed 1st February 2024). Available at: https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9978020

[38] National AI Policies & Strategies, *OECD.AI Policy Observatory* (accessed 25 January 2024). Available at: https://oecd.ai/en/dashboards/overview/policy

[39] Artificial Intelligence Act: Council and Parliament Strike a Deal on the First Rules for AI in the World, *European Council,* 9 December 2023 (accessed 25 January 2024). Available at: https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/

[40] EU AI Act: First Regulation on Artificial Intelligence, *European Parliament*, 19 December 2023 (accessed 25 January 2024).

[41] Commission Decision Establishing the European AI Office, *European Commission*, 24 January 2024.

AI and German Aleph Alpha by excluding them from regulatory burdens[42]. In other words, the provision might be driven by a desire to achieve an industrial policy goal of promoting European firms rather than protecting users from all AI risks, irrespective of the model size. Secondly, the proliferation of regulatory initiatives may result in regulatory inconsistency, leading to increased compliance costs and regulatory burdens. Consequently, some model developers might encounter challenges in scaling and competing, especially when compared to developers with greater compliance resources and the ability to benefit from economies of scale in regulatory compliance.

## 3.4 Competition

As GenAI is an emerging technology, firms vigorously compete along all the above value chains, from computing resources and models to data. Yet, some competition authorities have already voiced concerns that GenAI might be concentrated in the hands of a few large online platforms with access to computing resources, models, and data[43]. Moreover, they outline that some business practices, such as tying, bundling, exclusive dealing, and self-preferencing, might give rise to potential competitive concerns[44]. Competition authorities also closely monitor partnerships between large cloud providers and model developers. In Germany[45], the UK[46], and Europe[47], competition authorities investigate whether the partnership between Microsoft and OpenAI requires an obligation to review the transaction under their national merger control laws. If Microsoft/OpenAI undergoes a merger review, competition authorities could potentially accept or block the transaction or impose behavioural and structural conditions on how Microsoft should operate with OpenAI products and services. In the US, the Federal Trade

[42] Théophane Hartmann, Behind France's Stance Against Regulating Powerful AI Models, *Euractiv*, 29 November 2023 (accessed 25 January 2024). Available at: https://www.euractiv.com/section/artificial-intelligence/news/behind-frances-stance-against-regulating-powerful-ai-models/

[43] See in the US, Generative AI Raises Competition Concerns, *Federal Trade Commission*, 29 June 2023 (accessed 25 January 2024). Available at: https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns

[44] G7 Competition Authorities and Policymakers' Summit Digital Competition Communiqué 2023, *G7*, 8 November 2023 (accessed 25 January 2023). Available at: https://www.bundeskartellamt.de/SharedDocs/Publikation/EN/Others/G7_2023_Communique.pdf?__blob=publicationFile&v=2

[45] Cooperation Between Microsoft and OpenAI Currently Not Subject to Merger Control, *Bundeskartellamt*, 15 November 2023 (accessed 26 January 2024). Available at: https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2023/15_11_2023_Microsoft_OpenAI.html

[46] CMA Seeks Views on Microsoft's Partnership with OpenAI, *Competition and Markets Authority*, 8 December 2023 (accessed 26 January 2024). Available at: https://www.gov.uk/government/news/cma-seeks-views-on-microsofts-partnership-with-openai

[47] Commission Launches Calls for Contributions on Competition in Virtual Worlds and Generative AI*, European Commission*, 9 January 2024 (accessed 26 January 2024).

Commission even launched a sector inquiry into these partnerships, requesting detailed information from Alphabet, Amazon, Anthropic Microsoft, and OpenAI on the rationale and impact of the partnerships on competition[48].

Besides, GenAI spurs innovation in several sectors, potentially disrupting current markets and creating new ones, such as answer engines replacing search engines (Carugati, 2023a). As of January 2024, competition authorities have not yet launched market studies into generative AI and its impact on specific markets, such as advertising or cloud computing. However, as noted in the introduction, they announced sector inquiries into generative AI and competition.

