MINISTÈRE
DE L'ÉCONOMIE,
DES FINANCES
ET DE LA SOUVERAINETÉ
INDUSTRIELLE ET NUMÉRIQUE
Liberté
Égalité
Fraternité

Trésor-Economics

No. 354 • December 2024

Direction générale du Trésor

The Artificial Intelligence Value Chain: What Economic Stakes and Role for France?

Solal Chardon-Boucaud, Arthur Dozias and Charlotte Gallezot

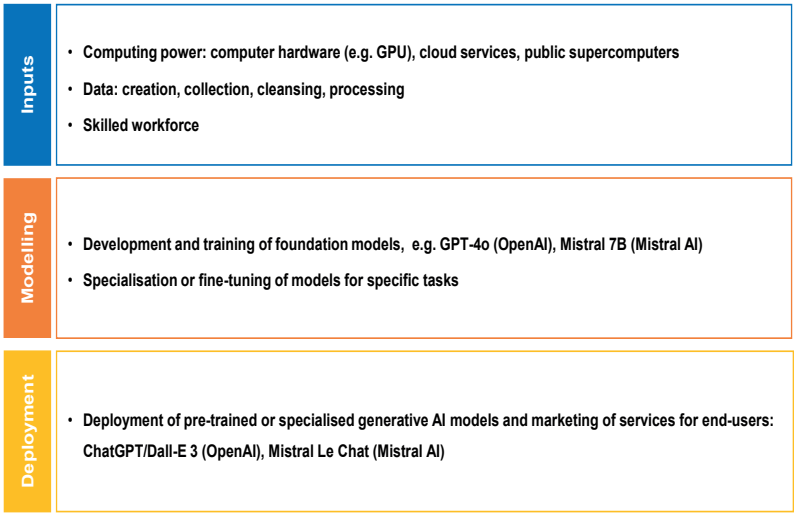
- Artificial intelligence (AI) refers to various techniques enabling machines to simulate human intelligence. The AI value chain is divided into three main segments (see Chart below): (i) the inputs required to develop AI systems and services (computing power, data, specialised workforce); (ii) modelling, which includes the development of general-purpose AI models (foundation models) and specialised models; and (iii) the deployment of such models to end-users.

- Regarding AI system inputs, France has no major companies on par with the global leaders in the chip manufacturing and computing power rental markets. However, it benefits from a skilled workforce and a thriving innovation ecosystem.

- As for AI model development, a few French companies are emerging, but the segment is dominated by the incumbent Big Tech firms established prior to the advent of AI technology. These incumbents enjoy vertical integration thanks to their preferential access to upstream inputs and downstream distribution channels for their AI solutions (e.g. office suite software). They have also partnered with emerging AI firms to integrate AI production processes horizontally.

- The dominance of the AI market by a small number of large, already mature firms raises issues of economic efficiency, fair competition, and sovereignty, with the risk of limiting the economy-wide diffusion of AI-related added value and productivity gains.

Artificial intelligence value chain



Source: Climate Finance Provided and Mobilised by Developed Countries in 2013-2022, OECD (2024).

1. AI system inputs: France has a skilled workforce, but is poorly positioned in the computing segment

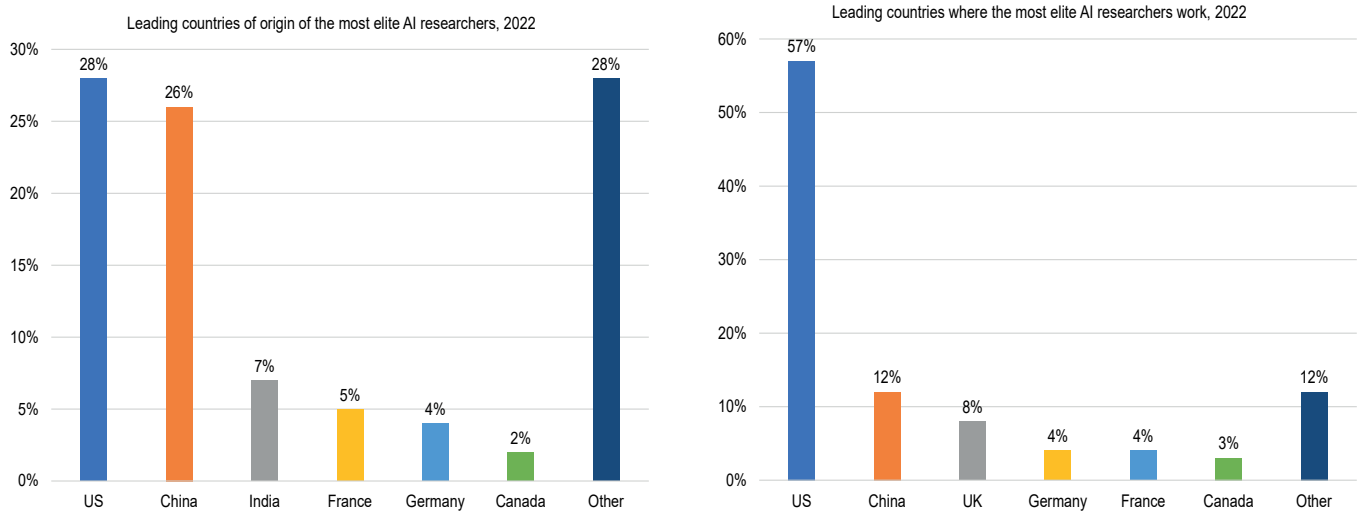
Developing AI systems requires substantial computing power, extensive quality datasets, and a specialised workforce. Although the computing market (infrastructure and rental) is highly concentrated among a few US companies, France benefits from a highly skilled workforce and a thriving research ecosystem. However, France could better harness certain pools of data, particularly to counterbalance the outsize influence of English-language model training data.

1.1 France has a thriving AI research and innovation ecosystem

The quality of the research and innovation ecosystem greatly affects the development of AI systems. Firstly, AI model development requires a highly skilled workforce that is able to innovate in order to develop competitive, state-of-the-art models in a fast-evolving field. In this respect, a crucial factor is the quality of the education and research system involved in training this workforce. Secondly, the emergence of innovators is fostered by an existing innovation ecosystem that pools the requisite talent and enables startups to grow and raise equity funding.

