



autorité de régulation
des communications électroniques,
des postes et de la distribution de la presse

RÉPUBLIQUE FRANÇAISE

GENERATIVE AI: ENVIRONMENTAL ISSUES AND CHALLENGES

May 2026

With the Centre of Expertise for Digital Platform Regulation – PEReN



Pôle d'Expertise
de la Régulation Numérique

ISSN No. 2258-3106

Table of Contents

Table of Contents	2
Executive Summary	4
Introduction 7	
Chapter 1 Environmental issues created by the accelerate development of generative AI	9
1.1 The tangibility of generative AI services.....	9
1.1.1 Like all digital services, generative AI relies on hardware and physical infrastructure ...	9
1.1.2 Singularities tied largely to computing capacity needs.....	11
1.2 Market dynamics that require resources	12
1.2.1 Massive adoption of AI services and growing investments	12
1.2.2 Innovation dynamic with uncertain outcomes	14
1.2.3 Risks of tensions over resources creating environmental and economic issues	17
1.3 Expected net environmental effects in need of assessment	18
Chapter 2 Assessing the environmental footprint of AI	25
1.4 Analysis by ICT building blocks: data centres as the foundational component of the environmental impact of generative AI, and lingering uncertainties over the effects of networks and devices.....	27
1.4.1 Data centres: the foundational component of the environmental impact of generative AI systems.....	27
1.4.2 Questions about the medium-term consequences for devices and networks.....	35
1.5 Analysis by phase: growing usage that is increasing the impact of the inference phase relative to model training.....	39
1.5.1 Training phase: impact that depends on the number of model parameters, the energy mix and the equipment used	39
1.5.2 Inference phase: in addition to the impact per prompt depending on the number of parameters, overall impact rising sharply due to the rapid spread of generative AI	45
Chapter 3 Our recommendations for making AI development compatible with planetary boundaries	54
Axis 1: Improve the assessment and understanding of AI's environmental impact.....	54
Axis 2: Promote the ecodesign of AI services as a strategic lever for European competitiveness ...	59
Axis 3: Enable users to choose their generative AI services on the basis of their environmental impact through an appropriate European regulatory framework.....	65
Axis 4: Construct a development strategy for data centres in Europe that combines sovereignty and sustainability.....	68
Appendices 71	
1. Bibliography.....	71
2. Interviews	76
3. Glossary	77
4. List of acronyms.....	80
5. Orders of magnitude and equivalences relating to energy consumption and greenhouse gas emissions	82

6.	Comparison of the training of the GPT-3 and BLOOM models based on the literature review	84
7.	Performance of AI models according to training time or memory footprint for different tasks	85
8.	Mapping of standardisation initiatives for methods to assess the environmental impact of AI	86

Executive Summary

1. The environmental challenges created by the accelerated development of generative AI

Generative artificial intelligence (AI) represents a major technological disruption, enabling the automatic creation of text, images, audio and video, based on simple instructions, known as prompts, given in natural language. Its swift and massive adoption by consumers (48% of people in France were using it in 2025, compared to 20% in 2023) and businesses unleashes tremendous potential for production gains and innovation.

Generative AI can also enable innovative actions to further the environmental transition and reduce the environmental impact of multiple sectors. Nevertheless, it is itself becoming an accelerating factor behind a range of environmental impacts.

As with any digital service, generative AI relies on tangible infrastructures (data centres, networks and devices) and stands out for the massive computing power it requires. This computing capacity is hosted in data centres and is driving an exponential increase in the electrical power needed to run them.

The surge in demand and the market's development outlook are spurring stakeholders to jostle for top positions along the value chain, and to make massive investments to develop large-scale models and access crucial inputs such as computing power (chips, servers, data centres) and energy.

This innovation race could very well trigger growing tensions over scarce resources (energy, water, metals, rare earths). One especially pressing issue is the market dynamics that are forcing the increasing use of data centres, whose capacities need to be constantly evolving to adapt to ever-growing compute workloads.

These market dynamics raise questions about the sustainability of current AI development models, and the levers that are likely to help attenuate their environmental impact. To answer these questions, a rigorous and transparent assessment of the direct impact of generative AI would appear an indispensable first step towards understanding and managing the sustainability of its development.

2. Assessing the environmental footprint of AI

There continues to be a dearth of available data on the environmental impact of generative AI, not least because of a lack of transparency from industry actors: there is no environmental information available for 84% of all AI models. A better assessment of the environmental impact of AI along the value chain requires robust, reliable (multi-criteria, multi-component, multi-stage) and harmonised methodologies, to create the ability to compare models and services. Standardised and internationally recognised life cycle assessment (LCA) based methodologies do exist, and need to be mobilised and disseminated to avoid a proliferation of disparate methods and instead rely on a common basis.

To contribute to the assessment of the environmental footprint of AI, Arcep has drawn on the results found in the scientific literature, supplemented by institutional and industry work.

Arcep also conducted its own work with the Centre of Expertise for Digital Platform Regulation (PEReN), which provides groundbreaking results on the environmental impact of AI's inference phase.

The elements gathered for the assessment of generative AI's environmental footprint are broken down by ICT building block (data centres, devices, networks), then according to the model's training and inference phases.

The analysis by building block reveals that data centres are the most influential component in the environmental impact of generative AI. Their development raises questions relating to energy consumption and, by extension, greenhouse gas (GHG) emissions, which depend on the electrical mix of the country where they are located. By way of illustration, on average in 2024, the electrical mix in Poland had a carbon content that was 1.65 higher than that in Ireland, and 17.6 times higher than that in France. It also raises questions about water use and land artificialisation. The development of AI is, in fact, considered to be the main driver of the increase in data centres' energy consumption, which could double worldwide by 2030. Big Tech companies have already stated that the ability to reach their initial decarbonisation targets had been compromised by the development of generative AI, and have scaled back their targets. In addition to industry actors' strategies, the increase in data centres' power consumption creates challenges for individual countries and their decarbonisation trajectories.

Among the uncertainties are the environmental effects of generative AI associated with devices (generative AI services that are potentially energy-intensive during the inference stage and risk of having to replace devices) and to networks (whose capacity is theoretically sufficient in the short term, but whose adequacy over the medium term remains unclear).

The analysis by phase shows that the environmental impact of the models' training stage depends chiefly on the size of the model (number of parameters), on the electrical mix and hardware used (chipsets, servers). It also reveals that the number of models and their size are constantly growing.

The environmental impact of the inference phase is less well documented, but could become a significant issue as usage increases. To deepen the analysis, groundbreaking work was done by PEReN on the power consumption of prompts. It reveals that, during the inference phase, the larger models invariably consume more energy, but also that models of widely varying size can consume the same amount of energy. In addition, some more energy-efficient models are capable of providing responses that are just as relevant as the large models. Limiting energy consumption does therefore not necessarily mean having to compromise the model's performance. The tests conducted also reveal that optimisation and compression techniques can deliver sizeable energy gains during the inference phase.

3. Our recommendations to render the development of AI compatible with planetary boundaries

To ensure that AI and its infrastructures are deployed under conditions that are compatible with planetary boundaries and enable future generations to take full advantage of its potential, Arcep has formulated four strategic areas of work covering nine recommendations. These recommendations align with the existing and forthcoming European legislative framework.

- **Axis 1: Improve the assessment and understanding of the environmental impact of AI**
 - **Implement the collection and publication of AI environmental data by public authorities**
 - **Use internationally standardised environmental impact assessment methodologies** to facilitate comparisons between AI systems
- **Axis 2: Promote the ecodesign of AI services as a strategic lever for European competitiveness**
 - **Integrate the ecodesign of AI services into European regulation of providers**
 - **Strengthen eco-conditionality** in innovation support and public procurement rules

- **Axis 3: Enable users to choose their generative AI services on the basis of their environmental impact, through an appropriate European regulatory framework**
 - **Require greater environmental transparency** from chip suppliers and the main AI model and service providers
 - **Ensure AI services** remain open
- **Axis 4: Construct a development strategy for data centres in Europe that combines sovereignty and sustainability**
 - **Inform government and investment choices:** make transparency and data-driven regulation assets for the development of data centres in Europe
 - **Strengthen European coordination between digital, energy and infrastructure policies to support the development of data centres**
 - **Encourage a coordinated territorial approach of data centres deployment**

Arcep will continue this work, as well as its dialogue with all of the stakeholders, to work together on ecodesigned generative AI services and sustainable digital infrastructures.

Introduction

In only a few years, generative artificial intelligence (hereafter “generative AI”) has taken hold as a major innovation, and has been widely adopted by both consumers and businesses.

While artificial intelligence encompasses a set of technologies that have been steadily developed over several decades, and resulted in a set of tools and applications that are an integral part of our daily lives (including GPS and navigation, search engines, content recommendation, chatbots and fraud detection), the advent of generative AI marks a major turning point. Generative AI creates the ability to automatically generate new content: not only text, but also images, sound and video. Through a new generation of tools, such as ChatGPT, Gemini, Claude and Le Chat, simple instructions, known as prompts, in natural language, either written or spoken, make it possible to rapidly generate content that had previously required complex human tasks to produce.

The new automation, assistance and creation capacities unleashed by these tools suggest the possibility of substantial gains in productivity and innovation, making them potential growth drivers¹. The ability to master these technologies has therefore become crucial not only to safeguard competitiveness, but also sovereignty. Falling behind in artificial intelligence could create the risk of becoming dependent on other countries or regions of the world. These issues are thus driving massive support for the technology, and for the infrastructures that undergird them, from governments around the world. With its “AI Continent Action Plan”, published in 2025², the European Commission announced its aim to become a global leader in AI through a set of initiatives, including a 200 billion euro investment plan. In France, the Government launched a national AI strategy back in 2018, to structure the country’s AI ecosystem for the long term, at every stage of technological development (research, development and innovation, applications, market release and intersectoral take-up, providing support and a framework for deployment). Close to 2.5 billion euros of the France 2030 plan have been earmarked for AI³. This type of initiative will help bolster the development of artificial intelligence, and especially generative AI-specific capabilities and skills.

The plethora of opportunities and the already massive take-up of generative AI services amongst consumers and businesses, along with the development outlook for this technology in the coming years, naturally raise questions about its environmental sustainability. Although AI, and generative AI in particular, can help further the environmental transition and reduce the environmental impact of multiple sectors, it is nevertheless also proving to be a potential accelerating factor in a range of environmental effects. Chipset production, the construction of data centres, servers’ energy consumption during model training, the energy and water consumed by each prompt: far from being “intangible,” running generative AI services relies on very tangible hardware and has a very real impact on the environment.

This situation is creating legitimate concerns which, today, are being exacerbated by the lack of transparency from industry actors along the value chain. Having access to information is vital to ensuring that AI is deployed under conditions that are compatible with planetary boundaries, thereby allowing future generations to benefit from its potential.

¹ In the report “AI: our ambition for France” published by the AI Commission in March 2024, A. Bouverot and P. Aghion estimate, for instance, that the automation of certain tasks could double annual economic growth in France with, an increase in GDP of between €250 and 420 billion in 10 years.

² [AI Continent Action Plan | Building Europe’s digital future](#)

³ [National strategy for artificial intelligence](#)

With this as its starting point, in its “Ambition 2030⁴” strategy, Arcep committed to “*deepening and sharing knowledge of the environmental impact of artificial intelligence*”. This action is an integral part of the work that Arcep has done in recent years on the environmental footprint of information and communication technologies (ICT), including the ADEME-Arcep report on the digital environmental footprint in 2020, 2030 and 2050⁵, and the collection of environmental data from digital industry actors which, since 2022, have been made publicly available as part of the annual Achieving digital sustainability survey⁶. By helping to increase understanding of the environmental impact of ICT, based on reliable data and methodologies, Arcep plans to propose levers for action that will help reduce this impact.

This report marks a new stage in Arcep’s commitment to obtaining a better assessment of the environmental impact of AI, and generative AI in particular. This work draws on interviews with a number of experts from the private and public sector, academia and NGOs, coupled with a review of the scientific and institutional literature. The report also draws on a groundbreaking report produced in collaboration with the Centre of Expertise for Digital Platform Regulation⁷ (PEReN), which made it possible to carry out tests on the energy consumed by several generative AI tools during the inference phase.

This report is structured as follows:

- It begins with an analysis of the environmental issues and challenges raised by the tangibility of generative AI and its massive boom, on both the supply and demand side;
- Next, it delivers a synthesis of current understanding of the environmental footprint of generative AI, enabling initial assessments and the ability to share a reliable knowledge base. The results drawn from academic literature, institutional and industry sources were supplemented by new findings from the work carried out in collaboration with PEReN, making it possible to analyse the key factors in generative AI models’ energy consumption during the inference phase;
- Lastly, based on this diagnosis, the Authority has formulated four strategic areas of work:
 - o Axis 1: Improve the assessment and understanding of AI’s environmental impact
 - o Axis 2: Promote the ecodesign of AI services as a strategic lever for European competitiveness
 - o Axis 3: Enable users to choose generative AI services on the basis of their environmental impact through an appropriate European regulatory framework
 - o Axis 4: Develop a European data centre strategy that combines sovereignty and sustainability

Translated into nine recommendations, these priorities seek to reconcile the development of AI, notably generative, AI, and sustainability, to ensure that this technology is deployed under conditions that benefit the whole of society.

This report thus completes Arcep’s work on how the development of generative AI affects the ability to maintain an open internet, which was the topic of a report published in January 2026⁸.

⁴ [Arcep Ambitions-2030 objectifs actions janv2025.pdf](#)

⁵ [ADEME – Arcep study on the environmental footprint of ICT in 2020, 2030 and 2050 | Arcep](#)

⁶ [Annual "Achieving digital sustainability" survey – 2005 edition \(2023 data\) | Arcep](#)

⁷ The Centre of Expertise for Digital Platform Regulation/Pôle d’expertise de la régulation numérique (PEReN) is a government service with national authority. It assists public authorities in the regulation of digital platforms and AI, by lending sovereign, shared and state of the art technical expertise. From algorithm audits to data analysis, by way of developing innovative tools; whenever possible, PEReN shares its output as open source products, for the sake of transparent and collaborative regulation.

⁸ [Generative AI: challenges for the future of the open internet](#)

Chapter 1 Environmental issues created by the accelerate development of generative AI

Like any digital service, AI runs on physical infrastructures and hardware, ranging from user devices to data centres, by way of electronic communications networks. This materiality of the infrastructure supporting generative AI services, combined with the rapid rise in the use of these services and intense market dynamics, are generating growing environmental issues and challenges. While it is true that AI can deliver encouraging environmental protection prospects, these opportunities must not ignore the adverse effects tied to these dynamics.

This chapter begins by exploring the tangible nature of generative AI services (1.1), as digital services underpinned by physical infrastructure and hardware (1.1.1), then goes on to describe the technical and tangible features of this technology that could escalate the environmental issues associated with these services (1.1.2). The chapter then analyses the issues resulting from the generative AI boom (1.2), how the massive adoption of this technology and the considerable investments it is attracting (1.2.1) are driving fierce competition dynamics at the different levels along the value chain (1.2.2). These market dynamics can create tensions over certain critical resources and become a real environmental issue (1.2.3). The final section examines the net environmental effects of artificial intelligence, which remain difficult to assess at this stage, largely due to the complexity of the indirect effects of ICT and the still fledgling nature of research in this area (1.3).

1.1 The tangibility of generative AI services

1.1.1 Like all digital services, generative AI relies on hardware and physical infrastructure

As with any digital service, generative AI services run on physical infrastructures and hardware that can be classified by three main categories or building block: user devices, networks and data centres.

Every digital practice – such as browsing social networks or a website – elicits all three of these building blocks. For a service to function, it requires a user device (e.g. smartphone or computer) on which it will be accessed. Although this building block is typically the most tangible, networks (ADSL, FttH internet routers, etc.) and data centres (that house servers) are also needed to relay, process and store the data, to be able to provide a digital service. These three building blocks are interdependent and can influence one another, which is why they must not be analysed in isolation. By way of example, the smartphone boom and the rapid expansion of the mobile app developer ecosystem gave rise to a range of services such as video streaming, video calling and online gaming, along with digital tools such as online banking and e-government services. This in turn drove up data traffic, which influences the scaling of networks and data centres.

INTERDEPENDENCE DYNAMICS BETWEEN INFRASTRUCTURES AND SERVICES

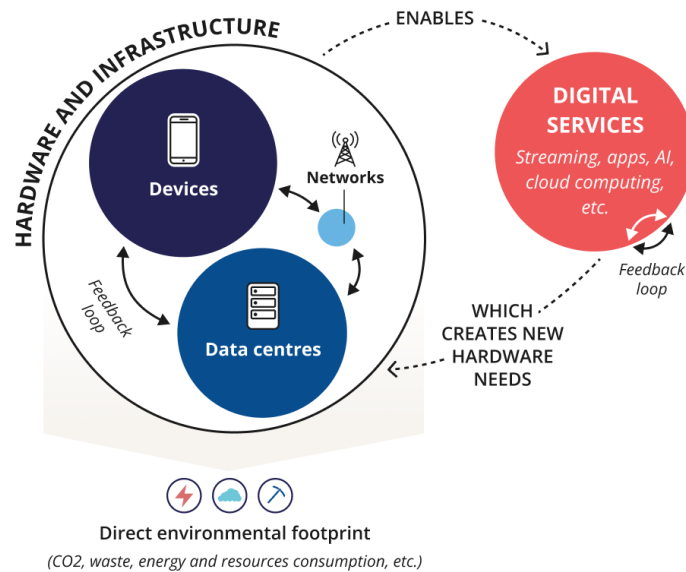


Figure 1: Interdependence between digital infrastructures and practices. Source: Arcep.

