Generative artificial intelligence: the Autorité issues its opinion on the competitive functioning of the sector

Published on June 28, 2024

Background

Since the public release of the ChatGPT chatbot (created by OpenAI) in November 2022, generative artificial intelligence (hereafter "AI") has taken centre stage in public and economic debate. The questions raised by generative AI range from ethics and respect for intellectual property to the impact on the labour market and productivity. The technology offers numerous possibilities to companies in terms, for example, of content creation, graphic design, employee collaboration and customer service.

The benefits of generative AI will only materialise if all households and companies have access to a variety of different models adapted to their needs. Competition in the sector must therefore be conducive to innovation and allow for the presence of multiple operators.

Against this backdrop, the *Autorité de la concurrence* decided on 8 February 2024 to start inquiries *ex officio* into the competitive functioning of the generative AI sector and to launch a public consultation. As part of this consultation, views from around 40 parties and 10 stakeholder associations were collected.

Scope

This opinion aims to provide stakeholders with a competitive analysis of the fast-growing generative AI sector, with a particular focus on the strategies implemented by major digital companies aimed at consolidating their market power upstream in the generative AI value chain (i.e. the design, training and fine-

tuning of large language models [LLMs]) or at leveraging this market power in order to expand in this booming sector. The *Autorité* looks in particular at practices implemented by operators already present in cloud infrastructure and at issues relating to access to cloud infrastructure, computing power, data and a skilled workforce. The *Autorité* also examines investments and partnerships by major digital companies, in particular in innovative companies specialised in generative AI.

Accordingly, the *Autorité* only incidentally addresses the practices implemented by operators downstream in the value chain (i.e. in contact with the end consumer) and does not touch on the consequences of AI for the competitive functioning of the economy as a whole – an issue of major importance that will merit further analysis in the future.

Recommendations that require no new legislative initiative

The *Autorité* makes a number of recommendations aimed at boosting competition in the sector:

- with no change to existing legislation, make the regulatory framework applicable to the sector more effective;

- in the event of harm to competition, use the rapid and effective tools of competition law and the law on restrictive competitive practices;

- foster innovation by ensuring better access to computing power;

- ensure a balance between fair remuneration for rights holders and access for model developers to the data needed to innovate;

- ensure greater transparency on investments by digital giants.

The generative AI sector

Definition

According to the European Parliament, artificial intelligence refers to any tool used by a machine "*to display human-like capabilities such as reasoning, learning, planning and creativity*". Generative AI refers to AI models capable of generating new content such as text, image, sound or video.

A growing priority for public authorities

The generative AI sector is attracting growing interest around the world.

In France, the government launched a national AI strategy in 2018, for which almost €2.5 billion of the "France 2030" plan has been earmarked. In March 2024, the French AI Commission (*Commission de l'IA*) launched by the Prime Minister presented a report containing 25 recommendations, suggesting in particular to make France a major centre for computing power, to facilitate data access and to establish global AI governance.

At European level, most of the provisions of the AI Act (which will soon be published in the EU Official Journal) will be applicable from 2026. Although published before the rise of generative AI, the Digital Markets Act (DMA) and the Data Act will have an impact on the sector.

A series of initiatives on AI have been adopted globally, such as the Bletchley Declaration in the United Kingdom in November 2023 at the AI Safety Summit. The next global summit will take place in France on 10 and 11 February 2025. Other initiatives have been taken by the G7, the United States, the United Kingdom and China, for example.

How the sector works

There are two key phases in generative AI modelling:

- **training**: the initial learning process of a model (often called "foundation model", which includes LLMs), during which its parameters, known as "weights", are determined. Training requires both significant computing power and a large volume of – generally public – data. The training phase may be followed by fine-tuning, during which the model is adapted to a specific task or a specialised dataset (e.g. legal or health-related data). Fine-

tuning is generally based on a smaller, proprietary dataset and may involve human expertise;
- **inference**: the use of the trained model to generate content. The model can be made accessible to users via specific applications, such as Open AI's ChatGPT or Mistral AI's Le Chat, or APIs for developers. The computing power required depends on the number of users. Unlike many digital services, the marginal cost of generative AI is not negligible, given the cost of the computing power required. New data that was not used for training may be added during the inference phase, in order to ground the model in recent data, such as news articles.