The education and research ecosystem is an advantage for France, which has a well-educated workforce and a broad talent pool of mathematics, computer science and engineering graduates.[1] In 2022, France was the fourth leading country of origin of the most elite AI researchers after the US, China and India, and the fifth leading country where they work in 2022, after the US, China, the UK and Germany[2] (see Chart 1). In addition, French AI research has a satisfactory level of scientific publications relative to other European countries: in 2021, in terms of number of publications, France ranked fourth in Europe after the UK, Germany and Italy and in tenth place globally.[3]

Chart 1: Leading countries of origin of the most elite AI researchers and leading countries where they work, 2022



Source: MarcoPolo (2023), The Global AI Talent Tracker 2.0.

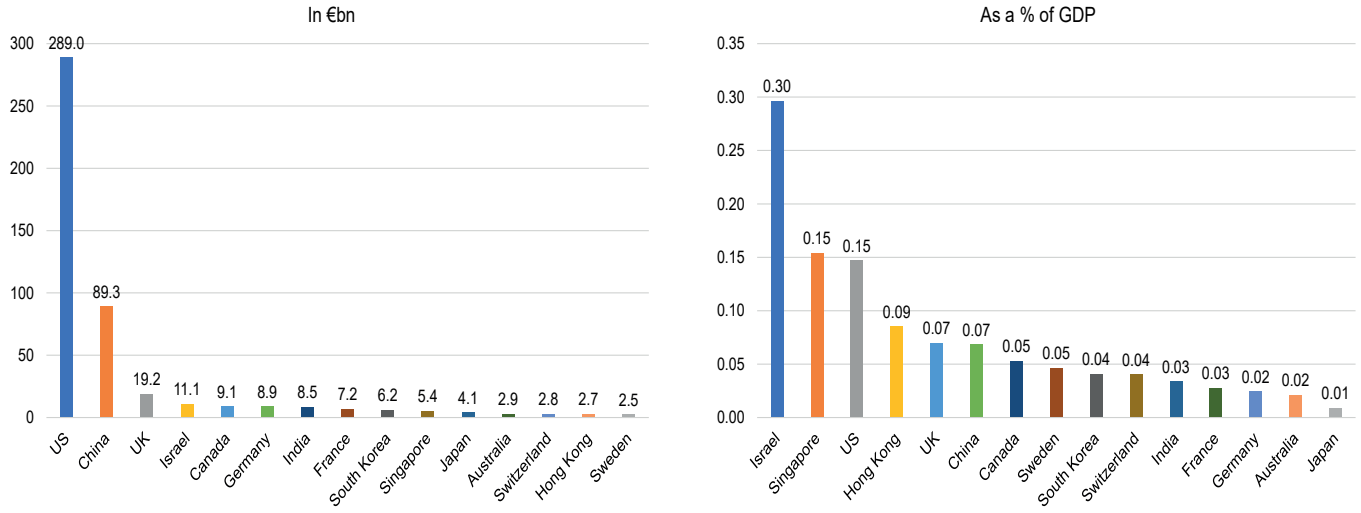
How to read this chart: The "most elite AI researchers" are defined as authors of papers selected for Oral Presentations at NeurIPS, which represent the most prestigious class of papers. The Oral Presentations acceptance rate was 1.8% in 2022.

(1) OECD (2023), Education at a Glance 2023: OECD Indicators. Table B5.3.
(2) MarcoPolo (2023), The Global AI Talent Tracker 2.0
(3) Source: OECD.AI (2024), based on data from OpenAlex. According to the same source, France maintains these rankings when taking into account "high-impact" publications only. In number of publications in all disciplines, France ranks ninth globally.

This ecosystem is built around world-class research organisations[4] such as the French Centre for Scientific Research (CNRS) and the National Institute for Research in Digital Science and Technology (INRIA), as well as outstanding educational programmes, particularly in mathematics.[5] However, French research centres are struggling to retain talent because they lack the appeal of a number of foreign institutions offering higher pay and better career opportunities.[6] Furthermore, a "culture of innovation" has taken root and grown in France's main innovation "hubs",[7] particularly the Paris innovation hub currently under development. These hubs generate positive externalities related to the concentration of complementary talent, especially researchers and those with business and entrepreneurial backgrounds.[8]

European AI startups have managed to attract rather substantial private funding in the last decade (from 2013 to 2023, €29.5bn of investments went to the European Union (EU), including €7.2bn for France and €8.9bn for Germany)[9] (see Chart 2), but this was much lower than in the US (€289bn), China (€89.3bn) and, to a lesser extent, several other Asian countries. Relative to GDP, France is ranked lower than tenth worldwide in AI startup funding, behind the UK, Sweden, Israel, South Korea, and India, while outranking Germany, Japan, and Australia (see Chart 2).

Chart 2: Private investment in AI funding rounds by country, 2013-2023



Source: For data on private investment in AI funding rounds: Stanford University (2024), AI Index Report 2024, data from Figure 4.3.9. For GDP data: World Bank.

How to read this chart: To express the figures in euros at current prices instead of in dollars at current prices, we used the average euro/dollar exchange rate for the 2013-2023 period. Given that 2023 GDP data is not available for some countries, we used GDP data for the 2013-2022 period as a proxy by adding an additional year of GDP data based on 2022.

(4) High Council for Evaluation of Research and Higher Education (HCERES), CNRS Assessment Report (2023), INRIA Assessment Report (2024).
(5) In 2023, three French universities ranked in the top ten of the Shanghai Ranking for mathematics: Paris-Saclay University (#2), Sorbonne University (#4) and Paris Cité University (#7).
(6) Artificial Intelligence Commission report (March 2024), "IA : notre ambition pour la France" (in French only), chapter 2.2.4.
(7) Geographical and disciplinary clusters of various resources and organisations (startups, SMEs, research organisations, etc.) help to spur innovation.
(8) L.A. de Vasconcelos Gomes, A.L. Figueiredo Facin, M.S. Salerno, R.K. Ikenami (2018), "Unpacking the innovation ecosystem construct: Evolution, gaps and trends", *Technological Forecasting and Social Change*.
(9) Stanford University (2024), *AI Index Report 2024*, data from Figure 4.3.9 for the France estimate and Figure 4.3.10 for the EU estimate.