A number of studies have been produced by a range of academic, private sector and institutional entities over the past several years on the environmental impact of ICT. These studies have helped to fill in some of the knowledge gaps on the subject. Although generative AI is not taken specifically into account in these studies – as its rise is an only recent phenomenon – their findings remain relevant for understanding and assessing the environmental impact of digital services as a whole.

In particular, the joint study conducted by ADEME and Arcep at the request of the French Government on the Digital environmental footprint in 2020, 2030 and 2050⁹ helped to highlight the relevance, when analysing the environmental impact of ICT, of an approach that must be:

- **Multi-stage:** assessing the impact of only the operational phase is not enough, as all stages of the hardware’s life cycle need to be taken into account (production, distribution, use, end of life);
- **Multi-component:** the infrastructure’s three building blocks (devices, networks and data centres) must be analysed;
- **Multi-criteria:** the environmental impact of ICT is not confined to greenhouse gases: water use, energy consumption, metal and scarce resource depletion also need to be analysed.

Using this approach, the study shows that digital technology accounted for around 4% of the national carbon footprint in France in 2022, and around 10% of national electricity consumption. It underscores that digital services and the “digitalisation” of communications that it enables, rely on infrastructures and devices that are both very tangible and consumers of resources.

⁹ To find out more about the environmental impact of ICT: [ADEME – Arcep study on the digital environmental footprint in 2020, 2030 and 2050 | Arcep](#)

The forward-looking analysis also highlights the fact that, if nothing is done to curtail the growth of ICT’s environmental impact in France, its carbon footprint could triple between 2020 and 2050, and energy consumption could double. This progression, which was forecast even before the recent generative AI boom, seems therefore clearly incompatible with France’s goal of a stabilised ICT carbon footprint by 2030-2035, which is the target stipulated in the “National low-carbon strategy”¹⁰.

The growth momentum of digital practices also raises pressing questions about the depletion of metals and other strategic resources and their future availability, not only for ICT but for other sectors as well, given that demand for metals for ICT use could increase by 179% between 2020 and 2050 (ADEME, 2024b). The consumption of these resources for information and communication technologies can also be observed through the generation of waste. By way of example, the joint study from ADEME and Arcep estimates that, for their digital practices alone, a person living in France in 2020 generated 300 kilos of waste per year (including electrical and electronic waste, as well as waste tied to raw materials extraction).

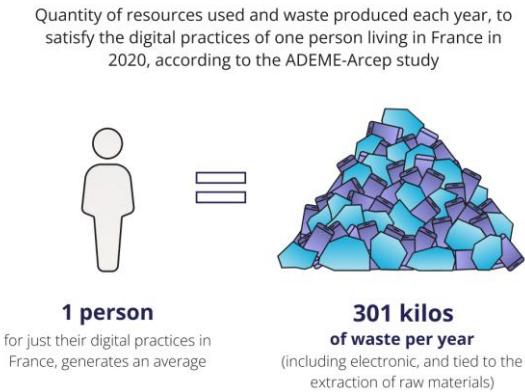


Figure 2: Digital goods and services generate waste. Source: Arcep.

The ADEME-Arcep study served to establish a shared diagnosis and to sound the alarm on the trajectory of ICT’s environmental footprint. While it provides methodological elements for identifying the sources of pollution associated with the training and use of generative AI services, it does not provide a detailed view of the impact by technology or by service. Other studies will need to be conducted specifically to isolate AI’s share of ICT’s environmental footprint.

The fast-paced development of generative AI over the past several years (see section 1.2.1) makes it vital to immediately begin to address the specific question of its environmental impact, especially as it embodies singular technical and economic features, compared to other digital services, that could affect its environmental impact.

1.1.2 Singularities tied largely to computing capacity needs

Generative AI differs from a “classic” digital service in particular because it requires massive computing power to create, train and fine tune AI models, but also to use the services based on generative AI.

Generative AI services are based on foundation models that require massive computing power during two key phases:

¹⁰ National low-carbon strategy/Stratégie nationale bas-carbone (SNBC3), [Summary of the SNBC3 project published in December 2025](#).

- model training, which refers to the foundation model's initial training process, during which data feed the model, and which can be supplemented by a fine-tuning phase, during which the model will be tailored to perform a specific task. During training, results are examined and the model's output is adjusted to improve its performance (quality of responses) and efficiency (computing time and quantity, energy costs, etc.);
- inference, which corresponds to users' actual use of the trained model, to create content.

During training, a generative AI model is exposed to massive datasets (e.g. texts, images, audio) to teach it the constants, which are the underlying structures that characterise them. This phase requires high-capacity computing resources to ensure that the model will be able to generate coherent and relevant content.

Once trained, the model is integrated into a service and used to respond to users' prompts and generate content (e.g. text or images, depending on the use case). This inference (i.e. daily use) phase typically consumes less computing power, but that consumption is tied directly to the volume of use. If each inference taken separately consumes far fewer resources than training, it is carried out on a very large scale (potentially billions of times) which requires massive overall computing power.

In physical terms, the computing power required not only to train the model, but also for its inference, corresponds to servers that are typically hosted in data centres, and powered by electricity¹¹. The AI boom is contributing to an exponential surge in the electrical power needed to run IT hardware and infrastructures, such as the graphics cards used in servers, and recent projects for data centres hosting these servers. Current data centre projects for AI thus represent, for the largest amongst them, gigawatt-scale¹² (GW) draws. To compare, projects that preceded the recent rapid rise of AI have a draw of several dozen or even hundreds of megawatts for the largest amongst them. At the same time, new central processing units (CPU) and graphic processing units (GPU), which are central components in AI servers, have increased their energy draw dramatically in recent years: some companies are planning on hardware with consumption that could reach 1400 W by 2026 and even more beyond that, compared to hardware that drew 300 to 400 W in 2017¹³.

1.2 Market dynamics that require resources

1.2.1 Massive adoption of AI services and growing investments

In addition to its particular physical requirements, generative AI also distinguishes itself from other digital services in the extremely rapid and massive development of its usage. According to the 2026 edition of the Digital Barometer (Arcep, 2026c), generative AI services enjoyed an unprecedented rate of take-up in only two years. In 2023, a fifth of the French population said they were using AI (20%), a percentage that climbed to close to half the population in 2025 (48% to be precise), or a 28-point rise in two years. And this is a global phenomenon. The speed of adoption for generative AI in France is comparable to that observed in the US (Bick et al., 2024).

The progression in usage is also being seen in the workplace. In the span of two years, the percentage of EU companies with a staff of 10 or more that report using at least one AI service, rose from 8% in

¹¹ Some companies plan on running the inference phase on the prompting device, but this would require computing power that a great many consumer devices do not yet possess. This is why the term "AI-compatible" is starting to be seen, especially for smartphones.

¹² By way of comparison, this corresponds to the order of magnitude of a nuclear unit.

¹³ Gauthier Roussilhe, Les GPU et les IA génératives : une nouvelle phase de l'histoire environnementale de la numérisation, March 2025

2023 to 20% in 2025, or by 12 points in two years. Larger companies, with a staff of 250 or more, stand out for having an even more rapid pace of adoption: going from 30% in 2023 to 55% in 2025¹⁴.

The expected boon from the development of this technology is reflected in the substantial and growing capital expenditures (Capex) by AI service providers. They are having to contend with very high costs, notably to access the computing power required (for AI-dedicated servers¹⁵ but also for the data centres to host them). To illustrate: for 2026, Amazon, Google, Meta and Microsoft announced a combined Capex of around 650 billion USD in AI, having spent 410 billion USD in 2025 and 245 billion USD in 2024¹⁶. If these colossal sums might raise questions about the ability to rapidly translate those investments into revenue (as business models remain uncertain), the sector's leading actors¹⁷ have the financial and debt capacities that allow them to assume these investments without necessarily having to worry about short-term profitability. As it stands, these Big Tech companies' business models are still diversified and in the process of being structured (see 1.2.2), and testify to their focus on promoting the massive adoption of their services, especially some of the most popular such as chatbots, which usually have a free version.

Aside from the announced investments from the private actors themselves, governments are also supporting this dynamic and joining the AI development race, aiming to capitalise on the economic potential of AI in terms of productivity gains and innovation, while also seeking to safeguard their sovereignty over key technologies and infrastructures. One prime example, the AI Action Summit in Paris, in February 2025, provided an opportunity for major investment announcements: 109 billion euros over the coming years in France. At that same summit, European Commission President, Ursula von der Leyen, also announced plans to earmark 200 billion euros, including 150 billion from a private-sector consortium called the "AI EU Champions Initiative", to make Europe a leading AI continent, saying that "*the race for leadership in AI is far from over*"¹⁸. This investment scheme is the cornerstone of the *AI Continent Action Plan*¹⁹ published in April 2025 and detailing the different measures, ranging from the installation of very high-capacity IT infrastructure dedicated to AI, to promoting AI in strategic sectors, by way of training highly skilled AI workers. These European announcements came on the heels of announcements from Stargate, a 500 billion-dollar programme backed by the Trump administration in concert with Tech giants, to bolster private investment in AI development in the United States.

Faced with this boom, some institutions are now voicing reservations and a desire to adjust the AI support strategy. In France, for instance, the Oversight Committee for Future Investments (CSIA), has observed an acceleration in spending on AI around the globe, which raises a number of questions "*over a possible bubble, as the business models for AI investments are not yet established*"²⁰.

¹⁴ [Eurostat](#), latest update: 27 February 2026

¹⁵ According to [Cyfuture cloud \(Cyfuture cloud, site accessed in April 2026\)](#), in 2025, a dedicated AI server, with eight "H100" processors can cost between 200,000 and 400,000 dollars, keeping in mind that several dozen if not hundreds of these servers would be required, along with storage and network capacity, to have an AI cluster capable of training a generative AI model (depending on its size).

¹⁶ https://www.lemonde.fr/economie/article/2026/02/07/intelligence-artificielle-la-folle-course-des-geants-de-la-tech_6665758_3234.html

¹⁷ In addition to Amazon, Google, Meta and Microsoft, model developers such as OpenAI, Mistral AI and Anthropic all have strategic partnerships with these Big Tech companies that enable them to benefit from venture capital funding.

¹⁸ [Artificial intelligence: Ursula von der Leyen announces 200 billion euros in investments in Europe](#)

¹⁹ [The AI Continent Action Plan | Shaping Europe's digital future](#)

²⁰ [rapport-2025-du-csia-france-2030.pdf](#)

1.2.2 Innovation dynamic with uncertain outcomes

The AI services market is booming. Growing from 189 billion dollars in revenue in 2023, it is forecast to be worth 4,800 billion dollars in 2033 according to the UN Conference on Trade and Development (UNCTAD). This means that AI could quadruple its share of the global high-tech market during that time, ballooning from 7% to 29%²¹. With this outlook in mind, the current race to innovation and services take-up is creating very strong incentives for market actors to seek dominant positions along the entire value chain and to make substantial investments in essential inputs, such as access to energy, computing power, data and technical expertise.

The generative AI value chain can be divided into three layers: infrastructure, modelling and deployment. It is currently populated by a large number of companies spread out amongst the different services. Depending on the layers, the inputs and investments needed will differ, and so result in varying degrees of market concentration. Some market segments – notably at the infrastructure level, such as supplying specialised AI chips (over which Nvidia has a virtual monopoly) and cloud computing services (which has been cornered largely by Amazon, Microsoft and Google) – appear to be more concentrated than others, particularly for the supply of AI-integrated services.

In its opinion on the competitive operation of the generative AI sector²², France's *Autorité de la concurrence* (Competition authority) underscores that the value chain is occupied by vertically integrated Big Tech companies which are therefore capable of controlling design, training and inference, accompanied upstream by myriad service providers, enterprises that integrate AI in their products and services. At the top of the pyramid is Big Tech: Alphabet (Google), Microsoft, Amazon, Meta and Nvidia. Some, such as Microsoft and Alphabet, are vertically integrated across the entire value chain, while others are focused on key sectors (GPU, cloud computing, data ingestion). Model developers such as OpenAI, Mistral AI and Anthropic also have strategic partnerships with the industry heavyweights, thereby increasing the sector's concentration. Lastly, companies positioned in adjacent markets such as device sales (e.g. Samsung, Apple) already integrate AI into their products and services to maintain their appeal.

²¹ UN Conference on Trade and Development (UNCTAD), 2025. AI

²² [Opinion, 24-A-05, 28 June 2024, Autorité de la concurrence](#) (in French)

Panorama of financing methods for generative AI services as of 13 april 2026

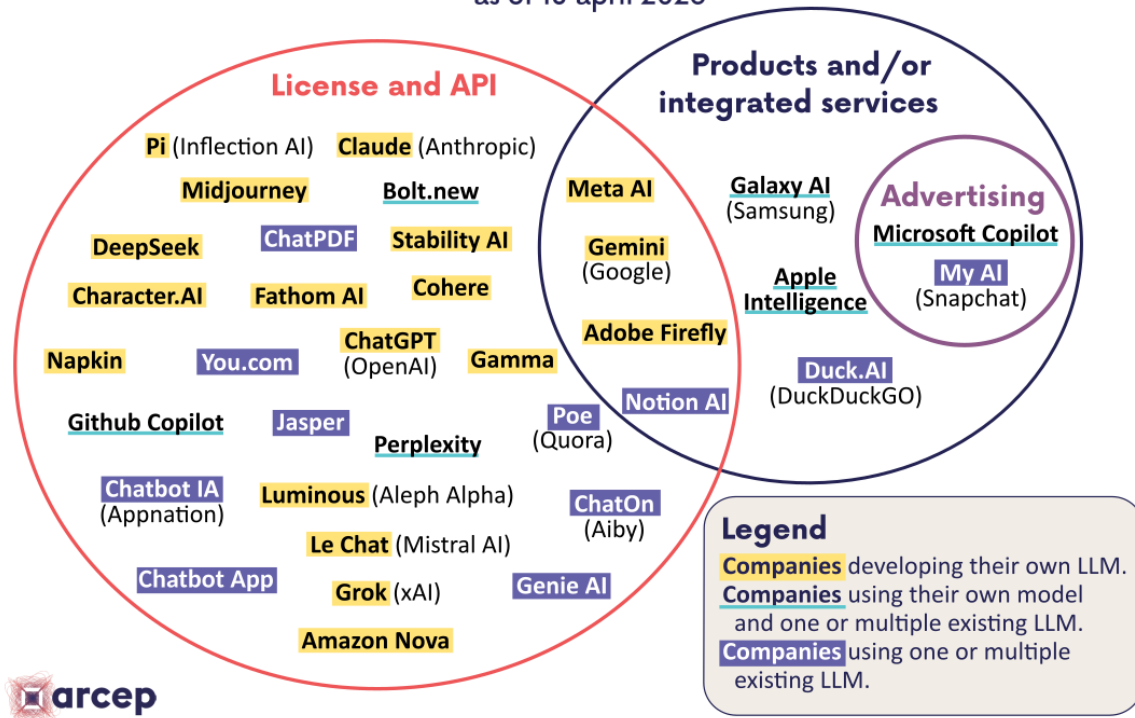


Figure 5: Panorama of financing methods for generative AI services. Source: Arcep

The market dynamics resulting from the innovation race therefore still contain a great many uncertainties over the market's structure over the long term. Today and for the foreseeable future, these dynamics mean intense competition driving up Capex and the consumption of resources. The effects of these dynamics are not confined to economics, but also hold environmental repercussions.

1.2.3 Risks of tensions over resources creating environmental and economic issues

As mentioned in section 1.1, far from being intangible, artificial intelligence relies on physical infrastructures and the consumption of natural resources. As a result of which, competition dynamics around the development of generative AI, which include an innovation race, a growing need for computing capacity and a steady increase in investments, are all consequential for the environmental impact of generative AI and can generate growing tensions over certain strategic resources.

The market dynamics described above involve increased use of data centres whose capacities need to continually adapt to the fast-rising computational load, along with steady demand for specialised services, notably for the training and inference of large-scale generative AI models. They can also translate into increased network equipment, depending on the chosen architecture (centralised or edge computing) or the early trading in of user devices for AI-compatible equipments.

These evolutions induced by the generative AI sector's competition dynamics rely on particularly resource-intensive supply chains. The production of semiconductors, which are ubiquitous in all three digital infrastructure building blocks (devices, networks and data centres) consume massive quantities of energy, high purity water and chemical products, which have a significant environmental impact prior to their utilisation, especially for the most powerful semiconductors (Roussilhe, 2021). By the

same token, growing demand for digital hardware is creating increasing tensions over critical raw materials. The extraction and refining of these resources often go hand in hand with tremendous environmental degradation, particularly in terms of soil pollution, water consumption and biodiversity loss. In this environment, the ongoing massive investments and increasing computing power are only exacerbating tensions over these resources, while also accentuating the environmental impact on a local scale.

Moreover, the geographical concentration of data centre infrastructures, for reasons of economies of scale and access to both energy resources and abundant connectivity, is also likely to exacerbate competition over access to local resources (notably energy, water and land), including increased tensions over infrastructures such as the electrical grid (e.g. capacity reservation) and potentially conflicts with other sectors over their use. These developments are thus part of a broader context of energy choices and coordination, where growing demand spurred by generative AI is likely to compete with other energy uses, and so raising the question of its compatibility with energy transition trajectories and public policies supporting the electrification of the economy.