The participants in the value chain

The operators in the generative AI value chain are:

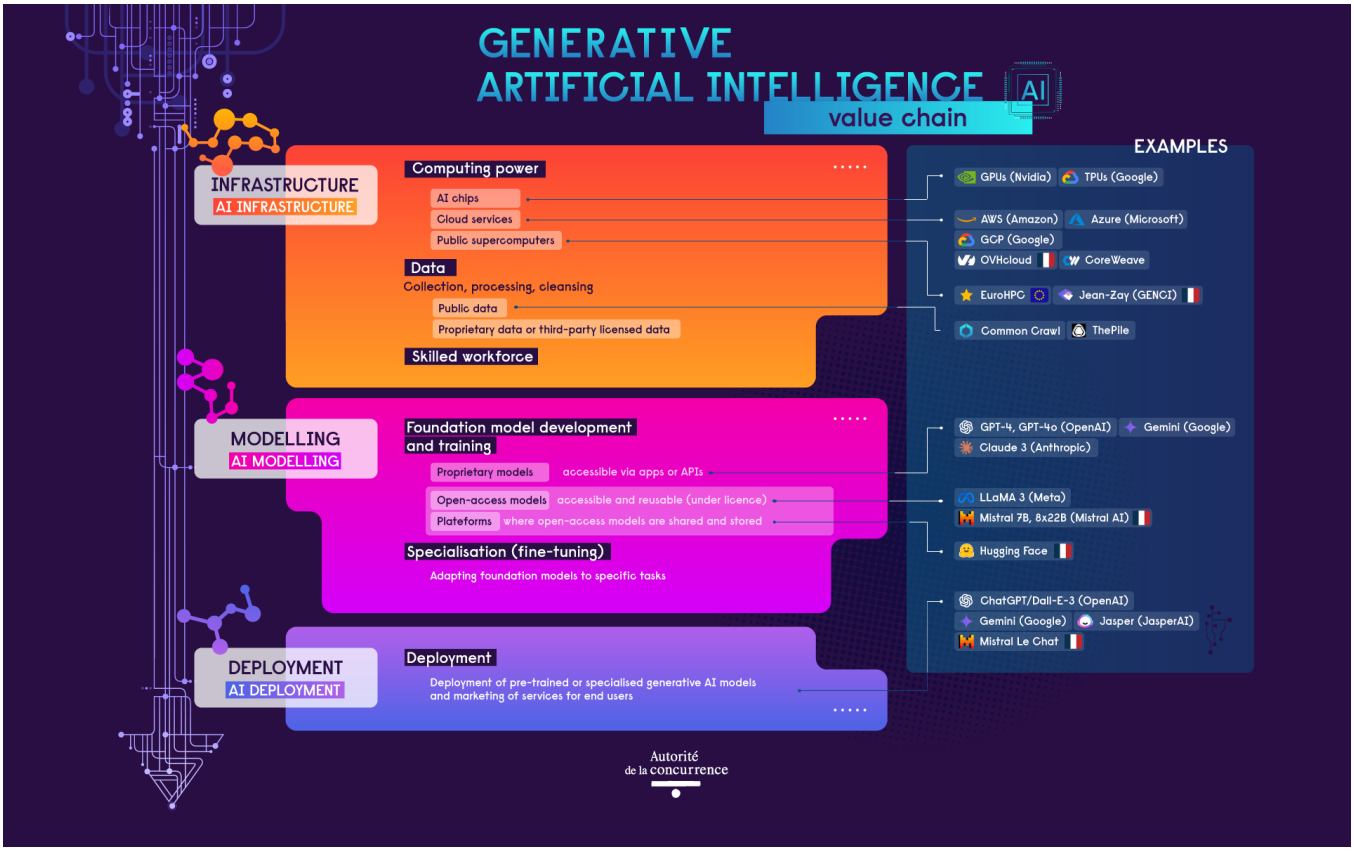
- **major digital companies**: Alphabet and Microsoft are present across the entire value chain, while Amazon, Apple, Meta and Nvidia are present only at certain specific layers;
- **model developers**: for example, start-ups or AI-focused research labs, such as Anthropic, Hugging Face, Mistral AI and OpenAI. They have often formed partnerships with one or more digital giants, such as OpenAI with Microsoft and Anthropic with Amazon and Google. For the distribution of their models, they may adopt either a proprietary or open-source approach.