Although the French and European venture capital ecosystem plays a significant role in these funding rounds and is seemingly effective for early-stage startups, the availability of French and European capital remains limited when raising €50m to €100m or more,[10] with these amounts generally being required to fast-track the global scale-up of tech firms. In France and the rest of Europe, companies in late-stage development are having significant trouble obtaining financing from European investors.[11] Between 2013 and 2023, among the funding rounds involving European scale-ups,[12] 82% had a foreign (i.e. non-European) lead investor.[13] Foreign buyers made up more than 60% of the deals among the European scale-ups that were acquired.[14] France's national Tibi initiative[15] and EU-level Scale-up Europe Initiative[16] seek to increase the size of French and European tech funds so that they can participate in funding rounds of more than €100m. Investment programmes in the mould of InvestEU[17] and, more broadly, the development of the Capital Markets Union[18] should also help to improve Europe's lagging participation in late-stage funding rounds.

1.2 France is not currently a major computing player

The training and deployment of AI models – and foundation models, in particular – require substantial, costly computing power. To meet these needs, a company can (i) buy computing power directly, (ii) rent computing power (through cloud computing services or possibly the rental of dedicated infrastructure) or, (iii) to a lesser degree, use public computing infrastructure (e.g. the Jean Zay supercomputer). AI models are trained and run mainly on graphics processing units (GPUs) that allow computations to be performed simultaneously (parallel processing), thus speeding up their processing.

The GPU market is highly concentrated around Nvidia, a Silicon Valley company that has a virtual monopoly with a global market share of around 85% in 2023. [19] AMD, the second-largest GPU producer, has a market share of 10%. Nvidia's dominant position is boosted by its software expertise, notably the software library Compute Unified Device Architecture (CUDA), which helps to optimise computing. Several firms are currently working to find alternatives to GPUs:[20] startups like Cerebra Systems, as well as Big Tech firms like Google, are seeking to optimise chips for AI, especially in order to reduce the high energy consumption of GPUs.[21]

The computing infrastructure and computing power rental markets also have a high concentration of US companies, called hyperscalers, while existing French firms play a negligible role in this ecosystem. To illustrate this, in 2021, Amazon (AWS) controlled 46% of France's cloud computing market, Microsoft (Azure) 17% and Alphabet (Google Cloud) 8%.[22] There are, however, more specialised cloud firms (pure players),[23] including French ones (e.g. Scaleway, OVHcloud, Outscale), that are competitive relative to major US firms in certain specific cloud services but are disadvantaged by their limited service offering.[24] In France, Scaleway announced it would invest in computing infrastructure for AI applications and partner with Nvidia to provide AI model training services. Other French cloud providers are specialising in AI inference,

(10) F. Hafied, C. Rachiq and G. Roulleau (2021), "Venture Capital and Development of French Start-ups", *Trésor-Economics,* No. 276.
(11) European Investment Bank (EIB) (2022), "EIB Group supports the pan-European Scale-up Initiative to promote tech champions".
(12) Defined as companies with a valuation of between $500m and $10bn at the time of the funding round.
(13) In private equity co-investments, lead investors are those making the largest investment and setting the deal's terms and conditions (e.g. valuation, governance).
(14) EIB (2024), "The scale-up gap: Financial market constraints holding back innovative firms in the European Union".
(15) Launched in 2019, Tibi is a market initiative seeking to increase funding for French tech startups by utilising the savings of institutional investors. In the first phase (2020-2022 period), €6bn was invested, helping to provide €30bn to the tech ecosystem. For the second phase, investors have pledged to invest €7bn by the end of 2026.
(16) This initiative spurred the European Tech Champions Initiative (ETCI), a €3.85bn fund of funds managed by the European Investment Fund (EIF) that helps funds to reach a critical mass of €1bn.
(17) As the successor to the Juncker Plan, which generated investments of €550bn from 2015 to 2020, the InvestEU Programme aims to harness €372bn of investment by 2027, to the benefit of a wide range of recipients, from social housing operators to startups and leading auto equipment manufacturers. This amount should be reached through leveraging, with a €26bn EU budget guarantee provided. The InvestEU Fund is implemented through the EIB Group (including the EIF) and through other financial intermediaries, such as Bpifrance and Caisse des dépôts.
(18) See recommendations of the Noyer Task Force (2024), "Développer les marchés de capitaux européens pour financer l'avenir" (in French only).
(19) G. Sastry et al. (2024), "Computing Power and the Governance of Artificial Intelligence".
(20) *Le Monde Informatique* (2024), "Des start-ups poussent des alternatives au GPU pour l'IA" (in French only).
(21) A. Shilov (2023), "Nvidia's H100 GPUs will consume more power than some countries", Tom's Hardware website.
(22) French Competition Authority (June 2023), Opinion 23-A-08 on competition in the cloud sector.
(23) A pure player is a company that operates in a single business sector without diversifying.
(24) French Competition Authority (2023), op. cit.

i.e. the deployment and use of a model after the training process.

Although the dearth of French players able to rival the size of the leading global firms in chip manufacturing and cloud computing does not currently hinder French firms' access to computing power (today's French players enjoy the same conditions as foreign companies), the fact that a small number of firms dominate these sectors raises many concerns.