In this context, the generative AI sector's most powerful actors may seek to pre-empt strategic resources, be it access to the energy required to power data centres or the latest generation of semiconductors that are crucial to the development of generative AI models, thereby accentuating environmental stresses. Here, the frugality of generative AI models could become both a major distinguishing factor between competing companies, and an asset to promote the sustainable development of generative AI. This incentive to create frugal systems could be undermined by the existence of highly concentrated AI services markets controlled by a small handful of heavyweights with a stranglehold on access to inputs (computing power, data, talent, energy).

Lastly, in addition to infrastructures and resources, the development of AI engenders systemic effects that are likely to amplify these tensions over resources. One of the main mechanisms at work is the rebound effect (see inset in section 1.3): the efficiency gains and decreased cost of accessing these technologies, some of which are currently being offered for free, stimulates their rapid dissemination, which in turn leads to an overall increase in demand for computing power and data. By way of example, because of the massive popularity of generating images inspired by Studio Ghibli-style animations, OpenAI was forced to limit access to its generative AI-based image generator tool, as offering it for free led to an energy draw that was too heavy for the company's GPUs to sustain. This example illustrates that the resources mobilised to provide generative AI services, whether computing power, electricity or infrastructures, cannot therefore be viewed as unlimited. The growing integration of AI in the many services that are used on a daily basis – from search engines to productivity tools – thus contribute to the exponentially growing demand for resources.

An environmental analysis of the effects of the economic mechanisms described above brings to light a set of direct and indirect effects that are closely bound up with the economic dynamics of artificial intelligence. Far from being marginal, these implications – of which a prime example is the tensions over limited resources – raise questions over the sustainability of the current AI development models, and the levers that might help attenuate the environmental impact. To this end, a rigorous and transparent assessment of this impact appears to be an indispensable prerequisite to an informed strategy for making this development path sustainable.

1.3 Expected net environmental effects in need of assessment

While digital technology generates direct effects, which correspond to the environmental footprint of the different phase of ICT's life cycle (see **Erreur ! Source du renvoi introuvable.**), it can also have indirect effects. Contrary to direct effects, indirect effects can have an either positive or negative environmental impact. In concrete terms, these are the effects that the use of ICT will have, in

particular on human behaviour and production processes (see inset “Understanding the direct and indirect effects of ICT” for more details).

One example of a positive indirect effect of ICT is efficiency, in other words the ability to optimise a process by digitalising it. One well-known use case for consumers is home automation, by adding connected equipment to the home. This creates the ability to optimise the functioning of heating equipment and so reduce the home’s energy bill and associated greenhouse gas emissions (GHG).

Multiple use cases of digital solutions, particularly in industrial and technical fields, can generate environmental gains, according to a recent ADEME study (2024a; 2025b). The study reveals, for instance, that the Dynamic Line Rating (DLR), for optimising an electrical system’s management, almost systematically leads to reduced GHG emissions in every one of the configurations tested²⁷.

In some cases, however, the energy or resource savings enabled by the use of new digital technologies can also lead to increased consumption. This rebound effect describes the situation where increased efficiency helps reduce the cost (energy or resource savings) of a given use and, as a result, drives an increase in usage that can cancel out the expected gains, or even have a negative impact if the increase in usage outweighs the original savings²⁸. To go back to the previous home automation example, optimising a home’s heating system helps reduce its energy costs (financial and environmental), but the resulting savings may create an “incentive” for the homeowner to turn up the usual temperature level and thereby cancel out the initial gains.

ADEME therefore delivers a reminder that, in certain instances, the net impacts of the digital solution in question will depend on the configuration and context in which the solution is implemented. For instance, in the case of connected street lighting, the net impacts are negative when the scope of assessment includes the environmental risks tied to the rebound and induction effects (see inset “Understanding the direct and indirect effects of ICT”). As a result, ADEME underscores the difficulty of extrapolating the results of a use case on a national scale, and so the impossibility of drawing conclusions on the effects of digitalisation on a global scale.

There is currently no scientific consensus on the net environmental impact of ICT, in other words on the net environmental footprint of its (positive or negative) effects. In particular, in the case of generative AI – which is undergoing very rapid technological development and whose adoption is still only nascent – no study to date has enabled a conclusion on its net environmental impact, not least because these evaluations must be made based on a use case and a specific deployment context for an AI solution.

Assessing the net environmental effects of ICT is therefore a complex undertaking. In practice, the massive dissemination of digital technologies limits the ability to make these evaluations, which require very large datasets, both on digital technology, but also on the multiple sectors in which digital solutions are deployed. This task is rendered all the more complex in a context where transparency is lacking. A case-by-case assessment therefore appears more relevant, as it is the only way to fully

²⁷ In France, the required distance between an electrical cable and different obstacles that could be found underneath it are regulated for safety reasons.. Weather conditions and the electrical current flowing through a cable affect its temperature, causing it to expand or contract and thereby altering its clearance from the ground. DLR systems create the ability to optimise the maximum transfer capacity of power lines’ current in real time (contrary to the Static Priority Rating which sets a value by taking fixed, conservative weather conditions into account). From a concrete perspective, DLR thus makes it possible, e.g. during windy weather that cools the cable, not to cap production on a wind farm unnecessarily, and foster the incorporation of renewable (hence decarbonised) energy sources in the network.

²⁸ We speak of the Jevons paradox when efficiency gains are quantitatively more than offset by the increase in consumption (referred to as a rebound effect), to the extent that total resource consumption exceeds the levels that existed prior to the appearance of the efficiency gains enabled by a technological innovation.

incorporate the indirect effects of a digital solution, thereby avoiding extrapolations to several insufficiently robust use cases. This analysis must factor in the digital solution's indirect and systemic effects²⁹. To achieve this, it is preferable to perform a consequential life cycle assessment (C-LCA), aimed specifically at understanding the overall consequences of a change (ushered in by a digital solution) rather than an attributional life cycle assessment (A-LCA), that provide a simple static picture of the impacts.

²⁹ G.Roussilhe explains that the problem with certain studies on the emissions avoided thanks to ICT is the extrapolation based on case studies, the representativeness of the selected case studies, and the lack of incorporation of the impacts of the life cycle and the structural impacts (Gauthier Roussilhe, Les émissions évitées de la numérisation, September 2023).

Attributional vs. consequential LCA: objectives and differences

Attributional LCA (A-LCA)

Attributional life cycle assessment (A-LCA) is a method used to assess the environmental impact of a system as it exists in a given moment in time. It is based on a descriptive and static approach: the goal is to estimate the environmental footprint associated with a – for instance, digital – product or service over the course of its life cycle (production, distribution, use, end of life). The impacts are typically spread out according to allocation rules.

Examples: Estimating the environmental footprint of a smartphone by incorporating the impacts tied to the different stages of its life cycle, or estimating the environmental footprint of one hour of streaming by allocating the impacts of the digital infrastructure needed (the device on which the content is streamed, the networks relaying the data, the data centres hosting the content) via allocation rules.

This approach is used in particular to compare products or establish footprint assessments, but one of its chief limitations is that it does not factor in changes or the indirect consequences of changes in usage.

Consequential LCA (C-LCA)

Consequential life cycle assessment (C-LCA) seeks to assess the environmental impact induced by a decision or a change, rather than to describe an existing system. It adopts a dynamic and comparative approach, focusing on the consequences of one scenario compared to another (e.g. the deployment of a digital solution compared to a situation where that solution does not exist). This method enables the incorporation of the effects of the studied system on its external environment, beyond the direct effects.

Example: What happens if a portion of in-store sales shift to e-commerce? If, in theory, this reduces the number of individual trips (by pooling trips with deliveries), it may, for instance, generate an increase in consumption thanks to an easier shopping experience. In other words, consequential logic must be applied, looking beyond just a snapshot of the system.

C-LCA is thus especially well-suited to help in decision-making and public policy or innovation assessment, but the results can be less tenable as they depend heavily on the chosen assumptions, notably in the scenarios selected.

Understanding the direct and indirect effects of ICT

The positive or negative environmental effects of ICT can be either direct or indirect.

The direct environmental effects of a digital solution correspond to the environmental impact induced by the life cycle of the digital hardware and infrastructures required for the solution to run. This therefore includes the impact of the user devices, networks and data centres mobilised to provide the service through the life cycle of the three building blocks (raw materials extraction, production, distribution, use and end of life).

The indirect environmental effects of a digital solution correspond to the environmental externalities generated by the deployment of digital technologies. This encompasses notably technical (e.g. using AI to facilitate smoke/fire detection) or societal (e.g. using AI for targeted advertising alters consumption behaviour) changes, and can be positive or negative.

Indirect effects can lead to a reduced environmental impact compared to a benchmark situation where the digital solution in question is not deployed. The positive indirect effects can include:

- **substitution effect:** a substitution effect with a positive environmental impact occurs when one technology is replaced by another more environmentally friendly one;
- **optimisation effect:** optimisation or efficiency effects correspond to changes made to a solution that allow it to use fewer resources (raw material, energy, etc.) to produce or consume the same quantity.

Indirect effects can also translate into an increased environmental impact compared to the benchmark situation. Negative indirect effects can include:

- **induced effects:** which correspond to situations where the digital solution induces new uses or the use of new resources (e.g.: innovations enabled by printers induce increased paper use);
- **rebound effects:** which correspond to situations where the efficiency gains enabled by a digital solution (e.g. smart heating system) lead to increased consumption of the activity that benefitted from the digital solution or of another activity (e.g. reallocating the savings on heating to another activity).

The net impacts (that factor in the direct and indirect effects) are therefore relative to a benchmark situation, which involves comparing scenarios where the digital solution is deployed with what would have happened had it not been deployed. **As ADEME stated in a recent report (ADEME, 2024a) on the environmental assessment of direct and indirect effects of ICT for use cases, the definition of the implementation context of each solution is therefore crucial when carrying out a consequential life cycle assessment (C-LCA). This requirement thus renders any general extrapolation of the findings on the net environmental impact of ICT very difficult and incomplete.**

As a result, while an assessment of the net effects is already a complex undertaking for older ICT innovations, it is all the more so when it comes to generative AI, due to its very rapid evolution and to a lack of the data needed to carry out these assessments. It is therefore difficult, as it is with ICT in general, to obtain a conclusive assessment of the net environmental effects of generative AI, whether positive or negative, by adopting a general conclusion.

The technical progress made by data centres³⁰- essential infrastructures for the development of generative AI – can be observed today through factual data (see inset “Spotlight on the work done on data centres in the annual “Achieving digital sustainability” survey in section 1.4.1), and helps to reduce the environmental impact of generative AI. The indirect effects of generative AI can, however, vary a great deal depending on the sectors in which it is deployed, and even within the same sector. In the energy sector, for instance, net effects can be negative when generative AI facilitates the extraction of fossil fuels, but are likely to be positive when it is deployed to support the development of renewable energies and their integration into the electrical grid. In some sectors, then, AI can help reduce GHG emissions (e.g. electricity, meat and dairy, light-duty vehicles) but Stern et al. (2025) advance that leaving it up to the market to determine AI applications and its governance could be risky. They conclude that interventionist governments are key to the “proper” use of AI in these sectors, as market forces alone may not suffice to drive the changes needed to unleash the full potential of AI, while also achieving climate targets.

Ultimately, under the current circumstances, analysing the net environmental effects of ICT (and especially of generative AI) involves having to contend with sizeable methodological limitations. The diversity of uses, the heavy dependence on sectoral contexts and indirect effects, such as rebound effects, make any overall assessment particularly uncertain and, at this stage, largely exploratory.

UNCERTAINTIES WEIGHING ON THE ASSESSMENT OF THE POSITIVE AND NEGATIVE EFFECTS OF GENERATIVE AI

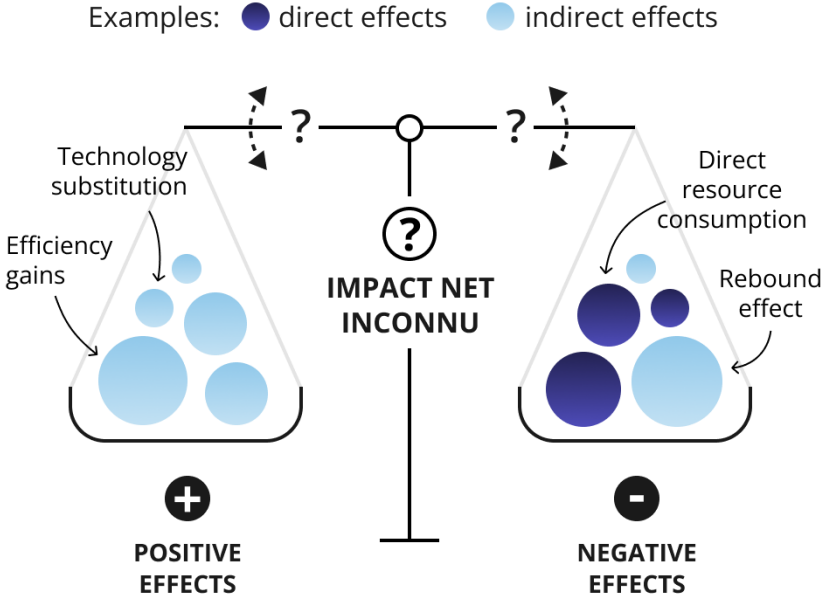


Figure 6: Illustration of the direct and indirect effects of generative AI. Source: Arcep.

³⁰ For instance, over the past 15 years the urbanisation of data centres’ server rooms, in other words the physical and logical organisation of IT infrastructure, notably with hot and cold aisles, has made it easier to cool them, and so to improve data centres’ overall energy efficiency.

This difficulty must not, however, be used to justify inaction. At the same time, the scale of the dynamics at play, starting with the rapid growth of the energy needs induced by generative AI (AIE, 2025), make these issues both pressing and foundational.

Without prejudging the net long-term environmental effects of generative AI, increasing knowledge and understanding of its direct environmental impact, which is still insufficiently documented, is of immediate necessity, to inform public debate as fully as possible.

*
**

Generative AI thus highlights, in an especially stark fashion, the materiality of digital technologies: their development relies on physical infrastructures, high-capacity computing power and massive investments in both data centres and IT components. The speed of its adoption and the intensity of the competition dynamics surrounding the development of AI are accelerating its requirements, and running the risk of exacerbating tensions over certain precious resources and amplifying the associated environmental issues.

Under these circumstances, while assessing the net environmental effects of AI remains a very difficult and complex undertaking, a vital first step would be to examine the direct environmental effects of AI, which are addressed in the next chapter.

Chapter 2 Assessing the environmental footprint of AI

This chapter provides a synopsis of current knowledge of the environmental footprint of generative AI. It provides initial assessments and shares an initial, reliable knowledge base, which can springboard levers for action. This work is particularly important in light of the very limited amount of data that is currently available, which is due in large part to a lack of transparency from the companies along the generative AI value chain, starting with model suppliers (see inset below). To contribute to the assessment of the environmental footprint of AI, Arcep thus drew as much as possible on the findings in scientific literature, supplemented by institutional and industry findings. Having ascertained the lack of reliable information, particularly on the inference phase, Arcep also collaborated with the Centre of Expertise for Digital Platform Regulation (PEReN), which provided groundbreaking results on the environmental impact of the inference phase.

AI models' lack of environmental transparency

Assessing the environmental impact of AI requires access to a reliable body of information, obtained from the suppliers themselves, which are the only ones to have a 360° view of their production.

As it stands, however, there are very few exact and reliable figures available on the different indicators on the environmental footprint of the models and use across their life cycle. Big Tech companies in particular do not share information on the environmental impact of their AI model development, training and inference operations. Luccioni et al. (2025) thus revealed that, of the 754 models analysed, which were deployed between 2010 and Q1 2025, 84% have not provided any environmental information, while environmental information is indirectly available on 14% of the models (e.g. data on the training or size of the model which are part of the technical but not environmental information available, but which can be used to carry out an environmental assessment), and only 2% of models provide information directly (e.g. energy consumption or carbon intensity assessments) as of May 2025. Information therefore remains very piecemeal.

Because of this lack of transparency from industry actors, very little assessment has been carried out on the environmental impact of AI along the value chain (Ligozat et al., 2022), which in turn limits the ability to anticipate and control said impact. Industry actors could already provide more environmental information on their models' entire life cycle, particularly on the training and inference phases, but the current lack of environmental transparency tends to prove that "natural" incentives to be transparent do not suffice. There may be several reasons for this situation, particularly the heavily imbalanced state of competition, the lack of shared methodology, and a poor accounting of demand-side issues.

Despite this lack of transparency from AI models, academic work has been carried out to obtain quantitative assessments of the technology's environmental impact. It typically relies on voluntary, albeit partial, environmental statements from generative AI service providers, or on less well-known open models. Average assumptions (e.g. carbon intensity of the average global energy mix, or the average power use of chipsets) are considered, in the absence of industrial data in the papers that were examined.

The elements collected to assess the environmental footprint of generative AI are presented, first, by ICT building block (data centres, devices, networks) (2.1) followed by a life cycle assessment of an AI

service, particularly during the model’s training and inference phases (2.2)³¹. Insofar as possible using available data, the proposed analysis examines a range of environmental criteria (including energy, carbon, abiotic resources and water).

While the following environmental analysis focuses as much as possible on generative AI services, the methodologies employed apply generally to any AI system³² since generative AI is only a sub-category of AI, as illustrated in the following graph, albeit with data and energy consumption levels that tend to be higher than those of other systems, according to the International Telecommunication Union (ITU)³³.