At the upstream level, several types of operators are involved:

- **IT component suppliers**, such as Nvidia, develop graphics processing units (GPUs) and AI accelerators, which are essential components for training generative AI models;
- **cloud service providers**, including digital giants, known as "hyperscalers", such as Amazon Web Services (AWS), Google Cloud Platform (GCP) and Microsoft Azure, cloud providers such as OVHCloud, as well as specialist AI providers such as CoreWeave. The necessary computing resources may also be provided by **public supercomputers** (such as Jean Zay in France).

At the downstream level, many operators are marketing new services based on generative AI to the general public (like ChatGPT), companies and public

authorities and/or integrating generative AI into their existing services (like Zoom).



Source: Autorité de la concurrence, inspired by ChatGPT, Bard & Co.: An introduction to AI for competition and regulatory lawyers by Thomas Höppner and Luke Streatfeild, 23 February 2023.

The competitive functioning of generative AI

High barriers to entry

1. The need for specialised AI chips

Access to sufficient computing power for performing a large number of operations in parallel, and with the high precision needed to determine several billion parameters, is essential for developing foundation models. The GPUs developed by Nvidia (combined with its CUDA software) or the AI accelerators developed by major digital companies (such as the tensor processing units [TPUs] developed by Google) are essential for the training, fine-tuning and inference of generative AI models. They are also very expensive. Since 2023, the sector has experienced shortages due to an explosion in demand.

2. The importance of cloud services

The cloud appears to be the only way to access the computing power needed to train models, and is also a vector for distributing models downstream on marketplaces. Such marketplaces enable developers to make their models easily accessible to cloud using companies, encouraging developers to make their models available on every cloud provider.

3. The need for large volumes of data

In the current state of generative AI technology based on LLMs, data is an essential input for operators in the market. This data – which can be text, image or video – is mainly obtained from publicly-accessible sources, such as web pages, or datasets like the Common Crawl web archive (an organisation that has been providing free data from the Internet since 2008). The cleansing and processing of this data is a differentiating factor, as operators need to filter the data in order to keep only qualitative content.

The stakeholders consulted as part of this opinion expressed concerns about data access, in light of the fear that publicly-available data will not be sufficient in the future and legal uncertainties linked to the actions brought by several rights holders, such as the complaint filed by the New York Times against OpenAI and Microsoft.

4. Rare, highly sought-after skills

Model developers must have advanced data science skills, for example in machine learning and deep learning, in order to be able to both develop the

code and set up the architecture to run this code efficiently. In addition to their theoretical training, engineers must have empirical skills that can only be acquired by working with the models.

5. Significant financing needs

The scale of the investments required creates significant barriers to entry, especially as repeat investments are required. The *Autorité* notes that investments in the sector increased almost six-fold between 2022 and 2023. Companies in the sector raised over $22 billion in 2023 (around €20 billion), compared with around $4 billion in 2022 (around €3.7 billion).

Technical and organisational developments and certain public policies may limit the barriers to entry:

- **the existence of public supercomputers**, which can be used free of charge in exchange for a contribution to open science, enabling a wide range of operators, particularly academics but also private operators, to access computing power for training or fine-tuning generative AI models;
- **the emergence of technological innovations that reduce the need for computing power and data**, such as smaller models or synthetic data (generated by AI) that can partially replace real data and reduce the risk of using personal data;
- **the existence of open-source models**, which can be reused or fine-tuned by other operators.

