Summary[1]

Only the French version is authentic. In the event of any discrepancy, the French version will prevail over the translation.

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.

Generative AI

According to the European Parliament, AI 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**.

There are two key phases in generative AI modelling:

- *training*: **the initial learning process of a model (often called "foundation model", which includes large language models [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, such as answering end users' questions, or to 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.

---

[1] This summary is for information purposes only. Only the following numbered reasons for the opinion are authentic.

The participants in the value chain

The generative AI value chain extends upstream from the design, training and inference of models to their use downstream by end users. The operators in the generative AI value chain are:

- **major digital companies**: Alphabet and Microsoft are present across the entire value chain (vertical and conglomerate integration), 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. They may adopt a more, or less, open approach as regards the information available about their models and the possibility of re-using and adapting them.

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

- **IT component suppliers** develop graphics processing units (GPUs) and AI accelerators, which are essential components for training generative AI models. In addition to Nvidia, the sector's leading operator and the world's most valuable publicly-traded company at the date of this opinion, and major digital companies that develop their own AI accelerators, the sector also includes traditional operators like Advanced Micro Devices (AMD) and Intel;

- **cloud service providers** play a key role in the development of new AI technologies, as they provide the storage, data processing and computing capabilities needed, in particular, by language model developers. They include both digital giants, known as "hyperscalers", such as Amazon Web Services (AWS), Google Cloud Platform (GCP) and Microsoft Azure, cloud providers such as 3DS Outscale, IBM and OVHcloud, as well as specialist AI providers such as CoreWeave. The cloud sector was described by the *Autorité* in Opinion 23-A-08 of 29 June 2023. The necessary computing resources may also be provided by **public supercomputers** (such as Jean Zay in France), which have historically been dedicated to high-performance computing and have diversified to accommodate AI research projects.

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).

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 aimed at equipping France with competitive research capabilities and deploying AI technologies throughout the economy. In March 2024, the French AI Commission (*Commission de l'IA*) launched by the Prime Minister presented 25 recommendations calling, in particular, for measures to make France a major centre for computing power, to facilitate data access and to establish global AI governance.

At the European level, several legislation governing the development of the AI sector have been adopted over the last two years. In particular, most of the provisions of the **AI Act** (which will soon be published in the EU Official Journal) will be applicable from 2026. Under the Act, providers of large generative AI models will be subject to transparency obligations and must implement policies to ensure compliance with EU copyright law when training their models. The obligations do not apply to free and open-source models, whose parameters are public (unless they present a systemic risk). 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.

High barriers to entry

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.

Aside from digital giants and a few companies with sufficiently large in-house data centres (like Meta or Samsung), **the cloud appears to be the only way to access the computing power** needed to train models. The cloud gives developers access to AI-specific infrastructure and platform services adapted to their needs, while avoiding massive initial investment in IT infrastructure. **The cloud is also a vector for distributing models downstream on marketplaces.**

In addition, training large generative AI models requires **large volumes of data**. Most of this data is 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**. On the one hand, models are getting bigger and bigger and training requires more and more data, raising fears that publicly-accessible data will not be sufficient in the future and that proprietary data held by a small number of operators will become more important. On the other hand, access to certain publicly-accessible data is creating **legal uncertainties**, as illustrated by the actions brought by several rights holders, such as the complaint filed by the New York Times against OpenAI and Microsoft.

Lastly, training large models also requires **highly advanced technical skills** in machine learning, as well as empirical experience that can only be acquired by working with the models.

Operators in the generative AI sector require **substantial funding** to meet their computing power, data and skills needs. Investment in the sector increased six-fold between 2022 and 2023, to more than €20 billion.

Barriers to entry potentially limited by technical and organisational developments and certain public policies

First, computing power can be accessed via **public supercomputers**. In return for contributing to open science (for example, publishing work in an academic journal), access to public supercomputers is free, which can help to reduce the barriers to entry for certain operators, in particular in the research world. For example, a team of researchers from the CentraleSupélec university has trained a model called CroissantLLM on the French supercomputer Jean Zay. The joint undertaking EuroHPC is working to develop supercomputers throughout Europe and plans to **install a new supercomputer in France** in 2025.

Second, a **number** of technological innovations are already reducing the need for data and computing power:

- **innovations in generative AI model architecture**, which are making the training and fine-tuning phases more efficient and less costly. Examples include Mixture of Experts (MoE) and Low Rank Adaptation (LoRA);

- **smaller models**, which are easier to use for the inference phase and can be used on smartphones, for example;

- **synthetic data** (also generated by AI), which can partially replace real data and reduce the constraints associated with the use of personal data. However, the use of synthetic data entails certain risks, such as bias or a higher error rate.

Lastly, many developers choose an **open-source approach** in order to contribute to overall knowledge about the technology, thereby enabling other operators to re-use or fine-tune the models. However, open source covers a wide range of scenarios, from open-weight models where only the model weights are made public (the most common scenario) to fully-open models where all the code, architecture, training data, weights and training process are made public. While publishing model weights can have a beneficial impact on competition for fine-tuning and inference, it does little or nothing to reduce the barriers for an operator wishing to train a foundation model. In order to reproduce an AI model, other elements would need to be made public, such as the code and data for training or the data used.

Advantages for some companies linked to their activities in other digital markets

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. Major digital companies are also starting to develop alternatives to Nvidia's CUDA software.

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 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 job prospects offered by major digital companies, given their reputation for innovation, their global positioning and their wide catalogue of services.