In terms of chip manufacturing, Europe is vulnerable to the risk of shortages and, in the long term, export restrictions,[25] at a time when surging demand has led to tight supply conditions. The French Competition Authority's recent report on generative AI[26] underscores the fact that Nvidia may be engaging in anti-competitive practices given its dominant market position (oligopolistic pricing, production restrictions, unfair contractual terms, favouritism towards a number of large customers who have been stockpiling a large share of the available chips).[27] Further concerns are emerging about Nvidia's CUDA chip programming software – the only system fully compatible with the GPUs that have become essential for accelerating computing – and over Nvidia's recent investments in AI-focused cloud service providers such as CoreWeave.[28]

Regarding cloud services, the dominant cloud-computing providers offer startups services that allow them to access computing resources at a lower cost (such as through cloud credit-based services, i.e. trial offers giving customers free access for a set period). Where these practices are combined with technical or financial lock-in (such as migration barriers), they threaten fair competition and could be considered as an abuse of dominant position.[29]

Against this backdrop, and in response to the financial and technical difficulties that are hampering certain organisations' (particularly public and private research entities) access to computing power,[30] France and Europe are investing in expanding public computing infrastructure, such as the Jean Zay supercomputer in France and the EuroHPC supercomputers in the EU (which will include the commissioning in 2026 of the new Jules Verne supercomputer in France). The recent Draghi report[31] on the competitiveness of the European economy calls for the development of these supercomputers across Europe.

While the development of this infrastructure is not intended to replace that of private-sector organisations, such additional capacity may help to ensure that public and private research entities have access to computing power. Accordingly, the Jean Zay supercomputer is accessible to private entities carrying out open research and publishing their findings. Furthermore, sovereign computing resources need to be created for specific sensitive uses of AI (e.g. defence, health) that pose a greater threat to national security, trade secrets, and privacy.[32]

1.3 French and European data pools could be better harnessed

Developing AI models necessarily requires access to datasets of sufficient quality and quantity, both for foundation models and related applications. Foundation models are primarily trained on large, general-purpose datasets containing text, images, and videos. Tailoring a pre-trained model for a specific use, whether through the fine-tuning method or retrieval-augmented generation (RAG),[33] requires more granular sectoral data. Data cleansing and processing are a differentiating factor in both cases, as low-quality data should be filtered out.

While some large datasets are accessible to the general public free of charge (for example, the non-profit Common Crawl has given the public access to its web archive since 2008), organisations with proprietary data have a competitive advantage in developing foundation models and related applications.

(25) Such restrictions are currently in force in the US for the sale of chips to China.
(26) French Competition Authority (2024), Opinion 24-A-05 on the generative artificial intelligence sector.
(27) J. Keegan (2024), "Just four companies are hoarding tens of billions of dollars worth of Nvidia GPU chips", Sherwood News.
(28) French Competition Authority (2024), op. cit.
(29) French Competition Authority (2024), op. cit.
(30) As highlighted in the recent Draghi report (see footnote on page 32 of said report).
(31) M. Draghi (2024), *The future of European competitiveness - A competitiveness strategy for Europe*.
(32) La Lettre (2024), "Le ministère des armées finalise le lancement de son cloud privé" (in French only) and General Inspectorate for Social Affairs (IGAS) report (2023), *Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire des données de santé* (in French only).
(33) During fine-tuning, a part of the model (including its parameters) is retrained on a new dataset that is representative of the target context. RAG, on the other hand, combines the model with a search engine linked to external knowledge bases, which helps to generate results that are more relevant without retraining the model's parameters.

This is true of Big Tech firms such as Facebook, Amazon, Microsoft or Google, which have preferential access to vast quantities of data via the content they host (for example, YouTube provides Google with a significant source of training data for AI models) and the data associated with the use of their services. This advantage is even greater now that the overall quantity of free, accessible and quality data appears to have peaked, at a time when models continuously require more data for training.[34]

In this context, Big Tech firms, along with most social media platforms and media companies, tend to restrict access to the data they publish. For example, YouTube prohibits third-party applications from using its videos. Similarly, some organisations[35] employ various strategies to limit third-party access to their users' data, sometimes abusing legal rules such as personal data protection or using security concerns as a pretext. In addition, Big Tech firms can use their financial power to enter into data exclusivity agreements with third-party data owners, an act that may constitute an anti-competitive practice.

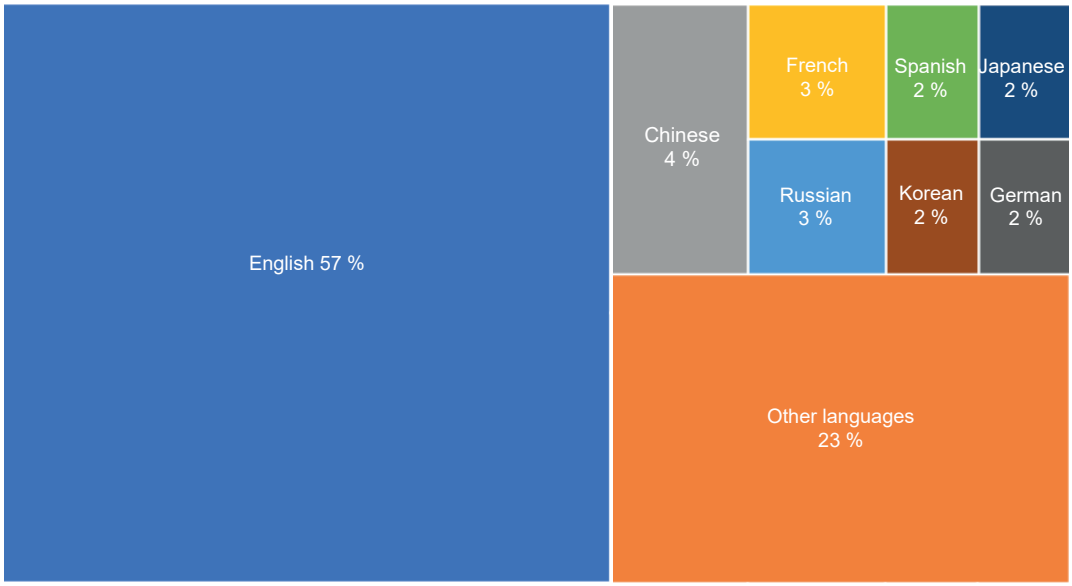
In response to this competitive advantage, several public initiatives are geared towards broadening companies' access to data, whether general-purpose or sectoral data, and making the data usable (standardising the nomenclature used by different sources, etc.). Regarding data derived from public or semi-public entities, France is at the forefront of open data at the EU and international levels.[36] The data.gouv.fr platform has made more than 40,000 datasets available on subjects such as property transactions (the Property Value Request, or *Demande de valeurs foncières*, database), land records, health insurance reimbursements, biodiversity and pollution. Regarding sectoral data, more often used for applications, France backs the Common European Data Spaces (CEDS) initiative, which aims to make integrated databases available in certain sectors (e.g. health, mobility, energy) while ensuring the protection of the data concerned.[37] Similar initiatives have also been introduced under the France 2030 Plan's Digital Commons for Generative AI call for projects. More specifically, in the health field, France could, in addition to its participation in the CEDS initiative, draw from the measures implemented by certain countries like the US and the UK, which each developed a framework to improve data reuse while ensuring its confidentiality.