DIFFERENCES BETWEEN GENERATIVE AI AND OTHER AI SYSTEMS

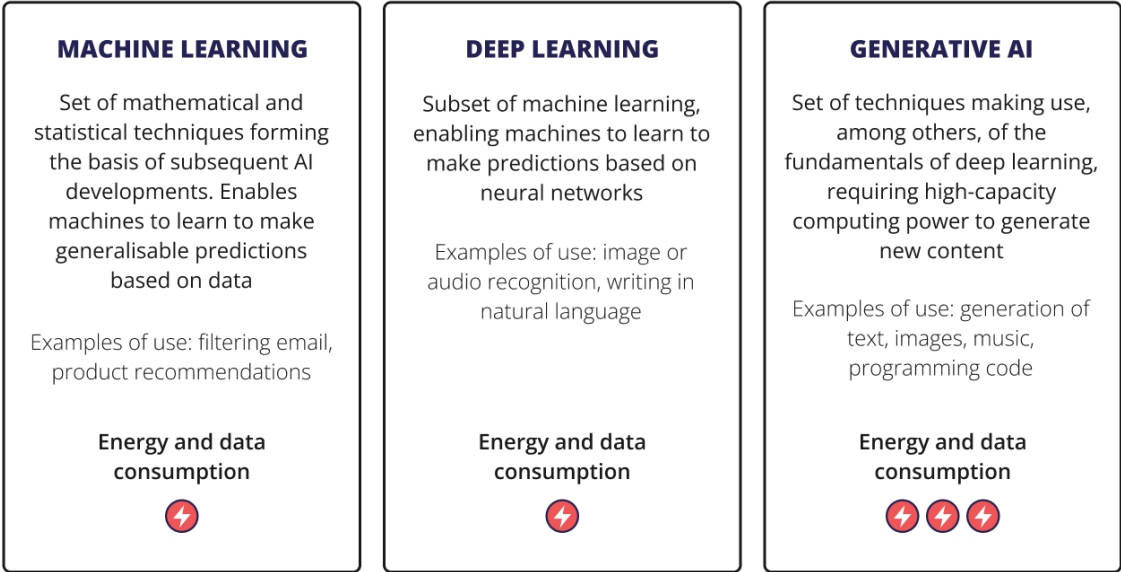


Figure 7 - Types of AI system. Source: Arcep.

³¹ A distinction needs to be made between the model’s training and inference phases, among others (collection, retraining, fine-tuning, etc.) required for the development of generative AI services such as ChatGPT. These two phase are examined in detail because of their sizeable environmental impact compared to the other phases. The stages of an AI system’s life cycle are detailed on page 5 of [ITU-T Recommendation L.1801](#). Specific key parameters must be examined according to these different AI systems (generative AI, deep-learning, machine-learning or expert systems).

³² According to the ISO/IEC 22989 standard, an AI system is an engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives.

³³ See this [pedagogical synopsis of the ITU-T L.1801 Recommendation](#).

1.4 Analysis by ICT building blocks: data centres as the foundational component of the environmental impact of generative AI, and lingering uncertainties over the effects of networks and devices

As stipulated in Part 1.1, assessing the environmental footprint of a digital service, such as generative AI, begins with an analysis of each of the ICT building blocks that underpin these services: data centres, networks and devices. This part takes a look at existing studies that help shed light on the impact of AI, starting with data centres (2.1.1), which currently account for the majority of identified environmental issue, followed by the devices and networks building blocks (2.1.2).

1.4.1 Data centres: the foundational component of the environmental impact of generative AI systems

According to the ADEME-Arcep study, updated in 2025, data centres represented 46% of the digital carbon footprint in France in 2022, or close to 13.5 million tonnes of CO₂e³⁴. This estimate was made before the generative AI boom; it does not make it possible to isolate its share of the digital environmental impact, nor even that of data centres. It is nevertheless probable that the rise of generative AI and its uses (see 1.2.1) will drive up the impact associated with data centres, largely because of generative AI's specific computing capacity requirements (see section 1.1.2).

The impacts associated with generative AI in the data centres building block can be assessed using LCA, through an analysis of the impacts tied to servers and their technical environment (a). If multiple local environmental issues can be linked to the existence of data centres (b), their energy consumption is raising especially weighty concerns at the global level (c).

1.4.1.1 Impacts tied to servers and their technical environment

The data centres building block includes the servers they house. The uses associated with generative AI typically require greater computing capacity as well as dedicated servers with specialised chips, notably graphics cards, with substantially higher computing capacity than traditional servers. These servers, and their components, have multiple effects on the environment, during the different stages of their life cycle.

- Server and chipset production

The production phase of servers and dedicated AI chips³⁵ requires, first, the use of metals which contributes to the depletion of abiotic resources, at a time when demand for metals for digital use could increase by 179% between 2020 and 2050, chiefly due to the tremendous increase in the number of total devices in circulation during that period³⁶ (ADEME, 2024b).

Chip production also uses a great deal of energy and generates substantial GHG emissions (Boavizta, 2021) in addition to requiring large quantities of water, which creates sizeable local pressures, especially during droughts³⁷. Based on the disassembly of an Nvidia A100 SXM 40GBG GPU card, Falk

³⁴ To compare: according to [CITEPA](#), France emitted 369 MtCO₂e in 2024.

³⁵ E.g. [NVIDIA GPU chipsets](#) or [those of Google which assemble their own TPU chips](#)

³⁶ Under the "business as usual" scenario, the number of devices increases by 522% (+53% excluding IoT equipment) between 2020 and 2050. It should be noted that the ADEME study (ADEME, 2024b) was produced based on data that did not take the massive development of generative AI into account, but rather based on the ADEME-Arcep study on the digital environmental footprint published in 2022 and 2023.

³⁷ Gauthier Rousilhe, [Eau et puces électroniques: avenir climatique et industriel de Taïwan](#), 2021.

et al. (2025a) estimate, for instance, the carbon footprint of the production of such a graphics card at 150 kgCO₂e and underscores the need to complete their analysis by factoring in, on top of the already well identified carbon impact of the inference phase, the depletion of abiotic resources, the impact on human health, and the ecotoxicity of the end of life phase.

- **Server use**

During the inference phase, servers need to be powered by electricity. Even if the embodied impacts, in other words those tied to the production of graphics cards, remain significant in terms of environmental criteria linked to the extraction of raw materials, a recent ADEME study, which analyses the composition of certain graphics cards³⁸, reveals that the inference phase of these cards for AI³⁹ has by far the greatest impact for most of the environmental criteria analysed, whether for training or inference (ADEME, 2026). These results corroborate those of Schneider et al. (2025), who carried out a life cycle assessment (LCA) for Google on the hardware used for generative AI, such as servers. The carbon impact of the inference phase for chips thus represented 70% to 90% of the total carbon impact of the life cycle of an AI system's training or inference (ADEME, 2026).

Moreover, the work of Zhang et al. (2025) highlighted that the location of the servers used for training or inference is a decisive factor in the carbon impact, due to the different energy mixes of the countries where they are located. To illustrate: on average in 2024, Poland's energy mix had a carbon content that was 1.65 times higher than Ireland's energy mix and 7.2 times higher than Finland's⁴⁰.

With the exception of the findings of these papers, there is very little recent environmental data on semiconductors (GPU, TPU) available. In particular, regarding their use in servers adapted to cloud computing applications which are also employed for generative AI, the lack of transparency on workloads⁴¹ (i.e. on the quantity of CPU/GPU mobilised to execute a task, or on their utilisation rate) makes it impossible to measure the environmental efficiency of these servers. These indicators are key to understanding whether the servers (and by extension the semiconductors) are underutilised or utilised at 100% of their maximum capacity, to be able to evaluate the energy profile chips under real world conditions (Schneider et al., 2025).

- **The technical environment of servers**

The impacts tied to data centres also stem from the infrastructure required to provide an appropriate operating environment for servers, particularly the cooling systems needed to keep equipment running efficiently.

The choice of cooling system for the servers heavily influences the energy consumed, with gains of up to 50% enabled by advanced systems such as liquid or immersion cooling, compared to traditional air or evaporative cooling systems (Patel et al., 2025), although often at the expense of increased water consumption. The environmental efficiency of cooling techniques is assessed through energy (and water) effectiveness indicators: "Power Usage Effectiveness" or PUE, and "Water Usage Effectiveness"

³⁸ This study compiled data from scientific literature and analysed the composition of some of the graphics cards that were recovered, by disassembling them and weighing each of their components. The goal was to conduct what is commonly referred to as a Life Cycle Inventory as part of an LCA, to model the environmental impact of a graphics card based on its composition.

³⁹ Modelling was carried out on a specialised LCA software programme, using data on the graphics cards' electrical power and composition.

⁴⁰ According to [Electricity Maps](#), in 2024 (site accessed on 28 April 2026), carbon intensity of Poland's energy mix stood at 581 gCO₂e/kWh compared to Ireland at 352 gCO₂e/kWh, or Finland at 80 gCO₂e/kWh). According to the International Energy Agency (IEA), the average carbon intensity of the energy mix around the world in 2024 was 445 gCO₂e/kWh.

⁴¹ A workload is the amount of time and IT resources that a system or network requires to complete a task or generate a given result or output.

or WUE⁴². Improvements in these efficiency indicators can be seen in the decreasing average PUE, both internationally, in industry actors' environmental reports⁴³, and in France where this trend can be tracked through the figures provided in Arcep's annual "Achieving digital sustainability" survey, with an average PUE of 1.42 in 2024 compared to 1.46 in 2023 for the data centres analysed (see inset "Spotlight on the work done on data centres in the annual "Achieving digital sustainability").

In addition to the issues tied to server cooling, there is also the issue of capturing the thermal energy generated by the servers. Servers and other IT equipment produce heat that, if not recovered, escapes into the atmosphere as waste heat. In France, it should be noted that any construction or major renovation project for data centres of over 1 MW, has an obligation to carry out a prior benefit-cost analysis of the "economic feasibility of improving the energy efficiency of heat and cold supply"⁴⁴. ADEME estimates that the potential capture of data centres' waste heat could reach between 4 and 13 TWh in 2035, depending on the data centre development scenarios considered (ADEME, 2026).

1.4.1.2 Local issues tied to the presence of data centres

As mentioned in Chapter 1, the extremely swift expansion of artificial intelligence has spurred the development of data centres, in both size and number, as testified by recent investment announcements in this area. Beyond their overall energy footprint, and the resulting forward-looking questions (see below), data centre development also creates more local issues, particularly with respect to access to energy, water consumption and land use. The feasibility of building a data centre depends on three main restrictions: access to connectivity, access to energy and access to land. This naturally pushes the industry to be geographically concentrated (e.g. in European hubs such as Paris, Dublin, London, Amsterdam, and more recently Marseille which is a landing for a number of submarine communications cables).

- Access to electricity

Because of the geographical concentration of data centres, energy grid managers are having to grapple with local issues created by requests for very large power installations. This led, for instance, to a moratorium (which has since been lifted) on new data centre grid connections from the manager of Ireland's national electrical grid, EirGrid, in 2021 for the greater Dublin area⁴⁵. Demand for data centre connections, which are already by nature very heavy energy consumers, can also compete with demand from other industrial sectors that are on a decarbonisation trajectory, for which access to electricity is vital for their electrification (e.g. the transport sector with private vehicle fleets and bus fleets that are steadily transitioning to electric) and, over the long term, a reduction in their GHG emissions. On this matter, in France, electricity system manager, RTE, published a document for public consultation regarding an overhaul of customer connections⁴⁶, notably to alter the "first come, first served" rule, to inform its thinking about this issue, in concert with stakeholders.

⁴² PUE is an indicator that measures how effectively energy is used, as defined by the ISO 30134-2 standard: the ratio of the total energy consumed by a data centre to the energy consumed by the IT equipment it houses. This means that, the higher the PUE, the lower the data centre's energy-efficiency. By the same token, WUE measures the ratio of the volume of water consumed to the energy consumed by the IT equipment. A data centre that does not use water for cooling therefore has a WUE of zero.

⁴³ [Uptime Institute](#) estimates that the average global PUE was 1.58 in 2023, compared to 2.5 in 2007.

⁴⁴ In accordance with [Law No. 2025-391 of 30 April 2025](#), known as the DDADUE ACT, transposing Directive (EU) 2023/1791 on Energy Efficiency into French Law.

⁴⁵ This moratorium was lifted in late 2025.

⁴⁶ RTE, Customer connection to the electrical grid, [public consultation on overhauling customer connections](#).

- Water consumption

There is currently very little specific data on the water consumption of generative AI in data centres, although some articles have estimated the amount of water consumed by the training of generative AI models, such as GPT-3 in Microsoft data centres (see section 2.2.1) and provide methodological approaches for making these estimates, and therefore to assess cooling systems' water use effectiveness (Li et al., 2023).

Without distinguishing the percentage tied to the use of AI, data collected by Arcep for its annual "Achieving digital sustainability" survey reveal that, in 2024, the volume of water abstracted or consumed by data centres increased by around 8%, due to a significant increase in overall power consumption. This volume is estimated at 6.5 million m³ in 2024, or the average annual water consumption for more than 100,000 people in France (Arcep, 2026d).

The water abstracted by data centres can nevertheless create conflicts over water use in the locations where data centres are installed. Data centres' water consumption is creating controversy and pushback from citizens, as has been the case in the Netherlands⁴⁷.

Recent papers show that the water cycle will also be more and more affected by climate change, making it a crucial issue for the country as a whole (France Stratégie, 2025).

- Land use

Lastly, while land artificialisation and land use can be observed at the local level, with the onset of citizen protests over certain projects⁴⁸, currently available literature remains slim but nevertheless reveals the importance of the land artificialisation issue (Lopez, Diguët, 2023; Diguët, Lopez, ADEME, 2019). Land-related issues can cause local controversies, even when there are existing programmes to bolster a region's attractiveness, supported by economic stakeholders. In some cities, such as Marseille, the development of data centres can clash with other projects in what is already a highly urbanised zone (Arcep, 2025c). It is also one of the reasons for moratoria on the installation of new data centres in cities like Amsterdam and Singapore, where land resources are increasingly scarce.

Spotlight on the work done on data centres in the annual "Achieving digital sustainability" survey

Every year, Arcep delivers an account of its environmental data collection work, carried out in collaboration with industry stakeholders, as part of its [annual "Achieving digital sustainability" survey](#). For the third consecutive year, in 2026 Arcep compiled the findings from the data it collects from data centre operators with an annual revenue equal to or above 10 million euros, excl. VAT, or whose installed IT equipment has a power draw equal to or above 100 kW, i.e. from 23 data centre operators in all. In 2024, these companies were operating close to 160 data centres in France, which represented around half of all collocation data centres in service in 2020, according to the ADEME-Arcep study. Examining the impact of these companies is crucial, as the number of ICT uses they host is surging.

The main findings of this edition of the survey include:

- Data centre consumption has been climbing steadily over the past three years, while the tertiary sector's power consumption decreased significantly in 2023, due to an increase in the price of electricity, then increased slightly in 2024, while still remaining below 2020 levels during Covid. The

⁴⁷ in 2022, according to the newspaper, Noordhollands Dagblad, figures on the water consumed by a Microsoft data centre far exceeded the company's initial estimates (84 million litres compared to a range of 12 to 20 million litres cited by Microsoft). This news was revealed at a time of extreme drought in summer 2022.

⁴⁸ Quadrature du Net, <https://www.laquadrature.net/moratoire-data-centers/>

power consumption of the data centres studied thus continued to increase steadily in 2024, to reach 2.7 TWh. After slowing in 2023 (+ 8% YoY), the increase picked up again in 2024, reaching + 12%, nearing 2022 levels (+ 14%).

- The data centre operators surveyed emitted a total 178,000 tonnes of GHG emissions in 2024⁴⁹. These emissions have been rising at an increasingly faster pace for three years in a row. After having grown by eight points in 2023 to + 13%, the rate of increase rose by 10 points in 2024 to + 23%.
- After two straight years of increase in 2022 (+ 17%) and 2023 (+16%), including an exceptional rise in 2023 due to external factors, including upgrading older installations, the volume of water withdrawn by data centres decreased in 2024 (- 15%) and dropped back to a level similar to that observed in 2022. The volume of water withdrawn or consumed nevertheless rose by around 8%, due to a significant increase in total energy consumption (see previous section 1.4.1.2 “Water consumption,”). This volume is estimated at 6.5 million m³ in 2024, or the average annual water consumption for more than 100,000 people in France.

These increases can be attributed, in part, to the growing number of data centres and their use. The pace of deployment has been accelerating since 2020, concentrated in the Île-de-France region, and involving centres with increasingly high IT capacities. As reported by RTE⁵⁰, there has been a surge in data centre construction in France in recent years. There are more and more projects for increasingly large data centres with, for instance, requests for connections of 100 to 200 MW, compared to previous draws of several megawatts⁵¹.

Drawing on the Decision on environmental data collection approved on 21 January 2026, Arcep expanded its data collection to include cloud computing service providers for the 2027 edition. It plans on collecting indicators from these companies to monitor the evolution of the number of servers and processors in data centres, and to evaluate their capacity to execute AI tasks. In particular, this information will make it possible to assess the impact of generative AI on data centre power consumption.

This work will continue, to enhance the data collected in concert with stakeholders.