These regulatory initiatives will inform how competition in GenAI works. They are not formal investigations into non-compliance with national competition laws or specific digital markets regulations, like the European Digital Markets Act (DMA). However, the findings will likely influence GenAI developments in delivering positive outcomes on competition. Competition authorities have already warned that they will intervene with formal enforcement powers where necessary (Competition and Markets Authority, 2023).

# 4 Policy Recommendations

The paper finds that GenAI leads to emerging markets and technologies in a context of regulatory instabilities in various jurisdictions and legal regimes. Against this background, competition authorities worldwide should follow the below policy recommendations.

First, competition authorities should cooperate in an international forum to ensure international coherence. They should do joint studies in a forum like the European Competition Network (ECN) or International Competition Network (ICN) to foster experience-sharing without resource duplication, given the borderless nature of the issues posed by GenAI.

Second, competition authorities should undertake in-depth studies of some critical elements of the value chain and markets. They should do priority inquiries into graphic cards and cloud computing sectors as GenAI developments depend on them. Market characteristics and business practices in these sectors might impact competition in the long term. Competition authorities should also closely monitor how GenAI impact competition in several important

[48] FTC Launches Inquiry into Generative AI Investments and Partnerships, *Federal Trade Commission*, 25 January 2024 (accessed 26 January 2024). Available at: https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships

markets, including search engines and online advertising, given their importance to content creators.

Thirdly, competition authorities should collaborate with relevant competent authorities to examine the impact of various legal regimes on competition. Considering the interactions between competition and other legal frameworks, they should ideally produce joint studies or, at the very least, joint statements addressing data protection, intellectual property rights, AI governance, and regulations in digital markets. The outcomes of these collaborations should contribute to greater regulatory stability, providing market actors with the assurance that GenAI can deliver its full benefits responsibly.

Last but not least, competition authorities should exercise formal enforcement powers and potentially update competition tools only when necessary and justified. They should do so only after evidence of enforcement gaps following market studies. They should resist the call for quick intervention to avoid critics of underenforcement in digital markets.

# References

Alizadeh, K., I. Mirzadeh, D. Belenko, K. Khatamifard, M. Cho, C. C. Del Mundo, et al. (2024) 'LLM in a flash: Efficient Large Language Model Inference with Limited Memory'

Autoridade da Concorrência (2023) 'Competition and Generative AI Intelligence'

Bradford, A. (2020) 'The Brussels Effect: How the European Union Rules the World', *Faculty Books*

Cabral, L., J. Haucap, G. Parker, G. Petropoulos, T. Valletti, and M. Van Alstyne (2021) 'The EU Digital Markets Act', *JRC Publications Repository*

Carugati, C. (2023a) 'Antitrust Issues Raised by Answer Engines', *Bruegel*

Carugati, C. (2023b) 'Competition in Generative Artificial Intelligence Foundation Models', *Bruegel*

Carugati, C. (2023c) 'The Competitive Relationship between Cloud Computing and Generative' AI, *Bruegel*

Competition and Markets Authority (2020) 'Online Platforms and Digital Advertising Market Study Final Report'

Competition and Markets Authority (2023) 'AI Foundation Models: Initial Report'

Gregory, R. W., O. Henfridsson, E. Kaganer, and H. Kyriakou (2021) 'The Role of Artificial Intelligence and Data Network Effects for Creating User Value', *Academy of Management Review*, 46:3, 534–51

Kojima, T., S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa (2023) 'Large Language Models are Zero-Shot Reasoners'

OECD (2023) 'AI Language Models: Technological, Socio-Economic and Policy Considerations'

Rae, J. W., S. Borgeaud, T. Cai, K. Millican, J. Hoffmann, F. Song, et al. (2022) 'Scaling Language Models: Methods, Analysis & Insights from Training Gopher'

Schick, T., and H. Schütze (2021) 'It's Not Just Size That Matters: Small Language Models Are Also Few-Shot Learners', in K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, et al. (eds), *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (presented at the NAACL-HLT 2021, Online: Association for Computational Linguistics), pp. 2339–52

Stucke, M., and A. Grunes (2016) 'Big Data and Competition Policy', *Oxford University Press*