A critical issue to address beyond access to data is ensuring the quality and diversity of the data used to train AI. The cultural origin and the language of the data used partly determine the extent to which the data is represented. This can lead to models performing less well in understanding and processing information in some languages. The vast majority of data available for training models are in English, although French is relatively well represented among "Other languages". On Hugging Face, a platform providing AI tools, models and databases to its community of users, 57% of the training data is in English, compared to 3% in French (see Chart 3). This means that general-purpose AI models trained on open-source data are optimised for processing information in English and that they are less efficient when processing languages that are less represented, limiting the usefulness of these AI models for the speakers of such languages.

There is therefore a need for French-language training data in order to increase the number of French references in AI models and to have models that work efficiently for French users. With this aim in mind, the Villers-Cotterêts initiative, launched in December 2023 by the French Audiovisual Institute (INA), the National Library of France (BnF) and the CNRS, will compile a vast quantity of French-language texts usable by AI models. The upcoming launch of the European Digital Infrastructure Consortium for the Alliance for Language Technologies (ALT-EDIC),[38] which will increase data availability in various European languages, represents another such effort to expand language diversity in AI models.

(34) P. Villalobos et al. (2024), "Will we run out of data? Limits of LLM scaling based on human-generated data".
(35) French Competition Authority (2023), "Advertising on iOS mobile applications: the General Rapporteur confirms having notified the Apple group of an objection".
(36) European Commission (2023), 2023 *Open Data Maturity Report*. OECD (2023), *OECD Digital Government Index 2023*.
(37) The CEDS initiative's success will depend on the participation and maturity of companies in managing their data, in order to set standards for intra-sectoral data sharing. On this front, Germany appears to be ahead of the curve concerning the involvement of companies.
(38) A European Digital Infrastructure Consortium (EDIC) is an instrument made available to EU Member States under the Digital Decade Policy Programme 2030 to speed up and simplify the implementation of multi-country projects.

Chart 3: Percentage breakdown of languages for AI training datasets on Hugging Face, from a list of 225 languages, 2024



Source: OECD (2024), OECD Digital Economy Outlook 2024 (Volume 1), Figure 2.2.
How to read this chart: This chart represents the language distribution of AI training datasets on Hugging Face from a list of 225 languages in 2024. Multilingual and translation datasets on Hugging Face contain more than one language and are thus double-counted. The languages included under "Other languages" account for less than 2% of all AI training datasets.

Box 1: Striking a balance between improving access to AI data and protecting intellectual property rights

The use of cultural content to train foundation models raises serious questions about respecting copyright-protected works. From an economic standpoint, the utility and value of an AI model depends on the quantity and quality of available data. Given the amount of data involved, negotiating with each individual content creator or rightsholder can entail significant costs for AI producers, even though maintaining an incentive to create quality French content is beneficial both for rightsholders and AI producers.

In Europe, the use of copyright-protected content to train an AI system is regulated by Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market.[a] Although the Directive introduces an exception for uses of text and data mining technologies by AI system providers, the rightsholder of the content can opt out of this exception and refuse permission for the use of their content. This option aims to restore rightsholders' negotiating power. Nevertheless, rightsholders are struggling to opt out in practice due to AI producers' lack of transparency about the training data they use.

Formally endorsed by the EU Member States in February 2024, the Artificial Intelligence Act, or AI Act, provides for a transparency obligation regarding AI systems' training data sources. Since data is a vital differentiating factor for AI system providers, such a requirement could adversely affect their growth. Consequently, the implementation of the AI Act's transparency obligation will need to strike a balance between the costs for AI model providers and the protection of rightsholders.

a. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

2. AI model development: despite new innovators, the dominance of incumbent tech giants poses competitive challenges

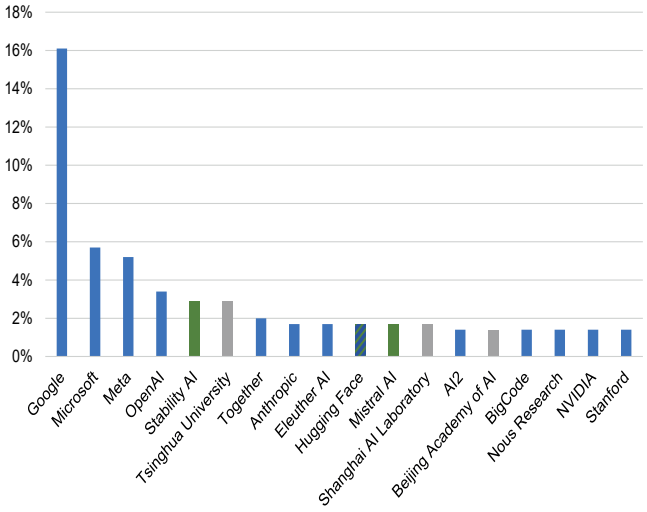
Silicon Valley tech giants dominate – directly and indirectly – the development of AI foundation models, although France has a successful startup, Mistral AI, in the field. Given their size and presence in various segments of the digital economy, Big Tech firms such as Google, Facebook, Amazon and Microsoft, have a competitive advantage since they have access to the requisite resources (data, computing power, specialised workforce, funding). This access represents a barrier to entry for other companies while enabling Big Tech firms to expand into the AI sector, especially by acquiring innovative firms or developing partnerships with them. There are fewer barriers to entry in specialised or downstream models, even if Big Tech firms benefit from their integration with their existing products. Tech giants' overall dominance poses many competitive and economic challenges, running the risk that the diffusion of AI-related economic gains (value added, profits, productivity) will be undermined or skewed by anti-competitive practices.