*
**

Arcep began working on the issues surrounding the digital environmental footprint in 2019, and has been entrusted with a steadily growing number of powers in this area by the legislature, including the ability to collect data from market stakeholders on their environmental footprint. This data collection, which began in 2020, with the country’s four main electronic communications operators, was expanded by the “REEN 2 Act” of 23 December 2021 to include data centre operators in particular. In 2024, the Law on Safeguarding and Regulating Cyberspace (“SREN” Act) further expanded those data collection powers to include cloud computing service providers.

Arcep’s work has been commended by the World Bank and the International Telecommunication Union (ITU). In 2025, these two institutions published a joint case study: “Measuring national ICT Sector Environmental Impact – Arcep case study” (World Bank and ITU, 2025), devoted to Arcep’s annual survey, and stating that Arcep is notable for being “*the first and only sector regulator to regularly publish, within its annual survey, indicators from data it collects directly from digital players to assess*”

⁴⁹ Scopes 1 and 2 of greenhouse gas emissions are measured. Regarding Scope 2, the methodology used to calculate these emissions the location-based methodology.

⁵⁰ RTE (*Réseau de transport électrique*) is responsible for operating national high and ultra-high voltage transmission system operator for Metropolitan France

⁵¹ Source: RTE, [Centres de données: 11 chiffres sur leur essor en France et leurs besoins en électricité](#)

and monitor their environmental impact". These institutions are calling for the conditions enabling this collection of environmental data to be implemented internationally.

1.4.1.3 Issues tied to energy consumption and greenhouse gas emissions

A rapidly accelerating energy footprint

There is a consensus over the expected exponential increase in data centre energy consumption due to the generative AI boom. In its latest report, the International Energy Agency (IEA), for instance, estimates that data centre global energy consumption reached 415 TWh in 2024, or 1.5% of total global electrical consumption, and that it could double by 2030, in large part due to the AI boom, and potentially reach close to 1,000 TWh, or equal to Japan's total power consumption in 2024⁵² (IEA, 2025). According to the IEA Energy and AI Observatory, in 2024, data centre power consumption represented 4.4% of total power consumption in the US and 2.3% in the European Union⁵³.

In France, the ADEME-Arcep study (2023) served to estimate that the overall power consumption of data centres in France totalled around 12 TWh in 2020 and could climb to close to 16 TWh in 2030, without factoring in the data centres abroad used by French services, or the specific recent momentum created by the generative AI surge.

Data on data centre power consumption collected by Arcep on a limited scope (see inset below: "Spotlight on the work done on data centres in the annual "Achieving digital sustainability" survey) also notes the increased power consumption over the past several years (+14% between 2021 and 2022 and +8% between 2022 and 2023 and +12% between 2023 and 2024). Although this upwards trend in power consumption has been observed on different scales, it is difficult to estimate the percentage that can be attributed to AI. Estimates do exist, but are currently difficult to countercheck due to a lack of data. ADEME, for instance, estimates that AI's share of data centre power consumption in France stands at 13% for 2025, taking into account data centres abroad used to provide French services (ADEME, 2026) and the Shift Project estimates this share at approximately 15% worldwide (The Shift Project, 2025).

As a result, the development of AI is viewed as one of the main drivers of the growing power consumption of data centres over the coming years. Forward-looking studies on their power consumption, notably in relation to AI, have been carried out in recent years by a range of stakeholders (institutions, industry actors, financial analyst) and supporting this hypothesis. In addition to the necessary caution regarding the findings⁵⁴, it is interesting to note that these different studies reveal a consensus on the significant increase to come, in the energy consumed by data centres to a surge in usage. As the following graph shows, the studies converge to highlight the fact that data centre power consumption is expected to increase exponentially between now and 2030, notably because of the development of generative AI.

⁵² Japan's Ministry of Economy, Trade and Industry, 2025 ([link](#)).

⁵³ IEA, 2025 [Energy and AI Observatory](#).

⁵⁴ These prospective analyses need to be taken with a degree of caution as they are based on partial data and assumptions that heavily skew the results. They are thus possible trajectories rather than forecasts.

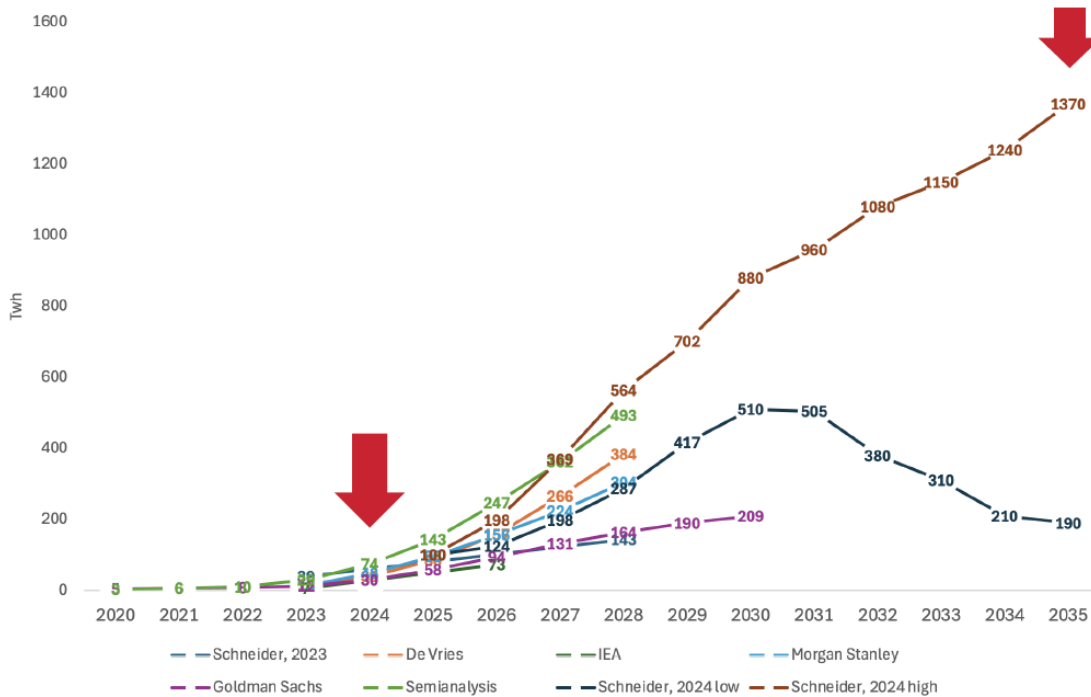


Figure 8: AI-related energy consumption, according to different sources⁵⁵

Looking beyond 2030, the evolution of energy consumption will depend in large part on data centres' rate of deployment and the efficiency gains. Here, in all of the scenarios considered by the IEA, the Agency estimated (see graph below) that data centres' global energy consumption would increase sharply up to 2030 (by between 700 and 1250 TWh, compared to 415 TWh in 2024), and to a less certain degree between 2030 and 2035 (between around 750 TWh and 1750 TWh). This evolution depends to an extent on the assumptions in energy efficiency gains, the evolution of demand for generative AI services and the previously cited restrictions, such as access to electricity or land, and on social acceptability. These assumptions are translated into the following scenarios: a *Lift-Off* scenario that considers a high take-up rate, a *Base*, business-as-usual scenario, a *High Efficiency* scenario that posits major technological progress in energy efficiency, and a *Headwinds* scenario that sketches out a low take-up rate.

⁵⁵ Source: Lynn Kaack, graphic published in 2025 for the AI Action Summit in Paris, based on data from the Roegen Centre for Sustainability and INFRAS based on data from Schneider Electric, 2023; Schneider Electric, 2024; De Vries, 2023; IEA, 2024; Morgan Stanley, 2024; Goldman Sachs, 2024; Semianalysis, 2024.

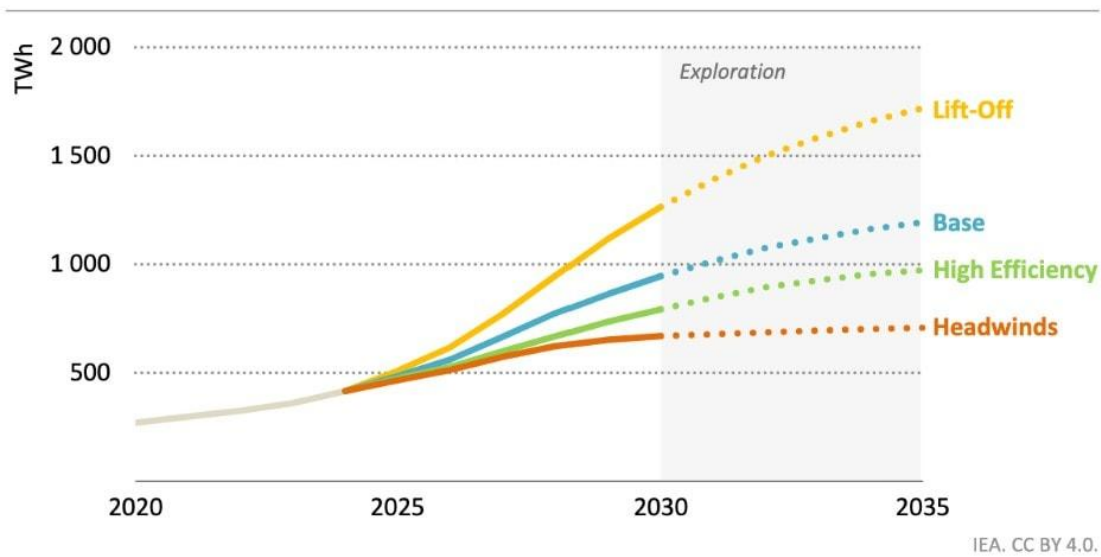


Figure 9: Projections of data centres' global energy consumption according to different scenarios
Source: IEA, 2025

Growing pressure of decarbonisation strategies and trajectories

This rapid increase in data centres' power consumption due to the growing use of generative AI is pushing some of the leading generative AI market actors, especially in the United States, to adjust their strategies. These Big Tech companies are tending to expand their positioning to become energy producers as well, to guarantee that their needs are covered – one prime example being the interest being expressed in reviving certain existing capacities, such as the nuclear reactor at Three Mile Island⁵⁶ and retired gas plants. The development of AI is also raising broader questions about the top generative AI and cloud computing companies' ability to reconcile the growth of their businesses and compliance with their climate commitments. Despite the dearth of environmental information supplied by generative AI companies, they have acknowledged that the environmental impact of AI models, and generative AI models in particular, is substantial and challenging their low-carbon strategy due to the development of data centres. In 2024, Google and Microsoft, for instance, recognised that their ability to reach their initial decarbonisation targets for 2030 had been compromised by the development of generative AI systems, and both scaled back those targets⁵⁷. In 2025, Melanie Nakagawa, Microsoft's Chief Sustainability Officer, said: *"in 2020, Microsoft leaders referred to our sustainability goals as a "moonshot," and nearly five years later, we have had to acknowledge that the moon has gotten further away"*⁵⁸.

Beyond industry actors' strategies, the rise in data centre power consumption creates issues for individual countries and their ability to keep to their decarbonisation trajectories. This has led, for example, to a massive revival of fossil-fuel powered infrastructures (notably natural gas in the US) due to limited low-carbon production capacities (The Shift Project, 2025). In light of this development and the lack of data for assessing any intensity, the US Federal Energy Information Agency (EIA), at the

⁵⁶ AP News, [Energy Department loans \\$1B to help finance the restart of the nuclear reactor on Three Mile Island](#), 19 November 2025.

⁵⁷ [Les Echos, 2024](#); [Novethic, 2024](#). See also the 2024 environmental reports from [Google](#) and [Microsoft](#).

⁵⁸ Microsoft, 2025, [Progress on the road to 2030](#).

behest of Senators Warren and Hawley, is planning to launch a mandatory nationwide survey of data centres focused on their energy use in the United States⁵⁹, after the launch of pilot projects in Texas and Washington state. The assessment will cover not only annual energy consumption, but also local energy production, often natural gas-based, and the efficiency of these infrastructures' cooling systems⁶⁰.

For some countries, such as Ireland, being home to a massive number of data centres is already undermining its decarbonisation targets. This growth momentum is creating increasing pressure on the national grid (of which a large percentage is currently based on natural gas): data centre consumption in Ireland rose by 24.7% between 2012 and 2022, spurred in large part by the development of cloud computing⁶¹. Today, these infrastructures already account for close to 20% of national power consumption, or around 18 points more than the global average, a percentage that could climb to 30% by 2028 according to certain projections (AIE, 2024). In this context, questions about hosting and setting rules for these installations has become a real issue for Ireland, as much from an energy as an environmental perspective.

1.4.2 Questions about the medium-term consequences for devices and networks

1.4.2.1 Devices: potentially greater intensity of energy use by generative AI services and the risk of forcing users to replace their devices

As detailed in section 1.1.1, the devices building block includes the hardware (e.g. televisions, smartphones, computers) used to consume digital services, including generative AI solutions. As a reminder, in 2024, households in France owned an average 9.6 digital devices with a screen, of which 1.8 were not being used, and could be recycled or sold⁶².

According to the ADEME-Arcep study, devices represented 50% of the digital carbon footprint in France in 2022, or close to 14.7 million tonnes of CO₂e⁶³. With respect to digital services, it is the building block with the greatest impact for most criteria.

The development of generative AI services is likely to affect this building block in various ways. First, this development will lead to an increase in the energy consumed by the use of devices, since these services consume a great deal of power. The use of generative AI services may indeed lead to prompts being processed in the cloud (increasing traffic and data processing with the device) or directly on the device, thereby draining devices' (especially smartphones) batteries more quickly⁶⁴ (The Shift Project, 2025). In addition, service providers could impose these energy-hungry services on end users (Arcep, 2026a; Beignon et al., 2025; Limites Numériques, 2025), with consumers having a limited capacity to choose whether they want these features added to the services they use (e.g. incorporation of generative AI features in instant messaging apps). This could thrust some devices into early

⁵⁹ Wired, 2026, [The US Government Will Ask Data Centres How Much Power They Use](#)

⁶⁰ EAI, 2026, [EIA launches pilot survey on energy use at data centres](#)

⁶¹ As mentioned by the *Autorité de la Concurrence* in its [Opinion on artificial intelligence](#) published in 2024, "use of the cloud appears to be a necessary route for accessing the computing power needed to train AI models. It is also a vector for disseminating the models upstream, on marketplaces. These marketplaces give developers the ability to make their models more easily accessible to enterprises using cloud services, which encourages developers to make their models available on every cloud service provider's platform".

⁶² Arcep, Arcom, CGE, ANCT, [Digital Barometer, 2025 edition](#).

⁶³ According to the update to the ADEME-Arcep study produced by ADEME in November 2024

⁶⁴ Greenspector, [Quels impacts environnementaux pour les AI locales sur nos smartphones ?](#), 2025

obsolescence, if they are not powerful enough to use these services comfortably, despite the energy efficiency gains sought by industry actors⁶⁵ (The Shift Project, 2025).

The development of generative AI services can also lead to the production of new, specifically adapted devices, installed with graphics processing units (GPUs) or other types of chips (NPUs). Devices with an embedded generative AI system⁶⁶ are already on the market. These new devices could have a greater environmental footprint, during the production and use phase, and above all drive a tide of early device upgrades amongst users whose devices do not have the computing capacity compatible with generative AI services, thereby speeding up the obsolescence phenomenon.

As it stands today, massive effects are not being seen on the market, and current services are by and large able to run on the same hardware, as prompts are primarily processed in the cloud, and relayed by the network. These trends could nevertheless evolve rapidly, especially if companies' strategies provide them with an incentive to start embedding processing power in the device.

⁶⁵ GSMA lists [actions](#) that operators are taking to make devices more energy efficient, with a dedicated initiative called [EcoRating](#).

⁶⁶ On-board generative AI services include Apple Intelligence (Apple), Galaxy AI (Samsung), Gemini Nano (Google), AI Engine (Qualcomm), HyperOS (Xiaomi) ou Harmony AI (Huawei). iPhone 16 Pro Max, Pixel 9 Pro XL and Galaxy S25 p smartphones make it possible to embed these generative AI services.

POTENTIALLY CASCADING EFFECTS OF GENERATIVE AI ON DEVICES

Depending on the strategies adopted by suppliers, the development of generative AI services can have multiple effects on user devices:

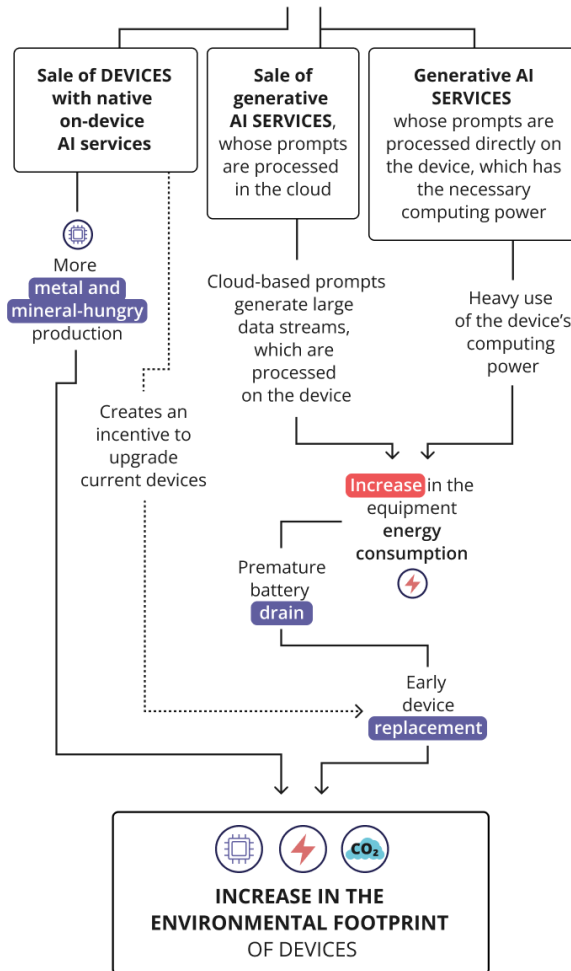


Figure 10: Simplified representation of AI's different effects on devices. Source: Arcep.