The position of certain operators in other markets linked to generative AI could give rise to a range of competitive advantages

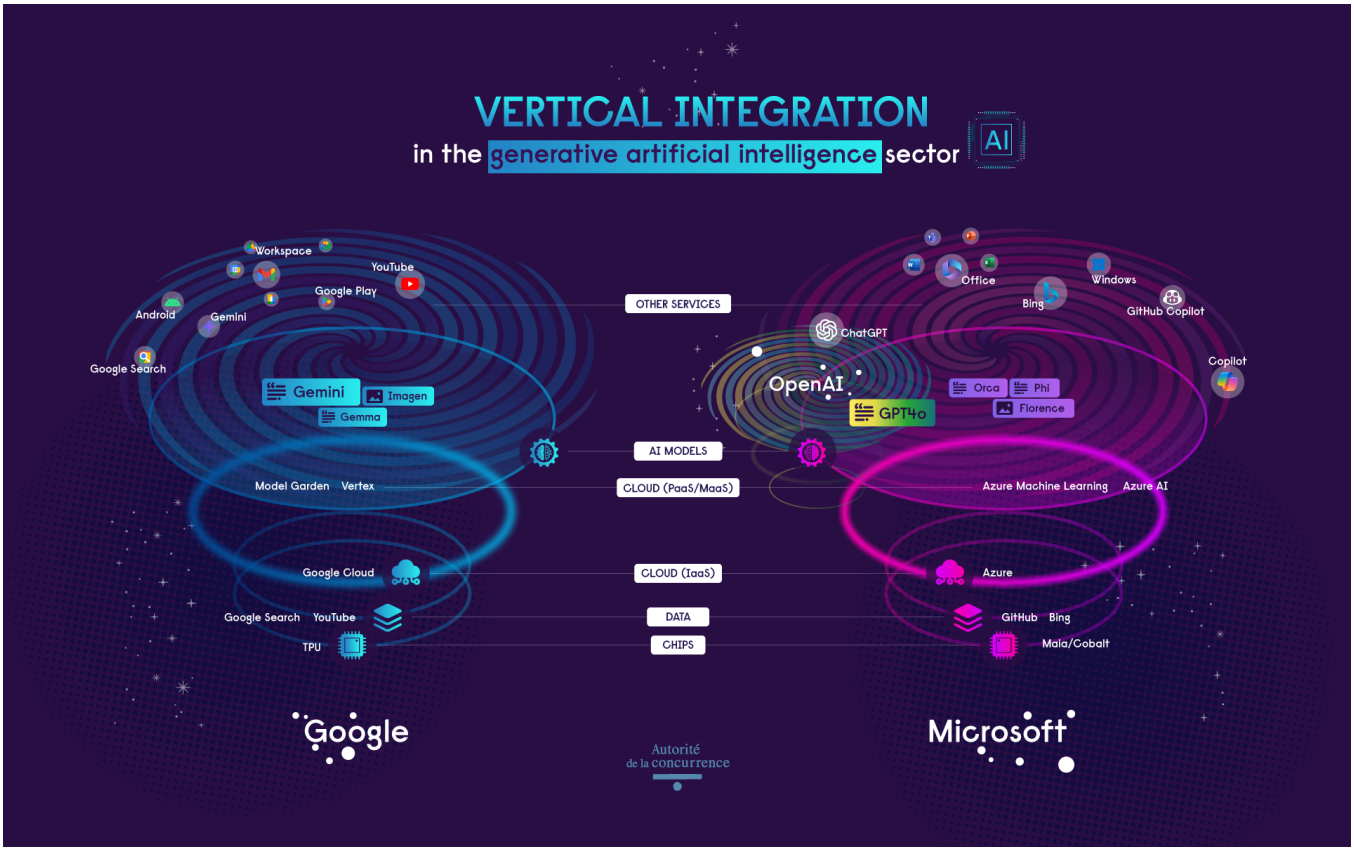
1. Preferential access to inputs

Major digital companies enjoy preferential access to the inputs needed to train and develop foundation models. Developers of competing foundation models, which do not have access to these inputs under the same conditions, cannot easily replicate these advantages.