2.1 AI's emergence is part of a unique landscape characterised by tech giants' dominance, particularly in foundation models

The AI landscape is characterised by the rapid growth of startups offering high-performance models, as illustrated by OpenAI, the firm that developed ChatGPT. In France, Mistral AI develops some of the world's most powerful foundation models[39] after raising more than €1bn during three successive funding rounds in 2023 and 2024.

However, foundation model development remains primarily dominated by US Big Tech firms, including Google, Microsoft and Meta. Almost one-third of recent models have been developed by these three companies (see Chart 4). In terms of model performance, according to the LLM Leaderboard,[40] Google currently has an edge over its competitors with its Gemini model, which outpaces OpenAI's GPT-4o model. Additionally, Microsoft developed a strategic partnership with OpenAI (see below) that bolsters the key role Big Tech firms play in the foundation model segment.

Chart 4: Percentage share by organisation, based on total number of foundation models, 1 January 2022-4 October 2024



Source: Stanford Ecosystem Graphs for Foundation Models, 4 October 2024.

How to read this chart: US organisations are shown in blue, Asian in grey, and European in green. Hugging Face is shown in green with blue crosshatching because it is a Franco-American company. It was originally founded in Paris by three French nationals, but it is currently headquartered in New York and incorporated in Delaware. Based on a total of 348 models, only organisations that have developed at least five models over the period are shown in the chart. Models developed jointly by several organisations are counted once for each organisation.

Big Tech firms' dominant role in developing foundation models is primarily due to their resources, derived from many years of growth in the digital sphere: substantial financial resources for investing, as well as the ability to attract top researchers by offering generous pay and favourable working conditions.

The vertical and conglomerate integration of tech giants across the entire foundation model value chain (see Chart 5) also gives them a natural competitive advantage:

- Before even designing a model, these companies enjoy direct access to the requisite computing power, which they already have to carry out their other activities. Tech giants have also invested in the design and manufacture of certain AI chips (see section 1) and in the ownership of software libraries that are essential for developers, such as Pytorch

(39) Chatbot Arena LLM Leaderboard, 20 November 2024.
(40) Chatbot Arena LLM Leaderboard, op. cit.

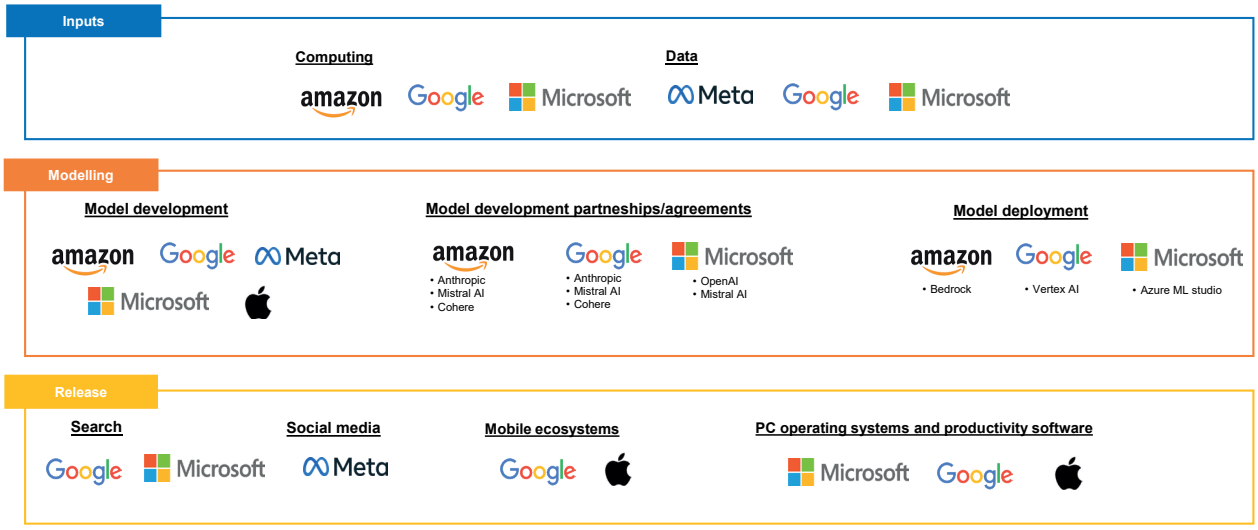
(Meta) and TensorFlow (Google), and which gives them standard-setting power in foundation model development.[41] In addition, they have access to massive volumes of data, which is a differentiating factor for foundation models.

- After designing a model, Big Tech firms can use their existing distribution channels to distribute it. Firstly, selling access to models is complementary to selling access to computing infrastructure, including cloud computing (with major synergy since using models requires substantial computing infrastructure). Big Tech firms thus distribute models using the "model as a service", or MaaS, approach in cloud marketplaces (e.g. Model Garden in Google Cloud and Amazon Bedrock in AWS). Secondly, foundation models can be directly integrated into existing products and services (e.g. software suites, search engines, social media platforms) with a large user base, thereby generating scale and network effects.

Ultimately, AI is unique in that it is a technological revolution where incumbent firms play a dominant role from the outset. In contrast, innovation is usually driven by new entrants that challenge incumbent firms' dominance or create new markets.[42] In the short term, investments made by established AI firms are favourable for the technology's development and make it possible to offer innovative functionalities rapidly. In the medium term, this situation could limit the development and diffusion of AI, relative to new, potentially more innovative entrants. The advent of large language models (LLMs) illustrates this paradox: Big Tech firms' AI models (Gemini, Llama, etc.) are compelling and developing rapidly. However, their diffusion among businesses and the general public has been significantly fast-tracked by the new entrants challenging the market (e.g. OpenAI's launch of ChatGPT in November 2022).[43]

In the longer term, the existence of open data and the growth of "energy-efficient" foundation models, which are less resource-intensive (e.g. computing power and data), could also help to reduce a number of currently identified barriers to entry, thus fostering the emergence of new players. The development of energy-efficient models is all the more necessary, as the energy consumption of AI services has increased exponentially in recent years, with some AI providers entering into agreements with suppliers of low-carbon energy, including nuclear power.[44]

Chart 5: Presence of Big Tech firms across the AI value chain



Source: UK Competition & Markets Authority, AI Foundation Models: Update paper.