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, which guarantees access to users, companies and consumers**. The sector is characterised by the high fixed costs involved in the initial training of a foundation model, which gives rise to **economies of scale** as operators seek to spread costs over as many users as possible. Generative AI products are also characterised by **economies of scope** because, once developed, a foundation model can be used for a wide variety of applications. The generative AI sector can also give rise to **cumulative network effects**, with feedback data from users being used to refine future models and improve performance or offer new services.

The *Autorité* also notes 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, "Copilot for Microsoft 365". 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.

Competition risks upstream in the value chain

While it seems premature at this stage to draw definitive conclusions about the definition of relevant markets and the market power of certain operators, vigilance is nevertheless required because major digital companies' access to key inputs and the advantages linked to their vertical and conglomerate integration create the conditions for strong concentration, to their benefit, and reinforce their power on distinct but linked or related markets, such as office productivity software, search engines or online advertising. **In certain cases, it may therefore be useful to perform the competitive analysis in terms of ecosystems being either created or reinforced, rather than market by market.**

The traditional tools of competition law, **such as cartel law and, above all, abuse of dominant position**, remain fully relevant. Other legal tools could also be used, such as **abuse of economic dependence**, where no position of dominance exists, or, with regard to contractual practices, the **law on restrictive competition practices**, the implementation of which falls mainly within the remit of the Directorate General for Competition Policy, Consumer Affairs and Fraud Control (DGCCRF) and the commercial courts.

Several risks of abuse identified by the *Autorité*

❖ **The risk of abuse by IT component providers**

France Digitale, an association representing a large number of French digital start-ups and investors, points to potential risks such as **price fixing, supply restrictions, unfair contractual conditions and discriminatory behaviour**. Concern has also been 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 a dawn raid in September 2023, is being closely scrutinised by the *Autorité*'s Investigation Services.

❖ **The risk of lock-in by major cloud service providers**

The *Autorité* notes 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.

First, **particularly high levels of cloud credits** are being offered to innovative companies in the sector. **Technical lock-in** practices have also been identified.

Such practices could be assessed under competition law, 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") or by the European Data Act.

❖ **The risks associated with data access**

Innovative companies in the sector may be confronted with practices of **refusal of (or discriminatory) access** by companies with significant access to data, such as a web index.

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**. In Decision 24-D-03 in the "related rights" case, the *Autorité* established that Google had used content from press agencies and publishers to train its foundation model Gemini (a chatbot based on the foundation model of the same name and formerly called "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.

❖ **The risks associated with access to a skilled workforce**

In competition law, supervisory authorities pay particular attention to practices in the labour markets. 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.

While it appears from the preparation of this opinion that such restrictions are not, for the time being, raising any particular concerns for stakeholders, the *Autorité* considers that vigilance is required.

❖ **The 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 re-use of models or some of their components can lead to users being locked-in.

❖ **The 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 **exclusivity agreements** between cloud service providers and foundation model developers. In their view, such agreements aim to make the developers exclusively dependent on the cloud service providers for access to the necessary cloud services and for customer distribution, and are therefore likely to have **an impact on innovation** and competition between providers, especially when a particular model occupies a significant position on the market.

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.

Through any of the above behaviours, certain companies could use their market power in distinct but related markets to the detriment of alternative operators, thereby restricting the choice available to users and the incentive to develop alternative solutions.

Competition concerns about minority investments and partnerships by digital giants

In a sector such as AI, where investment is very high given the cost of access to inputs, only a few major players have the financial capacity to enter into agreements with or invest in innovative start-ups. Investments and partnerships between operators in the sector are not problematic *per se*. They can give start-ups 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. On the one hand, the transactions may be subject to prior authorisation under merger control rules if they give investors *de facto* control and exceed EU and national notification thresholds. They may also be examined, under certain conditions, if they are below said thresholds, or as part of the analysis of a merger. On the other hand, they may be assessed *ex post* through competition law, on the basis of cartel law or abuse of dominant position (including collective dominance). 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.

The risk of collusion between companies in the sector

While almost all the stakeholders consulted during the public consultation did not express any specific concerns about the risk of collusion, the use of generative AI could potentially give rise to concerted practices that are already known and which 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.

Outlook

The *Autorité* notes that generative AI is far from having reached its potential. Less than two years after the launch of ChatGPT, many established operators have invested in the field and a multitude of start-ups have emerged to accelerate research and deploy the technology to companies and consumers.

The **race to innovate** and develop new generative AI models is likely to continue on two aspects: model size and optimisation at constant size. Model size is also a key factor in the **environmental impact of generative AI**.

The *Autorité* has also observed a trend towards **"platformisation"** in the generative AI sector. MaaS seems to be only way for model developers to reach consumers and AI using companies.

One of the main challenges for the healthy development of competition in the generative AI sector lies in the deployment of open-access resources. If the sector had more precise criteria for qualifying the degree of openness of a model, operators who so wished could use model openness as a competitive advantage.

Recommendations

Competition in the sector could be strengthened by the following recommendations, most of which do not require new legislative initiative at the French or European level.

The *Autorité* calls for full use to be made of the regulatory framework applicable to the sector.

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*.

In addition, at the French level, the *Autorité* encourages the DGCCRF to pay particular attention to the use of cloud credits in AI, in particular as part of the implementation of the SREN Law.

Lastly, the future EU AI Office and the competent national authority in France, which will be designated in accordance with Article 70 of the AI 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.

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.

With regard to access to computing power, the *Autorité*, like many public authorities, 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.

With regard to data, 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.

Lastly, the *Autorité* calls for **greater transparency** on minority investments in innovative companies, on the basis of Article 14 of the DMA, under which designated companies can be asked for information on their acquisitions.