- They have **easier access to computing power** as partners and competitors of AI chip suppliers. On the one hand, they are able to buy in large quantities and negotiate preferential agreements with GPU suppliers like Nvidia. On the other hand, most of them are also developing in-house AI accelerators specifically tailored to their ecosystems, such as Google's TPUs and AWS' Trainium. Alternatives to Nvidia's CUDA software are also beginning to emerge.
- They also enjoy **preferential access to large volumes of data** (as an example, YouTube provides Alphabet with a major source of training data for AI models). They can also access data associated with the use of their internal services, as well as use their financial power to enter into agreements with the owners of third-party data, as demonstrated by Google's agreement to pay $60 million (around €55 million) a year for access to data from Reddit, a US social news aggregation and forum social network.
- In addition, **many highly-skilled employees are enticed by the attractive salaries** and employment conditions offered by major digital companies.

2. The benefits of vertical and conglomerate integration

In addition to unrivalled access to the inputs needed to train generative AI models, major digital companies enjoy advantages linked to their vertical and conglomerate integration, such as **cumulative** economies of scale and scope and network effects, with feedback data from users being used to refine future models and improve performance or offer new services.



The *Autorité* also found that major digital companies are starting to integrate generative AI tools into their product and service ecosystems. For example, Microsoft deploys its own models and those of its partner OpenAI in the "Copilot" function to enhance Microsoft Bing's search functionality and offers an AI assistant designed to work with the Microsoft 365 offering. In addition, major digital companies' marketplaces (Model-as-a-Service [MaaS]) provide access to proprietary and third-party generative AI models designed to run in their ecosystems.

Thanks to their access to the key inputs for the development of generative AI, major digital companies have a huge advantage over their competitors. The advantage is reinforced by their integration across the entire value chain and in related markets, which not only generates economies of scale and scope, but also guarantees access to a critical mass of users.

Competition risks upstream in the value chain

1. The risk of abuse by chip providers

The *Autorité* found a number of potential risks, such as price fixing, production restrictions, unfair contractual conditions and discriminatory behaviour. Concern was also expressed regarding the sector's dependence on Nvidia's CUDA chip programming software (the only one that is 100% compatible with the GPUs that have become essential for accelerated computing). Recent announcements of Nvidia's investments in AI-focused cloud service providers such as CoreWeave are also raising concerns.

The graphics card sector, which was the target of an unannounced inspection in September 2023, is being closely scrutinised by the *Autorité*'s Investigation Services.

2. The risk of lock-in by major cloud service providers

The *Autorité* found that several financial and technical lock-in practices, already identified in Opinion 23-A-08 on competition in the cloud sector, appear to remain and even to be intensifying to attract the largest possible number of start-ups active in the generative AI sector. In addition to the particularly high levels of cloud credits offered to innovative companies in the sector, the *Autorité* identified a number of technical lock-in practices (such as barriers to migration).

The *Autorité* recalls that such practices could be assessed, in particular, on the basis of abuse of dominant position. Some of the practices are also governed by French law 2024-449 of 21 May 2024 to secure and regulate the digital space (known as the "SREN Law"), on which the *Autorité* issued an opinion, or by the EU Data Act.

3. Risks related to data access

Innovative companies in the sector may be confronted with practices of refusal of (or discriminatory) access to data across the entire value chain.

In addition, agreements under which major digital companies impose exclusive access to content creators' data, or pay them substantial remuneration that is difficult for their competitors to replicate, could constitute anticompetitive

practices (cartels or abuse).