1.4.2.2 Networks: theoretically sufficient capacity in the short term but uncertainties about the medium-term

According to the ADEME-Arcep study, the networks building block represented 4% of the digital carbon footprint in France in 2022, or close to 1.2 million tonnes of CO₂e⁶⁷. Of the three building blocks, this is the one with the least influence on most of the environmental impact criteria being examined.

⁶⁷ According to the update to the ADEME-Arcep report produced by ADEME in November 2024.

The impact of generative AI services on networks is likely to materialise in several ways⁶⁸. First, a number of papers reveal that AI models' training phase relies on high-capacity computing power and communication between learning nodes that draw more heavily on network resources than average digital practices⁶⁹ (Arcep, 2025c).

Next, during the inference phase, AI applications, and particularly the generation of data-intensive content (especially videos or interactive content that requires low latency and increased synchronisation with the AI agent on the server) could account for a growing share of internet traffic, apace with the increased number of prompts. In particular, according to telecom equipment supplier estimates⁷⁰, the development of "indirect AI traffic", in other words traffic generated by the creation of hyper-personalised content tailored to users (based on their browsing history, location, time of day and weather) by recommendation algorithms that can use generative AI, could create a snowball effect whereby every prompt generates a host of requests or calls to other AI systems, thereby exponentially increasing the quantity of data being circulated. Indirect AI traffic could also overtake direct AI traffic (The Shift Project, 2025).

Lastly, looking ahead, a number of questions remain about the distribution⁷¹ of computing power: it can be deployed in either a centralised or distributed fashion, which would entail different levels of network scaling, depending on the situation (Arcep, 2025c). There are also lingering uncertainties about devices' capacity to contain local AI capabilities (see above), which would help offload the networks but would still need to be synchronised with the cloud. It is also worth noting that uncertainties over development, network scaling and the associated effects will also vary depending on each company's business model and strategies (see 1.2.2).

Generally speaking, even with the massive deployment of generative AI, network capacity could suffice in the short term to meet this demand: generative AI does not appear to currently be a factor in network saturation. Based on the literature studied and the interviews conducted as part of the Future Networks cycle of inquiry, generative AI does not currently appear to be affecting fixed network scaling (Arcep, 2025c). There are nevertheless questions regarding mobile networks in the medium term. The growth of AI-powered user services could drive up edge computing capacities that would affect certain specific sections of the network. It could also amplify the volume of uplink traffic (from users to the network), thereby creating a need to adjust network scaling which is typically designed with downlink traffic (network to users) as the dominant direction of traffic. Standardisation work is also underway on 6G, which is due to establish AI-native networks. Lastly, AI integration into the networks could help improve network performance and create a lever for optimising capacity.

*
**

This analysis by building block reveals that the main environmental issues are tied to the rapid development of AI and the prospect of massive data centre deployment. The AI boom is indeed likely to go hand in hand with an exponential increase in energy demand from data centres and their high-capacity computing infrastructures. It is also expected to translate into increased consumption of metal and mineral resources, which are required for the production of the chips used in servers. On a local level, access to electricity and the impact on water use could lead to conflicts over usage, while access to land could trigger citizen protests. If the issues surrounding data centres are, in theory, the

⁶⁸ This was a targeted analysis carried out on fixed and mobile networks as part of Arcep's forward-looking "Future Networks" cycle of inquiry (2025), for the "Artificial intelligence and electronic communications networks" briefing note.

⁶⁹ The Statista website estimates that, in 2023, Azure was providing bandwidth of 1600 Gbit/s for AI computing power needs.

⁷⁰ Nokia. (2024). Global network traffic report. <https://onestore.nokia.com/asset/213660>

⁷¹ In other words, the location-based distribution of computing power

largest and the least certain, the medium-term environmental impact of AI created by users replacing their devices and the re-scaling of networks, and mobile networks in particular, must be taken into account.

1.5 Analysis by phase: growing usage that is increasing the impact of the inference phase relative to model training

Recent studies on generative AI have helped establish an initial knowledge base on the environmental impact of AI services' training and inference phases, despite the lack of transparency from service providers. This section focuses on the studies that shed light on these two phases. It should nevertheless be noted that the stages in an AI model's life cycle are not limited to training and inference. Model training that does not lead to the marketing of a model – along with the retraining, continuous training or fine-tuning of models – consumes energy and generates its own environmental footprint.

1.5.1 Training phase: impact that depends on the number of model parameters, the energy mix and the equipment used

Multiple studies have been done on the environmental impact of the training phase, both academic and by the sector itself. While academic papers rely on transparent methodologies and assumptions, the vast majority of studies by the private sector rely on opaque methodologies and therefore deliver results that cannot be compared or challenged.

1.5.1.1 Number of parameters and length of training: two key factors in the power consumed to train a model

As detailed in Chapter 1, the number of model parameters has increased exponentially (see Figure 4) over the past several years (Varoquaux et al., 2025)⁷², which raises questions about the increase in computing power, hence the power consumption needed to train them.

Academic work shows that the duration of the training and the computing power employed are decisive in determining the amount of energy consumed during the models' training phase (Strubell et al., 2019; Patterson, 2021). The number of parameters also has a direct effect, however: the larger the number of parameters, the longer the training phase and the greater the quantity of resources required.

How to measure training phase power consumption?

A number of academic papers provide figures on the environmental impact of the training phase using transparent and freely accessible tools, or based on estimates drawing on a small amount of publicly available industry data. In their synopsis of the academic studies on the environmental impact of the training phase, in terms of energy consumption and greenhouse gas emissions, Kim et al. (2025) observe that a number of assessment tools already exist for quantifying and monitoring greenhouse gas emissions generated by models' training phase. These can include estimates (empirical or statistical models) such as the ML Emissions Calculator (Lacoste et al., 2019) and Green Algorithms (Lanelongue et al., 2021), sensor-based tools such as Carbontracker (Anthony et al., 2020), and hybrid tools such as Code Carbon (Schmidt et al., 2021) and Eco2AI (Budenny et al., 2020). Assessments of

⁷² 120 million parameters for the first ChatGPT model released in 2018 compared to 1,000 billion for the GPT-4 model released in 2024.

the training enabled by these tools provide a good starting point for discussing the sustainability of generative AI services.

Strubell et al. (2019) carried out the first quantitative study assessing the carbon impact of the training phase of natural language models (Transformer, ELMo, BERT and GPT-2). This study, which includes assessments of the energy consumption of GPUs and CPUs, revealed that the size of the model is key to its energy consumption: lightweight models emit far less carbon than the largest models. By way of example, the greenhouse gas emissions generated by the training of a big Transformer model with a neural architectural search were estimated at 284 tCO₂e compared to 0.087 tCO₂e for a more lightweight, Transformer big model with no neural nets.

Budenny et al. (2022) estimated the greenhouse gas emissions generated during the training of the Malevich (1.3 billion parameters) and Kandinsky (12 billion parameters) ML image generators, and quantified the ratio of the greenhouse gas emissions of the two models at 17:1. According to their estimates, factoring in Russia's energy mix, the first model emits 330 gCO₂e⁷³ (corresponding to energy consumption of 1.37 kWh) for a training phase lasting 4 hours 19 minutes, and the second 5.89 kgCO₂e⁷⁴ (or energy consumption of 24.5 kWh) for a training phase lasting 9 hours 45 minutes. Although these models are not comparable to the large generative AI models, these estimates of the carbon footprint of their training do shed light on the impact of parameter numbers.

In the assessment of an open-source model, the BLOOM general-purpose LLM⁷⁵, trained on 176 billion parameters, Luccioni et al. (2023) also demonstrated that the training phase is a particularly heavy consumer of resources (servers, GPUs) and electricity: 348 GPUs (NVIDIA A100) were mobilised for 118 days, consuming a total of 433 MWh, and generating a carbon footprint resulting from the servers' power consumption of 30 tCO₂e. By adding the carbon footprint of the hardware production, Luccioni et al. (2023) estimate that the total carbon footprint of the training of this model was around 124 tCO₂e, a relatively small order of magnitude compared to the training of GPT-3, for instance, due to the use of hardware considered efficient and France's low-carbon energy mix (see Annex 5.5 that compares orders of magnitude for training GPT-3 and BLOOM). It should nevertheless be noted that this is a relatively small model as it was only trained on a 1.6 Tb dataset (compared to an estimated 45 Tb to train the GPT-3 model⁷⁶).

Some studies also sought to identify avenues for reducing the power consumption of the training phase. Although the size of the model is a key factor, Patterson et al. (2021) revealed, in work carried out for Google, that progress in the energy efficiency of Google data centres and in the design of models (notably architecture) helped reduce the greenhouse gas emissions caused by the training of the models that were examined by Strubell et al. (2019) by a factor of 80. Although this study may be questionable from a methodological standpoint⁷⁷, it also helps to underscore the importance of factoring in efficiency gains made possible by technological developments in recent years. The following table recaps the orders of magnitude of the different academic papers cited in this section: these figures must be taken with some caution as not all of this information is systematically communicated, and there may be concerns about the reliability of certain input data.

⁷³ i.e. 1.27 gCO₂e/min training (or 0.005 kWh/min training).

⁷⁴ i.e. 10 gCO₂e/min training (or 0.04 kWh/min training).

⁷⁵ This is a model initiated as part of the *BigScience* project, chiefly by Franco-American company HuggingFace, CNRS, GENCI and France's Ministry of Higher Education and Research. The was trained in France on the Jean Zay supercomputer.

⁷⁶ French Senate, 2024, [ChatGPT, et après ? Bilan et perspectives de l'intelligence artificielle](#)

⁷⁷ In particular, the choice to only communicate market-based results is questionable, as the location-based results of life cycle assessments must also be shared (See section 1.4.1.2).

Model trained	Number of parameters	Training length	Resources mobilised	Electricity consumed by servers for training	Greenhouse gas emissions associated with server use
GPT-3	175 billion	15 days	10,000 GPU	1 287 MWh	552 tCO ₂ e
Transformer big with neural architecture	213 million (with 979 million training steps)	10 hours (equal to 274,120 hours on 8 GPU P100)	1 TPUv2 core	656 MWh	284 tCO ₂ e
BLOOM	176 billion	118 days	384 GPU	433 MWh	30 tCO ₂ e
Transformer big with no neural net	213 million (with 300,000 training steps)	3.5 days	8 GPU	0.201 MWh	0,087 tCO ₂ e
Kandinsky	12 billion	9 hours 45 minutes	N/A	0.0245 MWh	0,00589 tCO ₂ e
Malevich	1.3 billion	4 hours 19 minutes	N/A	0.00137 MWh	0,00033 tCO ₂ e

Table 1 – Recap of the orders of magnitude of energy consumption and greenhouse gas emissions generated during the training phase of the models analysed, based on a review of the scientific literature

Tech companies such as Meta and Mistral AI have also conducted their own impact studies, but they are either incomplete, as they do not meet the requirements of an LCA, or they lack transparency on the assumptions considered, and are therefore difficult to countercheck. Meta, for instance, released information in July 2024 on the carbon footprint of the training phase of LLaMa3-405B⁷⁸: it is reported to have employed 16,000 GPU H100 and emitted around 9,000 tonnes of CO₂e over a limited scope of operations, as only the energy consumed by the GPU chips was calculated, i.e. failing to factor in the power consumption of the CPUs and network equipment, the non-IT equipment used to cool the servers, and during the servers' production phase⁷⁹. In July 2025, Mistral shared some of the findings of an LCA of its Large 2⁸⁰ model: it is reported to have generated 20,400 tonnes of CO₂e during its 18-month training phase (Mistral, 2025)⁸¹.

Model trained	Number of parameters	Training length	Resources mobilised	Electricity consumed by servers for training	Greenhouse gas emissions associated with server use
Mistral Large 2	123 billion	18 months	N/A	N/A	20 400 tCO ₂ e

⁷⁸ See the [analysis](#) on the Hugging Face website, based on available data, [using this methodology](#).

⁷⁹ [Carbone 4 estimates that if the complete life cycle is taken into account, the carbon impact would be 19,600 tonnes of CO₂e.](#)

⁸⁰ This LCA, performed by Carbone 4 in partnership with ADEME, was based on the General framework for frugal AI (AFNOR, 2024), a benchmark initiative developed in partnership with France's Ministry of Ecological Transition. Only a single news release was made available on the Mistral website, which makes it impossible to verify the information provided, and particularly the key parameters of the modelling.

⁸¹ For the sake of comparison, the average annual carbon footprint of a person in France was 8.5 tonnes of CO₂ equivalent, per adult, in 2023, [according to Citepa and ABC](#). Reminder: ICT accounted for 4.4% of the country's carbon footprint in 2022, or 29.5 MtCO₂e of GHG emitted in 2022 (ADEME, 2025c).

Meta LLaMa3-405 B	405 billion	N/A	16,000 GPU H100	N/A	9,000 tCO ₂ e
-------------------	-------------	-----	-----------------	-----	--------------------------

Table 2 – Recap of the orders of magnitude of energy consumption and greenhouse gas emissions generated during the training phase provided by AI model suppliers

In addition to the energy consumption and associated greenhouse gas emissions, other papers have analysed the impact of training a model in terms of water consumption. Li et al. (2023) thus estimated that training GPT-3 required 700,000 litres of water, over 15 days, to cool the data centres, for a total water footprint (including indirect abstractions tied to energy production⁸²) of 3.5 million litres⁸³.

1.5.1.2 Carbon intensity of the training country’s energy mix: a key element to be factored in at the design stage

As mentioned in the previous section, the training phase is especially energy intensive as it requires massive computing power. The greenhouse gas emissions generated by the use of servers depends to a great extent on the carbon intensity of the energy mix used (see section 1.4.1). To the extent that a model is typically trained a single time, without the major latency issue, the training location is an important factor for model providers to consider from the design stage.

There is a lack of transparency on training locations: it is rarely included in the information released, and the details that generative AI services providers do share on greenhouse gas emissions do not make it possible to deduce the location where the model was trained, and can be deceiving. The greenhouse gas emissions tied to model training: indirect emissions tied to energy consumption, representing Scope 2 emissions, can be communicated in different ways, and stakeholders tend to opt for a market-based (economic) approach, which minimises their impact, over a location-based (physical) one, which reflects actual physical outputs⁸⁴. A market-based approach serves to report on the progress that enterprises achieved when signing renewable energy contracts⁸⁵. This methodology allows them to declare zero greenhouse gas emissions for a portion of their energy consumption equal to the quantity of renewal energy covered by these contracts. The GHG Protocol recommends using both of these methodologies at once, to establish the most complete and accurate documentation and evaluation possible.

The ability to evaluate the greenhouse gas emissions generated by training continues to be shaped by Tech companies’ transparency and reporting rules, and the use of market-based approaches that might not reflect that actual level of emissions generated by training.

1.5.1.3 Hardware used: training power consumption depends heavily on the features of the chips

Generative AI models can be trained using different hardware, and particularly different chip models, which will affect the energy needed to power the training process. Technological progress has spurred

⁸² The water footprint is based notably on EWIF (Energy Water Intensify Factor) which quantifies the amount of water needed to produce electricity. According to Microsoft’s annual report for 2022, the water footprint was 6.4 billion litres.

⁸³ For the sake of comparison 2021, according to [INSEE](#), the volume of water abstracted stood at 29,543 million m³ in France. According to [Eurostat](#), this figure stands at around 218,000 million m³ in the European Union.

⁸⁴ The Guardian, [Data centre emissions probably 662% higher than big tech claims. Can it keep up the ruse?](#), 2024

⁸⁵ These disparities can be due to: 1/ renewable energy certificates (REC) which are low-carbon electricity credits that can be used up to one year after purchase, anywhere in the world; 2/ direct renewable energy Power Purchase Agreements (PPA) which can be physical, involving the physical delivery of a quantity of electricity (via the public grid or not) or virtual PPAs which are financial agreements, with no physical energy supply.

a tremendous evolution in the chips' features over the past several years. Their power draw has increased⁸⁶ but they have also made sizeable gains in computing power efficiency. According to an analysis from Epoch, total installed NVIDIA computing power doubles each year, as illustrated in the following graph. In light of these gains, the chips being used will heavily influence the energy required to train a model, all else being equal.