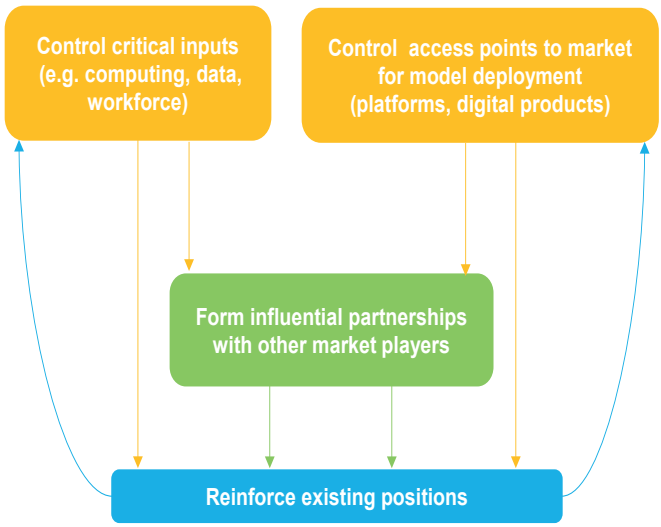
(41) Foundation models perform better when they have access to vast amounts of data and computing power (scale effects); their value increases if they are utilised by other users (network effects) because user queries will improve model performance.
(42) Aghion et al. (2005), "Competition and Innovation: An Inverted-U Relationship".
(43) See Nasse Seminar summary, "Intelligence artificielle : enjeux concurrentiels, enjeux économiques et applications pratiques dans le secteur des médias" (in French only), organised by DG Trésor and the French Competition Authority, 19 September 2024.
(44) *Le Monde* (2024), "Les géants du numérique se convertissent au nucléaire pour étancher les besoins énergétiques toujours plus importants de l'IA" (in French only).

2.2 Tech giants' dominance can create risks in terms of AI's market structure and development

Regulators, and competition watchdogs in particular (in both Europe and the US),[45] are monitoring the development of AI in this unique situation of Big Tech firms' dominance, as firms like the "Magnificent Seven" (Microsoft, Alphabet, Meta, Amazon, Apple, Nvidia and Tesla) could further consolidate the oligopolistic structure of digital markets. The French Competition Authority recommends paying close attention to strategies that tech giants could implement to consolidate their market power or leverage it to expand in the AI sector (see Chart 6). This due diligence should be applied to any potential anti-competitive practices related to inputs or distribution channels,[46] as well as to anti-competitive practices through acquisitions of innovative AI companies,[47] or through forming influential partnerships with emerging players in model development.

In certain partnerships, Big Tech firms allow their partners to use their resources (financial[48] as well as in terms of data and computing power) in exchange for preferential access to models developed by startups. This enables tech giants to distribute the models through their platforms or integrate them into their products. This is the prevailing strategy in a partnership between OpenAI and Microsoft – with Microsoft investing a total of more than $13bn in the company developing GPT models. As competition watchdogs have reported,[49] this type of agreement poses significant risks to competition: it could weaken competition between the two entities and lead to vertical effects or the lock-out of certain competitors. At the same time, this type of agreement can also help startups grow and, more upstream, create incentives to innovate with a view to the startup's sale or to a future partnership.

Chart 6: Competitive dynamics in the foundation model segment



Source: UK Competition & Markets Authority, AI Foundation Models: Update paper.

Big Tech firms can also, more simply, hire employees from startups without buying a particular company.[50] Such an approach is possible in segments where human capital is the main asset. It can also provide a way for tech giants to maintain their dominant position, but competition rules are more difficult to enforce in these circumstances.[51]

Big Tech firms' dominant position also weighs on business models and value chain organisation. In the medium term, the AI model market could become "platformised", meaning concentrated among a few platforms pooling ecosystems of tools that facilitate the development and use of foundation models.[52] However, major uncertainties remain as to the business models that could be rolled out. Platforms such as Amazon Bedrock, Hugging Face and Google

(45) For example, the US Federal Trade Commission (FTC) launched an inquiry into the competitive practices of Microsoft and OpenAI, while the US Department of Justice (DoJ) did the same for Nvidia. In the UK, the Competition & Markets Authority (CMA) submitted a report on AI foundation models in September 2023 and launched an investigation into Microsoft's partnership with OpenAI. In June 2024 in France, the Competition Authority published an opinion on the artificial intelligence sector (cited on several occasions in this paper). In October 2024 in Rome, Italy, the G7 Competition Authorities and Policymakers' Summit also underscored the importance of AI-related competition issues.
(46) See section 1 for computing power and data. See below for access to a skilled workforce.
(47) CB Insights (2021), The Race For AI: Which Tech Giants Are Snapping Up Artificial Intelligence Startups.
(48) Some startups have recently asked their investors to refrain from funding other companies, although to date it is not known if investors acquiesced to such requests. See Financial Times (2024), "OpenAI asks investors not to back rival start-ups such as Elon Musk's xAI".
(49) French Competition Authority (2024), op. cit. Competition & Markets Authority (2021), op. cit. Both competition watchdogs underscore in their reports the potential risks associated with the Microsoft-OpenAI partnership.
(50) Microsoft hired employees from Inflection, Google recruited workers from Character.AI and Microsoft offered to hire Sam Altman and other OpenAI employees when a disagreement erupted between Altman and OpenAI's board of directors. These potentially anti-competitive practices do not appear, however, to raise particularly serious concerns among firms in the sector at this stage (source: French Competition Authority, op. cit.).
(51) Especially as the implementation of non-poaching or non-compete clauses, which would prevent companies from being able to hire employees en masse, can also be considered anti-competitive in certain ways.
(52) Competition & Markets Authority (2023), *AI Foundation Models: Initial Report*.