Access to user data is also a major challenge. Several stakeholders reported that major companies in the sector continue to use various strategies to restrict third-party access to their users' data, by abusing legal rules, such as personal data protection, or security concerns.

Lastly, content publishers are very concerned about the use of their content by foundation model providers without the authorisation of rights holders. The recent decision by the ***Autorité*** in the "related rights" case established that Google had used content from press agencies and publishers to train its foundation model Gemini (formerly "Bard"), without notifying them and without giving them an effective possibility to opt-out. While this question raises issues relating to the enforcement of intellectual property rights that go beyond the scope of this opinion, competition law could, in principle, address these issues based on an infringement of fair trading, for example, and therefore, exploitative abuse.

4. Risks related to access to a skilled workforce

In addition to wage-fixing agreements, no-poach agreements may also constitute prohibited anticompetitive practices.

An additional area of concern is the recruitment by digital giants of entire teams (such as Microsoft's hiring of most of start-up Inflection's 70-person staff) or strategic employees of model developers (such as Microsoft's brief recruitment of Sam Altman, the founder of OpenAI, before he was eventually hired back by OpenAI). While this type of practice may be examined under merger control rules, it can also be analysed as an attempt to exclude competitors from the sector.

5. Risks associated with open-access models

While open-access models can help to lower barriers to entry, they can also raise competition concerns. In some cases, the conditions of access and reuse of models or some of their components can lead to users being locked-in.

6. Risks associated with the presence of companies on several markets

The vertical integration of certain digital operators and their service ecosystems may give rise to a number of abusive practices.

At the upstream level, model developers could be **denied or given limited access to the chips or data needed to train competing foundation models.** This practice could lead to delays or the introduction of less ambitious models, thereby undermining effective competition in the market. Several stakeholders are also concerned about existing **exclusivity agreements** between cloud service providers and the leading foundation model developers.

Other risks arise from the downstream use of generative AI models, through **practices of tying**. Companies holding pre-eminent or dominant positions in AI-related markets could tie the sale of products or services to that of their own AI solutions. **In particular, the integration of generative AI tools on certain devices, such as smartphones, is raising concerns.** This type of practice could permanently consolidate the generative AI sector around already dominant digital companies.

Downstream competitors could also be harmed by **self-preferencing** practices of vertically integrated operators, affecting the ability of developers of non-vertically integrated models to compete with those operators.