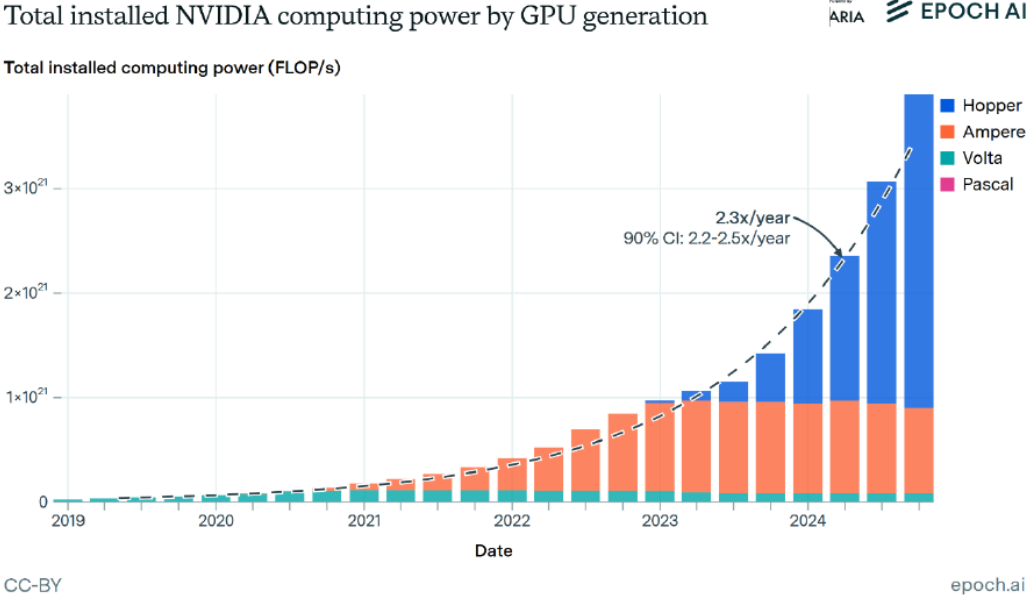


Figure 11 - Installed NVIDIA computing power by GPU. Source: Luke Emberson and David Owen (2025). Published on epoch.ai

It should nonetheless be noted that the technological progress that has been made, while considerably increasing computing power, makes it possible to train every larger models with an ever larger number of parameters. As a result, the power needed to train AI models continues to double each year, according to another Epoch AI study, as shown in the graph below. This dynamic illustrates the rebound effect: efficiency gains made by chips are today translating into increased overall energy consumption.

⁸⁶ The energy draw of an NVIDIA H100 chip is 700 W (or the power of a microwave oven) compared to 400 W for an NVIDIA A100 chip. The new chips can have a draw of up to 1 400 W in 2026, as mentioned in an [article](#) by Gauthier Roussilhe. By comparison, the average energy draw of a consumer CPU varies between 75 and 150 W but remains relatively stable according to the researcher. A [graph](#) traces the history and potential evolution of this power for chips.

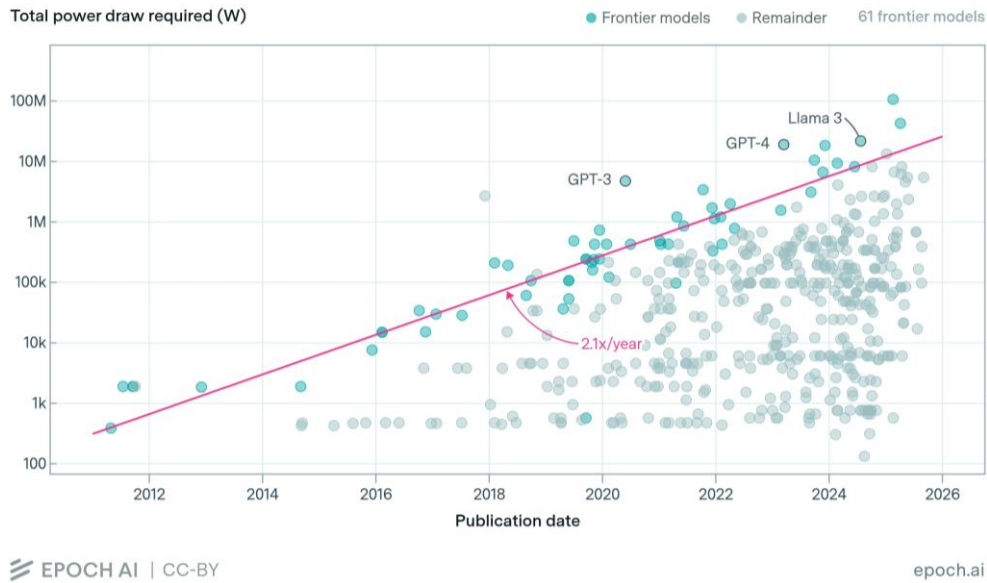


Figure 12 – Estimated computing power needed to train AI models. Source: Epoch AI

Lastly, the number of chips used to train AI models is also on the rise. In addition to their use, chip production too has an embodied carbon impact, and requires a substantial number of metals. Falk et al. (2025b), for instance, estimated that training GPT-4 required between 1,174 and 8,800 GPU A100 graphics cards, or between 176 tCO₂e and 1 320 tCO₂e. The study estimates that the embodied carbon impact represents close to 13% of the total carbon impact associated with the model’s training.

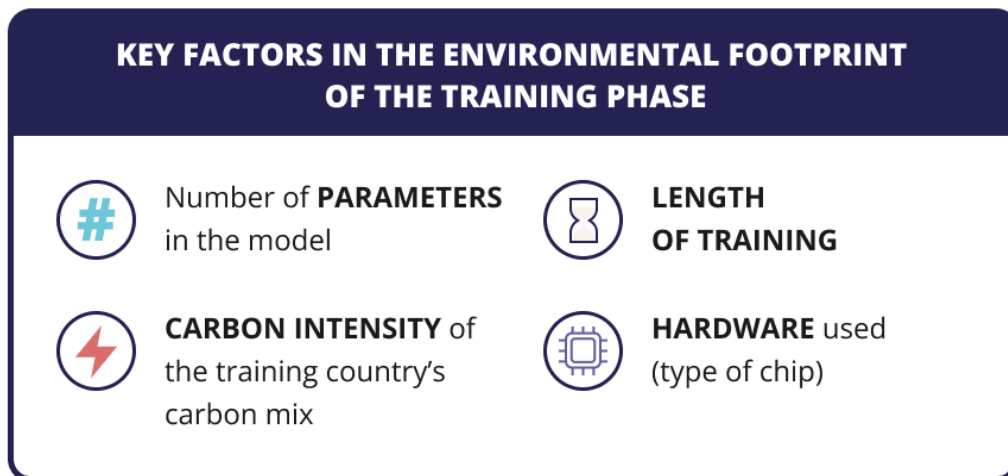


Figure 13: Key factors in the environmental footprint of the training phase for a generative AI model. Source: Arcep.

1.5.2 Inference phase: in addition to the impact per prompt depending on the number of parameters, overall impact rising sharply due to the rapid spread of generative AI

There are fewer estimates of the environmental impact of inference phase than the training phase. There are nevertheless studies that make it possible to pinpoint the key factors of the environmental impact of the inference phase. Especially noteworthy is that these studies show an inference impact that remains limited when assessed per individual prompt but, when taken together, could become significant with the exponential growth of uses (a), in addition to a widely varying impact depending on the type of content generated (b). To assess the impact of other factors, which have been largely overlooked in the literature up to now, supplementary studies were carried out by PEReN on behalf of Arcep and help to flesh out the analysis (c).

1.5.2.1 A small impact per prompt, but an issue that is volume-dependent

Several academic papers provide estimates on the energy consumption of a prompt (Elsworth et al., 2025). De Vries (2023) estimates that a prompt on ChatGPT-3.5 consumes around 3 Wh and You (2025) that a prompt on ChatGPT-4o consumes around 0.3 Wh. Meanwhile, Rincé and Banse (2025) estimate that a 50-tokens long⁸⁷ response would consume between 1.83 Wh and 6.95 Wh according to the tested model. These figures come from estimates, and not assessments performed *in situ*.

Other scientific papers made it possible to measure the energy consumption of a prompt based on tests carried out *in situ*. Luccioni et al. (2022), using the CodeCarbon tool to assess the BLOOM model, evaluated the energy consumption at 4 Wh/prompt (for close to 230,000 prompts evaluated) and greenhouse gas emissions at 1.5 gCO₂e/prompt. Samsi et al. (2023), meanwhile, performed tests on LLaMA (7B, 13B and 65B) models on an MIT supercomputer using NVIDIA V100 and A100 GPUs and estimated that the LLaMA-65B model consumed 0.3 Wh/prompt.

If estimated impact per-prompt seems relatively low, it also needs to be set in the context of absolute environmental impact (e.g. with direct and indirect greenhouse gas emissions over the course of year, for instance) and the total number of prompts. In July 2025, OpenAI reported 912.5 billion prompts per year, every year, or an estimated average 2.5 billion prompts per day⁸⁸.

Other estimates and orders of magnitude on the impact of prompts are often relayed in the press, but the validity of the data used to obtain these estimates has been disputed (Vanderbauwhede et al., 2024) and the findings are difficult to countercheck, because of methodologies that may lack transparency and data that cannot be compared⁸⁹. For instance, Open AI had stated in 2020 that a GPT-3 prompt had a draw equal to 4 Wh (OpenAI, 2020), and in 2025 that an average prompt on ChatGPT which was equivalent to a GPT-4 prompt consumed 0.34 Wh⁹⁰. Meanwhile, Mistral and

⁸⁷ Tokens are units of data processed by generative AI models to enable prediction, generation and reasoning when required. For more details, see the definition in the glossary.

⁸⁸ Les Numériques, July 2025, "[ChatGPT expose absolument tous les compteurs: voici combien de requêtes sont traitées chaque jour par l'IA](#)". By considering that a prompt consumes between 0.34 Wh (according to OpenAI's assumption for GPT-4) and 4 Wh (according to OpenAI's assumption for GPT-3 and the orders of magnitude provided by scientific articles), total energy consumed by prompts in 2025 would be between 310 GWh and 3650 GWh, or GHG emissions of between 138,000 tCO₂e and 1,600,000 tCO₂e (factoring in the average carbon intensity of the global energy mix of 445 gCO₂e/kWh which is the 2024 figure provided by the IEA).

⁸⁹ For more transparency and hindsight on the methodology used and data employed, either a countercheck or a critical review performed by a panel of experts would be required, as with any LCA.

⁹⁰ Sam Altman, 2025, <https://blog.samaltman.com/the-gentle-singularity>

Google published the results of their own assessments of the impact of a prompt during inference. Mistral (2025) evaluated the impact of the inference of the Le Chat chatbot per 400-token long⁹¹ response (around one page of text) at 1.14 gCO₂e emitted, 45 mL of water consumed and 0.16 mg of abiotic resources consumed. Google (2025) evaluated the impact of a median Gemini textual prompt at 0.24 Wh of energy, emitting 0.03 gCO₂e and consuming 0.26 mL of water. While these last estimates are more fully documented, the lack of transparency (no details on the methodology or primary data used) and the use of certain assumptions (e.g. use of a market-based approach to assess the greenhouse gas emissions tied to energy consumption in inference phase) limit the scope of their relevance.

By and large, the lack of benchmarks and a harmonised methodology make it difficult to compare the results of corporate news releases, which are not based on the same assessment methods. This also makes it difficult to assess the environmental impact of substituting a non-AI practice (e.g. querying a search engine) with a generative AI practice AI (see inset below).

Can the environmental impact of a search engine query be compared to that of a generative AI prompt?

Several studies on digital behaviour point to a growing trend towards replacing search engines with generative AI tools, particularly chatbots. According to an Ipsos study⁹² conducted in 2024, 48% of the French respondents surveyed reported using generative AI services to search the web. The 2026 Digital Barometer⁹³ likewise found that information retrieval is the most common use of generative AI: 73% of users enter this type of prompt at least once a month and 21% do so every day. Arcep highlighted this trend in its January 2026 report on the relationship between AI and the open internet.⁹⁴

This shift raises questions about the environmental consequences of replacing one type of service with another and, more specifically, about how the impacts of search engines and generative AI chatbots compare. However, the lack of transparency from both search engine providers and generative AI developers makes such comparisons difficult. Drawing on a range of estimates, Vanderbauwhede et al. (2024) conclude that a ChatGPT prompt could consume between 50 and 90 times more energy than a Google search, depending on the assumptions used. If generative AI were to become widely adopted for routine tasks such as web searches, this could therefore lead to higher overall energy consumption.

Other experts, however, contend that chatbots may provide more accurate answers from the outset than search engines. Duprat (2026)⁹⁵ argues that generative AI can compare favourably when the analysis focuses on the service actually delivered rather than on a simple one-to-one comparison between a search engine query and a generative AI prompt. Search engines rely on an advertising-based business model and the associated advertising auction system accounts for between 15% and 44% of the total energy consumed by a query without providing any information to the user. In the author's view, what should be compared is the total energy required to obtain a piece of information. Several search engine queries may be needed to reach a result that a generative AI service can provide in a single interaction.

⁹¹ Tokens are units of data processed by AI models during training and inference to enable prediction, generation and reasoning.

⁹² Ipsos & CESI, 2025, [Intelligence artificielle : quels sont les usages des Français ?](#)

⁹³ Digital Barometer, 2026 ([link](#)).

⁹⁴ Arcep, 2026, [“Generative AI: Challenges for the future of the Open Internet”](#).

⁹⁵ Charles Duprat, 2026, [The Thermodynamic Efficiency Inversion](#)

Moreover, rather than replacing one another, generative AI and search engines increasingly appear to be used together. The most popular search engines now include AI-generated summaries alongside lists of websites, making it more difficult to distinguish and compare the environmental impact associated with overall generative AI use (via chatbots) and that associated with overall search engine use.

There is therefore currently no consensus on the comparative environmental impact of these two services, owing to both the difficulties of evaluation and the wide variety of use cases. Nevertheless, these challenges do not prevent the identification of certain good practices that can help users reduce their environmental footprint, such as avoiding the use of generative AI for simple searches and favouring resource-efficient AI systems for more complex queries⁹⁶. More precise prompts, which shorten inference time, could also enable users to reduce the energy consumption associated with their prompts by as much as 50% (UNESCO, 2025).

1.5.2.2 Impact that varies considerably depending on the type of task performed

Although assessing and comparing impacts remains difficult, scientific research broadly agrees that the environmental impact of AI during the inference phase depends greatly on the nature of the task.

Available estimates suggest that image generation consumes around 60 times more energy than text generation, on average (2.9 kWh compared with 0.047 kWh for text generation). General-purpose models also tend to consume more energy than specialised models (Luccioni et al., 2023). In addition, generating text is estimated to be around 25 times more energy-intensive than text classification⁹⁷ (Luccioni et al., 2023). According to preliminary estimates from the International Energy Agency, generating a six-second video would require approximately 115 Wh (IEA, 2025).

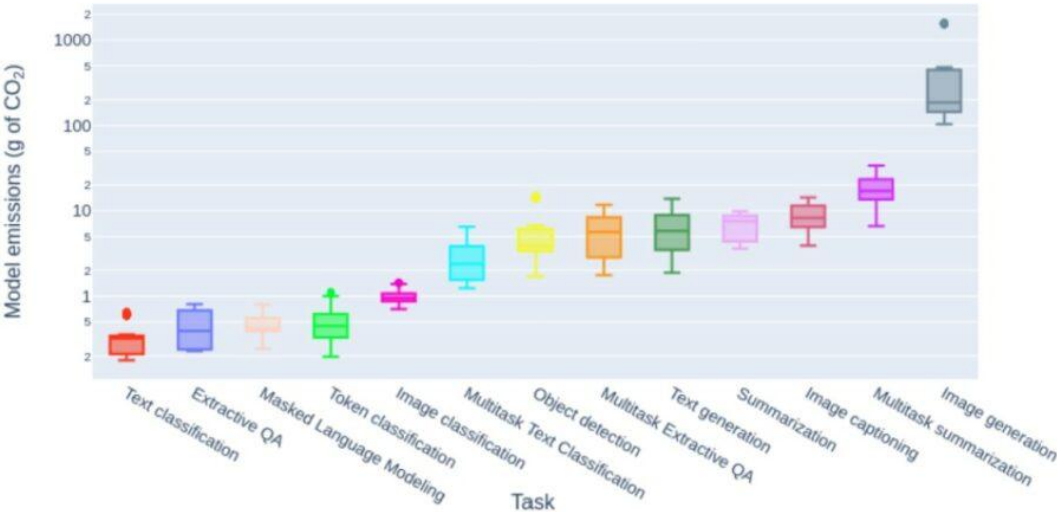


Figure 14: CO₂e emissions by task type. Source: Luccioni et al., 2023

⁹⁶ In a recent article, *Le Monde* highlighted seven practical ways of reducing the carbon footprint associated with AI use ([link](#))

⁹⁷ Text classification is a machine-learning task in which predefined labels are assigned to text data so that they can be automatically grouped into categories

1.5.2.3 The groundbreaking study conducted by PEReN, taking a closer look at the key drivers of energy consumption in generative AI models during the inference phase

To deepen understanding of the environmental impact of inference and provide more robust data, Arcep commissioned PEReN to carry out an *in situ* study. The aim was to measure the energy consumption of generative AI models during the inference phase and analyse the relationship between energy consumption and model performance (understood as the model's ability to deliver the expected response) for specific use cases. By assessing around twenty models under controlled conditions, this groundbreaking study sought to identify the main factors underlying energy consumption during inference.

The experimental methodology used in the PEReN study

PEReN's research was carried out on the Jean Zay supercomputer,⁹⁸ using a methodology described in detail in the full study report published separately in the appendix⁹⁹. The use cases selected for the study were based both on a state of the art of the scientific literature covering consumer practices (including everyday uses, coding assistance, business applications and creative uses) and on data from Compar:IA.¹⁰⁰ Once these themes had been identified, benchmarks¹⁰¹ were selected to assess how well the models performed on each task, in other words their ability to provide the expected response.