Cloud already provide access to a plethora of tools and resources (e.g. models, databases, libraries) for developers of foundation models and their related applications. Platforms enjoy network effects given their role as an intermediary between developers and the businesses using these applications. Similarly to operating systems and smartphone app stores, only a small number of platforms – those that can provide the most complete ecosystem – will be able to survive in the medium term. Quite a few potentially anti-competitive practices on the part of these platforms (for example, related to the fact that the companies owning them can favour their proprietary models) have, moreover, already been identified by competition watchdogs. To prevent such practices, the French Competition Authority recommends[53] that MaaS services be included in the core platform services regulated by the Digital Markets Act (DMA).[54]

The concentration of foundation models among Big Tech firms also runs the risk of limiting the diffusion of AI-related productivity gains.[55] Indeed, although in the medium term a provider has an incentive to offer its model as open-source because the innovations arising from open models are beneficial to the provider and contribute to building an ecosystem around the model's architecture,[56] in the longer term, once the provider has consolidated its market leadership, it has an incentive to close the ecosystem and make it profitable. Productivity gains from using the model could then be captured as rent by model providers, with a more limited diffusion of the technology in the economy and higher costs for users. Consequently, the implementation of a proactive competition policy is essential to simultaneously ensure innovation in the AI sector and AI's diffusion throughout the economy.

The question of Big Tech firms' dominance in the value chain should also be addressed in relation to broader issues concerning the competitiveness of the European economy. The recent Draghi report on the matter[57] indeed shows that the growing productivity gap between the EU and the US is primarily driven by digital technology: the EU is first and foremost specialised in mature technologies, whereas the US is specialised in new innovative tech, especially digital tech, which generate greater economic growth.[58]

Lastly, the fact that the current global leaders are non-European firms raises questions regarding the location and taxation of AI-related profits and salaries. Against a backdrop of systemic under-taxation of large multinational enterprises, the presence of leading French or European firms would help to strengthen our tax sovereignty, thus ensuring a fairer and more appropriate tax policy for AI-related income. The reforms under negotiation at the OECD on the taxation of large multinational enterprises should help to provide solutions to this issue, bearing in mind that to harmonise taxation, it must be addressed on the broadest possible scale (at the EU level at a minimum), in order to ensure the effectiveness of such harmonisation. In this respect, reforms aimed at harmonising the business taxation environment are also currently under negotiation at the EU level.

2.3 The specialised model segment has lower "natural" barriers to the market's proper competitive functioning

Structural aspects that may explain the already highly oligopolistic character of the foundation model value chain are less applicable to the specialised model segment. Indeed, provided that the main foundation models remain accessible (i.e. without the principal foundation model owners using lock-out strategies), applications based on these models should remain affordable to develop. Tailoring a pre-trained model for a specific use requires less data and computing power than training a foundation model.

The main barrier to entry for new competitors in this segment may be sales materials, with Big Tech firms benefitting from their integration into their existing products, such as office software suites and social media platforms. This preferential access to essential distribution channels creates de facto barriers to entry

(53) French Competition Authority (2024), op. cit. Recommendation 1.
(54) Self-preferencing is referred to in Article 6 of the DMA, which compiles the practices of gatekeepers that may be prohibited on a case-by-case basis after assessment.
(55) L. Besson et al. (2024), "The Economic Implications of Artificial Intelligence", *Trésor-Economics*, No. 341.
(56) Meta (Facebook) made its Llama 2 model available in open source in mid-July 2023.
(57) M. Draghi (2024), "The future of European competitiveness": "The key driver of the rising productivity gap between the EU and the US has been digital technology ("tech")", part A, chapter 2.
(58) C. Fuest, D. Gros, P.L. Mengel, G. Presidente, J. Tirole (2024), "EU Innovation Policy: How to Escape the Middle Technology Trap", *EconPol Policy Reports*.

for deploying specialised models. The fact that Big Tech firms can favour their proprietary models in their products that make up most of the potential market poses, furthermore, the risk of abuse of dominant position.

France has many startups[59] and major companies[60] that are investing in new specialised use cases for AI. One of the main differentiating factors in this market will be using quality sectoral data tailored to specific needs. France can meet these needs by capitalising on its industrial firms, their data, and their use cases to spur the development of applications, especially as the relatively low cost[61] and quality of France's skilled workforce in AI (see section 1.1) are assets for its partners. These major companies can also play a relevant role with positive externalities in intra-sectoral data sharing.

Nevertheless, it must be underscored that irrespective of the nationality of the firms developing specialised models, their adoption by economic stakeholders, particularly among SMEs, is a major priority. Firms' use of powerful AI tools is indeed a source of growth that is potentially more significant than the direct gains generated by the creation and growth of French AI companies.[62]

(59) France Digitale (2024), "Mapping 2023 des start-ups françaises de l'IA" (in French only).
(60) L'Usine Nouvelle (2024), L'IA générative dans l'industrie française en 15 cas d'usage (in French only).
(61) According to Glassdoor, the average annual salary of an entry-level Machine Learning engineer is €55,000 gross in France compared to $150,000 in the US. This pay gap may nevertheless also be a barrier to attracting top talent.
(62) L. Besson et al. (2024), op. cit.

Publisher:

Ministère de l'Économie,
des Finances
et de la Souveraineté
industrielle et numérique
Direction générale du Trésor
139, rue de Bercy
75575 Paris CEDEX 12

Publication manager:

Dorothée Rouzet

tresor-eco@dgtresor.gouv.fr

English translation:

Centre de traduction
des ministères économique
et financier

Layout:

Mimose Mellia
ISSN 1962-400X
eISSN 2417-9698

Recent Issues in English

November 2024

No. 352 Japanification: a Risk for China's Economy?
Louis Bertrand, Thomas Carré, Patrick Kanda Tunda and Eloïse Villani

October 2024

No. 351 Minerals in the Energy Transition
Alban Aubert, Sary Zoghely and Xavier Le Guennec

https://www.tresor.economie.gouv.fr/Articles/tags/Tresor-Eco

Direction générale du Trésor

@DGTresor

To subscribe to *Trésor-Economics*: bit.ly/Trésor-Economics

This study was prepared under the authority of the Directorate General of the Treasury (DG Trésor) and does not necessarily reflect the position of the Ministry of Economy, Finance and Industrial and Digital Sovereignty