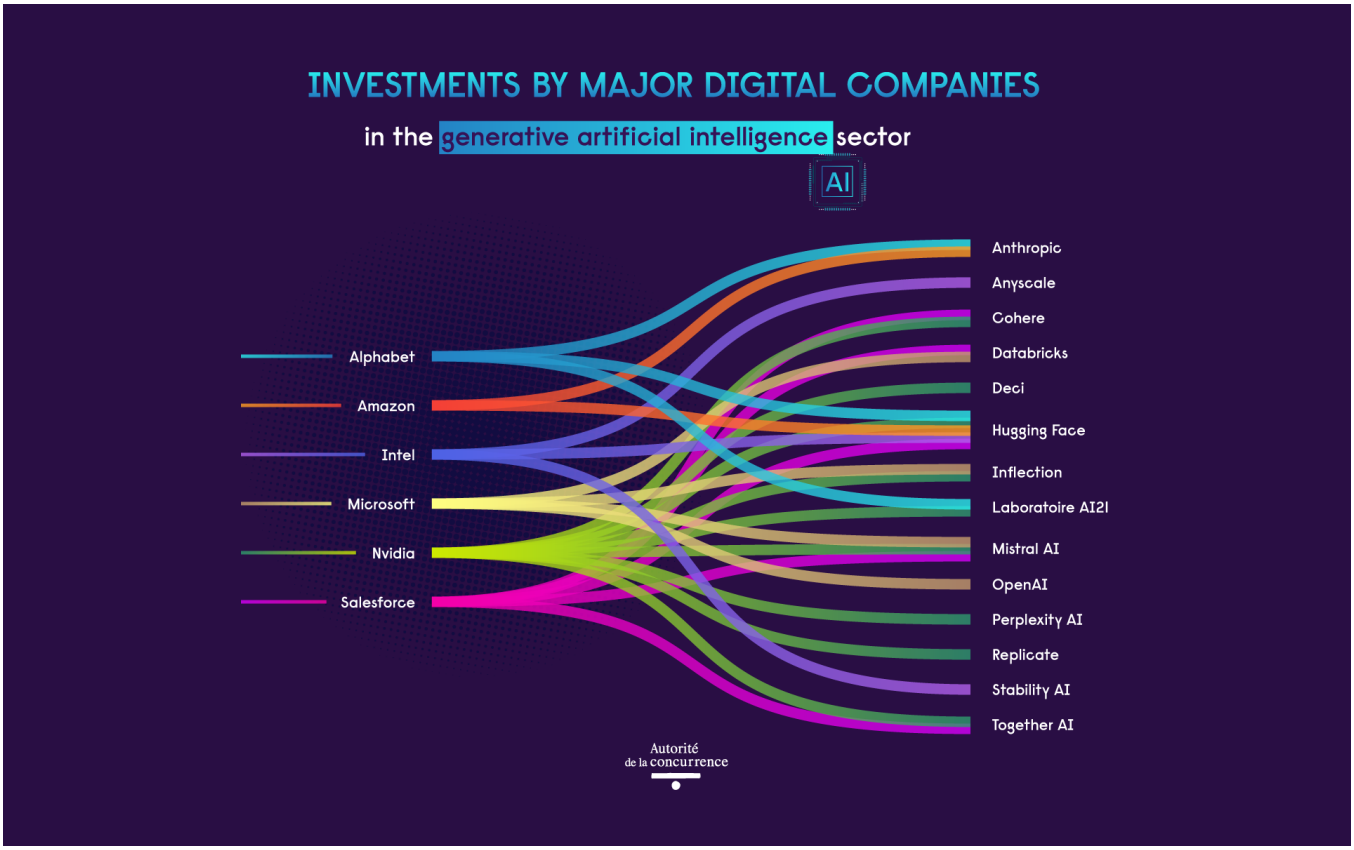
7. Risks associated with minority investments and partnerships by digital giants

Investments and partnerships between stakeholders can give start-ups in the sector the opportunity to benefit from the financial and technological resources of major companies, and thus foster innovation. For the buyer, such investments enable diversification or access to innovative technologies to improve the quality of its services. For example, Microsoft has entered into an exclusive partnership with OpenAI in the form of a multi-year investment.

Nevertheless, they present significant risks that call for particular vigilance by competition authorities. They may **weaken competition** between the two entities, lead to **vertical effects**, **increase market transparency** or **lock-in** some parties.

Minority investments by major companies may be assessed by competition authorities on several legal grounds, for example under merger control rules or competition law.

However, the *Autorité* notes **a lack of transparency** in agreements, which can make it difficult to determine whether they are likely to harm competition and hence consumers. These concerns are shared by competition authorities around the world, as evidenced by ongoing investigations into Alphabet, Amazon, Anthropic, Microsoft and OpenAI.



Source: Autorité de la concurrence, *inspired by S&P Global, Untangling the web of strategic tech investments in generative AI, 22 February 2024.*

8. Risk of collusion between companies in the sector

While almost all the stakeholders consulted during the public consultation did not express any specific concerns on this issue, the use of generative AI could potentially give rise to the concerted practices that were the subject of a joint study in 2019 by the *Autorité* and the German *Bundeskartellamt*, such as the parallel use of separate individual algorithms or the use of machine learning algorithms. Here too, vigilance is essential.

The recommendations made by the *Autorité*

Make the regulatory framework applicable to the sector more effective

The Commission should pay particular attention to the development of services that give access to generative AI models in the cloud (MaaS) and assess the possibility of designating companies providing such services as gatekeepers specifically for those services, under the DMA. Some of the problematic behaviours identified above would therefore be prohibited *ex ante*.

> Proposal no. 1: the Commission should pay particular attention to the development of MaaS services to assess the possibility of designating companies providing such services as gatekeepers under the DMA.

In addition, at the French level, the *Autorité* encourages the Directorate General for Competition Policy, Consumer Affairs and Fraud Control (DGCCRF) to pay particular attention to the use of cloud credits in AI, in particular as part of the implementation of the SREN Law.

Proposal no. 2: at the French level, in implementing the provisions of the SREN Law on cloud credits, the DGCCRF should pay particular attention to the use of such credits in AI.

The *Autorité* calls for vigilance regarding the effects of the AI Act on competition in the sector.

Proposal no. 3: the future AI Office, established under Article 64 of the AI Act, and the competent national authority in France, which will be designated in accordance with Article 70 of said Act, should ensure on the one hand that the implementation of the Act does not hinder the emergence or expansion of smaller operators, and on the other hand that the largest operators in the sector do not misuse the text to their advantage.

International coordination is necessary, given the various initiatives underway in France, Europe and the rest of the world, to ensure that such initiatives do not create distortions and additional costs for companies. The AI Summit to be hosted by France in February 2025 will be an opportunity to strengthen global AI governance.

Use the full extent of competition law tools

The *Autorité* also calls for the support of the relevant authorities and for the use of all available tools. The *Autorité* will remain vigilant in the generative AI sector,

alongside the DGCCRF, in order to use all their respective tools, if necessary, to act swiftly and effectively.

Proposal no. 4: the authorities responsible for enforcing competition in the markets must remain vigilant in the generative AI sector and, if necessary, use all the tools at their disposal to act swiftly and effectively.

Increase access to computing power

Like many public authorities, the *Autorité* supports the development of public supercomputers, which are an alternative to cloud providers and give academics, in particular, access to computing power, which is beneficial for innovation. The *Autorité* is also in favour of opening supercomputers to private operators, under certain conditions, for a fee.

Proposal no. 5: continue to invest in the development of supercomputers at European level, to give as many parties as possible access to computing power.

Proposal no. 6: the government and/or companies responsible for managing supercomputers could look into how to propose an open, non-discriminatory framework that would enable companies to use public supercomputer resources for a fee, while maintaining priority for research, particularly academic research.

Proposal no. 7: in conjunction with the AI Act in particular, set criteria for the openness of generative AI models trained on public supercomputers.

Take account of the economic value of data

Agreements between rights holders and developers should reflect the relative importance of the data for the developers according to the use case, and specify in which circumstances the data may be used.

> Proposal no. 8: public authorities, in particular as part of the mission entrusted by the French Ministry of Culture to the French Higher Council for Literary and Artistic Property, could encourage rights holders to take account of the economic value of data according to the use case (for example, by introducing differentiated pricing), and to propose bundled offers to reduce transaction costs, in order to safeguard the innovation capacities of model developers

Wherever possible, the public sector should play a leading role in making public data more open. Such initiatives can also help to ensure better representation of French (and European) language and culture among generative AI models, where English currently predominates.

> Proposal no. 9: make public and private data available for the training or fine-tuning of generative AI models, and encourage public and private initiatives to distribute French-language data, whether text, image or video.

Ensure greater transparency on investments by digital giants

With no change to existing legislation, there should be greater transparency of minority investments in the sector.

Proposal no. 10: the Commission could request further information on minority investments in the same sector as the target, in the template relating to the obligation to inform about a concentration pursuant to Article 14 of the DMA.

OPINION 24-A-05 OF 28 JUNE 2024

on the competitive functioning of the generative artificial intelligence sector

See the full text of the opinion in English

Presentation slides
See the press conference sldies

Contact(s)

Nicola Crawford
Communications Officer
+33155040151
Contact us by e-mail

Maxence Lepinoy
Chargé de communication, responsable des relations avec les médias
06 21 91 77 11
Contact us by e-mail