The 22 generative AI models included in the study were open-weight models containing between 3 and 123 billion parameters. They were available through Hugging Face, the leading platform for hosting AI models, and released between June 2024 and September 2025. Using open-weight models enabled PEReN to measure energy consumption directly during inference on a single well-known infrastructure, making it possible to produce exploitable results on the main drivers of energy consumption during inference. However, this methodological choice meant that neither the largest models (with more than 125 billion parameters) nor the most widely used consumer-facing models, such as ChatGPT, could be included in the tests. It also restricted the number of specialised models that could be examined.

⁹⁸ This work used AI computing and storage resources provided by IDRIS under resource allocation 2025-AD011016936, awarded by GENCI on the A100/H100 partitions of the Jean Zay supercomputer.

⁹⁹ This test was conducted using a constant hardware configuration on a single machine type. For any given model, the choice of GPU has a significant effect on energy consumption, as each GPU has its own energy consumption profile.

¹⁰⁰ Compar:IA is a platform for comparing generative AI models. It enables the creation of preference datasets based on real-world use cases expressed in European languages and focuses on themes such as cultural and linguistic bias, environmental impact, model diversity, critical thinking and societal issues.

¹⁰¹ An evaluation dataset that comprises: (i) a set of questions relating to a specific topic, (ii) the expected answers and (iii) a methodology for assessing the quality score (referred to throughout the remainder of this report as performance) by comparing the tested model's output with the expected response.

Energy consumption during inference was measured using the CEEMS¹⁰² tool integrated into the Jean Zay supercomputer. The results present only GPU energy consumption, as CPU power consumption displayed considerable variability even when carrying out the same task.¹⁰³

The study finds that, when both the use case and hardware remain unchanged, the number of model parameters is the main factor driving electricity consumption during inference, although it is not the only one. Model architecture and the use of a dedicated "reasoning" mode also have a significant influence on energy consumption.

The results confirm that the largest models are generally the most energy intensive. However, they also show that models of very different sizes can consume similar amounts of energy. In some cases, large models consume as much energy as, or even less than, smaller models with fewer total parameters because of the influence of other factors discussed below.

In addition, reducing energy consumption does not necessarily come at the expense of performance. Some low-energy models achieve performance levels equivalent to those of models that consume significantly more energy.

PEReN's findings show that certain model architectures developed in recent years¹⁰⁴, notably the "Mixture of Experts"¹⁰⁵, approach, can substantially reduce the energy consumption of large models. Models using this architecture consume, on average, 45% less energy than dense models of equivalent size. The study also finds that model-compression techniques (quantisation)¹⁰⁶ can significantly lower electricity consumption during inference. The effect of quantisation (8-bit or 4-bit) on energy consumption is substantial, with an average reduction of 39%. The impact of reasoning, however, varies considerably depending on the task being performed, and can increase energy consumption by as much as 849% in code-generation tasks.

The activation of a reasoning mode¹⁰⁷ is a major factor affecting energy consumption during inference because it increases token generation. Across the models tested, enabling reasoning mode increased energy consumption by up to 92% on average, while the resulting performance improvements varied significantly from one task to another.

Overall, the PEReN study concludes that the best performing models are not necessarily the most energy-intensive, and that the most energy-intensive models are not necessarily the best performing. Reducing the energy footprint of AI therefore does not always involve a compromise in performance.

¹⁰² The tool is freely available on GitHub [here](#).

¹⁰³ It should also be noted that excluding CPU energy consumption from the scope of the analysis reduces the model's total energy consumption during inference by approximately one third. PEReN concluded that CPU energy consumption was too variable for the same task to be reliably included in the experiment, since two assessments carried out under identical conditions could produce substantially different CPU energy consumption figures.

¹⁰⁴ In 2024, relatively few dense (non-MoE) models had been released, whereas MoE models saw a marked increase in adoption from September 2024 onwards.

¹⁰⁵ The Mixture of Experts (MoE) approach consists of dividing a model into separate subnetworks (the experts), each specialising in a subset of the input data, in order to jointly perform a task. Models that do not use this approach are referred to, by contrast, as "dense". During inference, an MoE model uses only the part of the network that it requires. As a result, the number of so-called "activated" parameters is lower than the model's total number of parameters.

¹⁰⁶ These techniques consist of reducing the size of the model's parameters (and therefore their precision) without reducing performance. For example, a model may be compressed from 16-bit to 8-bit precision. By analogy, this is similar to reducing the number of decimal places used to represent a number in mathematics.

¹⁰⁷ Reasoning models became increasingly common on the market from September 2024 onwards.

Finally, although the tests conducted by PEReN used a fixed hardware configuration to isolate the effects of different variables, the choice of computing infrastructure remains an important factor. Both the literature review and the empirical results indicate that the type of GPU used is a key determinant of energy consumption (Chen et al., 2024) and that innovations in chip technology can play a significant role in reducing energy use (Kachris, 2025).

Optimisation techniques to reduce AI models' energy consumption

A number of advances in reducing the energy consumption of AI models have already been achieved through optimisation techniques. Among these, the Mixture of Experts (MoE) architecture consists of dividing a model into several specialised sub-models that are activated according to the prompts received. There are also model-compression techniques such as quantisation, highlighted in a recent article (Delavande et al., 2026) and by experimental research conducted by University College London (UCL) for UNESCO¹⁰⁸ (2025), which show that the technique could reduce energy consumption by as much as 44%, thereby lowering the operating costs of Large Language Models (LLMs). Other techniques, such as Low-Rank Adaptation (LoRA), distillation and pruning, seek to reduce or simplify model weights in order to lessen the hardware resources required for use. In addition, Retrieval-Augmented Generation (RAG) allows models to access external information that was not present during training, thereby avoiding the need for full end-to-end retraining. It should be noted that although such optimisation techniques are necessary, they may not be sufficient if rising levels of use lead to a rebound effect.

Another factor to be considered is the carbon intensity of the electricity mix powering the servers used for inference. Unlike model training, however, inference is subject to latency requirements. As a result, there is less flexibility in choosing where inference servers are located than there is for training infrastructure.

¹⁰⁸ The prompts were tested on NVIDIA GeForce RTX 3090 Ti GPUs paired with a 12th Gen Intel Core i9-12900K processor and 128 GB of RAM. The protocol is described in detail in the appendix to the report.

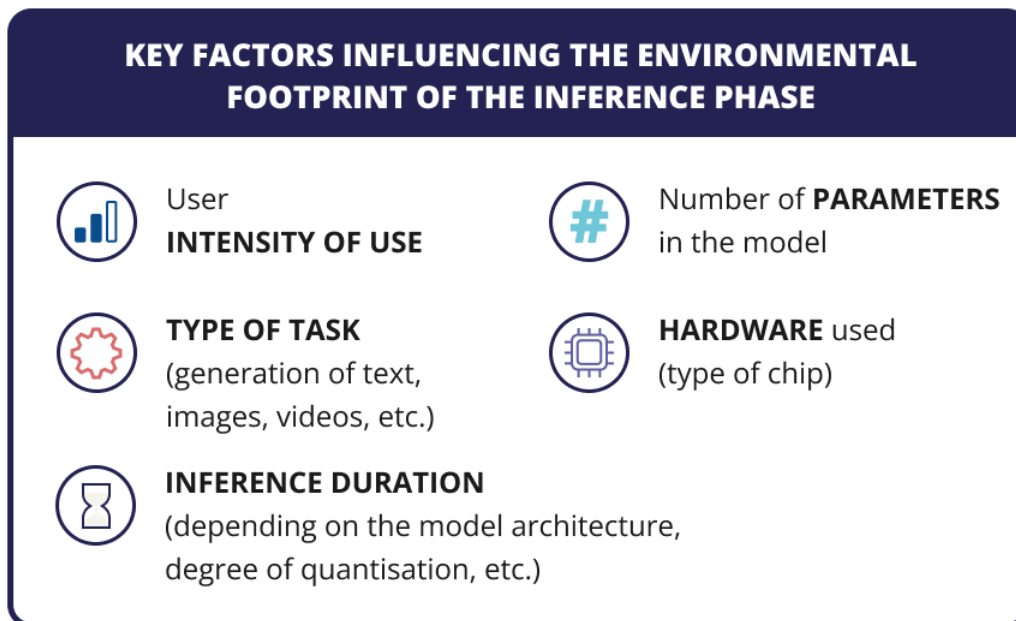


Figure 15: Key factors influencing the environmental footprint of the inference phase of a generative AI model. Source: Arcep.

*
**

In conclusion, the phase-by-phase analysis suggests that although the training of generative AI models generates very substantial environmental impacts, inference is likely to become the most frequently occurring phase as AI adoption expands, particularly for widely used services such as ChatGPT. Beyond a certain number of prompts, the greenhouse gas (GHG) emissions associated with usage may therefore exceed those generated during training, depending on the size of the user base and how frequently the service is used (Data For Good, 2023). At present, it is difficult to conclude that either training or inference has a greater environmental impact than the other, as this depends both on the model analysed (and the number of times it has been retrained¹⁰⁹) and the number of users (De Vries, 2023).

More broadly, the environmental impacts associated with both training and inference across the AI sector will depend on how the market evolves over the coming years (see Section 1.2). Innovation pathways could either continue the current trend towards increasingly large models with ever more parameters, or lead to the development of smaller models (see boxed text).

Looking ahead, important questions also arise regarding innovations such as agentic AI services¹¹⁰, which could significantly amplify the environmental footprint of generative AI during inference. This new generation of services combines the conversational capabilities of LLM with automation functions (Arcep, 2026a). Initial studies suggest that agentic AI systems may generate greenhouse gas emissions and water consumption up to 60 times greater than those associated with a conventional generative

¹⁰⁹ A model may be retrained several times, or fine-tuned, in order to improve its accuracy.

¹¹⁰ These are generative AI systems that incorporate the ability to interact with other digital services (search engines, calendars, messaging platforms, office productivity tools, social media platforms and third-party applications).

AI service (Green IT, 2025). These findings should, however, be interpreted cautiously given the very recent nature of literature available, as well as the lack of transparency surrounding such systems.

"Small is beautiful": Are smaller models a better choice from a performance and environmental perspective?

Ahead of the AI Action Summit in Paris in February 2025, the National Institute for Research in Digital Science and Technology (Inria) published a white paper whose central recommendations included favouring specialised models over general-purpose models, as they typically contain fewer parameters. According to the Institute, emerging technologies often follow a natural path towards increasingly general-purpose tools. Yet specialised models may offer a way to combine innovation with environmental sustainability (Inria, 2025; Varoquaux, 2025).

Research shows that the volume of data used to train a model, and therefore the number of parameters, is not necessarily a reliable indicator of model performance, understood here as the ability to deliver the expected response. Sophisticated AI systems designed for specific tasks can sometimes outperform fine-tuned general-purpose models while requiring fewer resources. Varoquaux et al. (2025) found that, across a range of applications, extending the number of parameters, training time or GPU resources devoted to training does not automatically lead to better performance (see Appendix 7).

Small Language Models (SLMs) can also achieve very strong results. According to UNESCO (2025), using SLMs for targeted applications could reduce energy consumption by as much as 90% compared with Large Language Models (LLMs), while maintaining a high degree of accuracy. Beyond their energy advantages, SLMs are better suited to environments with limited resources or connectivity, and provide faster response times.

By way of illustration, researchers from Nvidia and the Georgia Institute of Technology consider that SLMs are powerful enough to support agentic AI systems, and that replacing LLMs with SLMs in these contexts could generate significant environmental benefits through reduced resource consumption (Belcak et al., 2025). However, the substantial investments already made in high-performance data centres used for LLM training and inference may act as a barrier to adoption. These investments can create a form of techno-economic lock-in, encouraging organisations to continue relying on LLMs in order to maximise returns on these investments. Belcak et al. (2025) also note that SLMs receive far less attention and promotion than LLMs.

Nevertheless, the deployment of SLMs may have limitations. First, many smaller models are themselves derived from LLMs¹¹¹, which reduces their environmental advantage compared to families of smaller models that do not derive from LLMs¹¹² and that have therefore been trained using a smaller number of parameters. Second, according to experts interviewed, wider deployment of SLMs could lead to SLM usage being layered on top of existing LLM usage, thereby reducing the environmental benefits that SLMs might otherwise provide.

¹¹¹ This is the case for models such as BERT Mini, Llama3.2-1B, Qwen2.5-1.5B and DeepSeek-R1-1.5B.

¹¹² This is the case for the Gemma family developed by Google, the Ministral family from Mistral AI and the Phi family from Microsoft.

Chapter 3 Our recommendations for making AI development compatible with planetary boundaries

As digital technologies and the infrastructures that support them become ever more important to the economy and to everyday life, ensuring their sustainability is becoming increasingly necessary. In the case of generative AI, although information on its environmental impact remains incomplete, current observations and projected trends already point to the need for action to ensure that this technology is developed and deployed in a manner that remains sustainable for future generations.

A better understanding of the issues surrounding generative AI market dynamics, together with its environmental footprint, helps to identify relevant courses of action. In this final chapter, Arcep sets out four strategic areas of work aimed at making the development of AI compatible with planetary boundaries.

These proposals combine soft-law and regulatory instruments and build on the current and forthcoming European legislative framework. They are intended to inform public debate and support future work by regulators and other stakeholders on generative AI.

Axis 1: Improve the assessment and understanding of AI's environmental impact

As emphasised throughout this report, there is currently a clear shortage of information needed to fully assess the environmental impact of AI, largely because of the lack of transparency among stakeholders at all levels of the value chain. Although academic and institutional research provides valuable insights, identifying, prioritising and steering the actions required depends on a better assessment of impacts. Establishing a robust evidence base and assessing impacts using standardised methodologies therefore requires reliable data gathered from stakeholders throughout the value chain.

Recommendation 1. Establish the collection and publication of AI environmental data by public authorities

In response to the rapid acceleration of artificial intelligence use, the European Union has embarked on an ambitious programme of investment to support AI development. Yet the environmental impact of these technologies remains only partially documented (see Chapter 2), despite the fact such knowledge is essential if the objectives of competitiveness and sustainability are to be reconciled for an artificial intelligence from which future generations can benefit.

Against this backdrop, the ability of public authorities to make informed decisions depends directly on access to reliable and harmonised data. Data-driven regulation therefore represents a key instrument for delivering on European ambitions. Without a robust framework for measuring impacts, it becomes difficult both to give practical effect to, and to assess progress towards, the objective of sustainable AI.

The European Union should therefore implement the means to produce a robust evidence base on the environmental footprint of AI by giving national electronic communications regulators responsibility for collecting environmental data across the AI value chain. In this regard, provisions currently being considered under the Digital Networks Act (DNA), particularly Article 115(1)(c), which envisages the collection of data from digital-sector actors, including on sustainability matters, represent a promising development and open a pathway towards effective data-driven regulation at the European level.

Achieving this objective requires going beyond existing mechanisms, which remain insufficient to support a structured and usable system of data collection. While certain provisions of the AI Act introduce documentation requirements for AI providers and transparency obligations for relevant actors (see the inset on the AI Act below, particularly Articles 53(1)(a) and 95), they do not amount to a genuine framework for collecting environmental data. Instead, they rely largely on the capacity and willingness of market participants to produce such information, without ensuring its availability, reliability or consistency across the European Union. In particular, they leave unanswered the issue of access to key information, such as the energy consumption associated with training AI models or with inference, which remains a significant blind spot in the Regulation where providers do not systematically assess or estimate these impacts themselves.

Recent experience with the regulation of data centres clearly demonstrates the limits of this type of approach. In response to the lack of reliable information on data centres' energy consumption and environmental footprint, the European Union introduced, as part of the revision of the Energy Efficiency Directive (EED), a specific requirement for harmonised reporting at the European level (see inset below). This framework requires the collection of detailed data using standardised formats in order to build a robust evidence base for public policy. While this clear legal framework should help reduce information asymmetries, it applies only to data centres, which are just one part of the AI value chain. Other critical segments, including semiconductor manufacturers and AI model providers, remain largely outside any structured data-collection framework. The absence of harmonised information for these segments makes it impossible to obtain a comprehensive view of AI's environmental footprint. In this context, giving regulatory authorities an explicit mandate to collect environmental data appears to be an appropriate way of improving the assessment of AI's environmental impacts across the value chain as a whole.

Nevertheless, any such mandate should be implemented in a proportionate and administratively efficient manner. In line with the "once-only" principle, care should be taken to avoid creating multiple reporting obligations for businesses and to make the greatest possible use of existing mechanisms. In this respect, some of the proposals currently under consideration in the DNA could provide a useful basis for action. Alongside the proposed data-collection role for regulatory authorities, the DNA also envisages the development by BEREC of European guidelines governing procedures and requests for information. Building on these measures by assigning regulatory authorities a clearly defined role and explicit remit for collecting data across the entire AI value chain (including AI model providers) would enable them to operate as independent and transparent trusted intermediaries. Such a development would significantly strengthen the assessment of AI's environmental footprint across the whole value chain, and would therefore contribute directly to the European objective of ensuring that AI develops in a manner that is both competitive and sustainable.

The Energy Efficiency Directive

Directive (EU) 2023/1791 on energy efficiency seeks to reduce businesses' energy consumption and greenhouse gas emissions. It establishes a target of reducing final energy consumption across Europe by at least 40% by 2030 compared with 2007 levels.

Following its revision in 2023, the Directive now requires data centres with an IT power draw of more than 500 kW to report information on their energy consumption together with the environmental indicators specified in its Article 12. The European Commission is required to create a European database of data centres that will consolidate the information reported by operators.

Member States need to encourage the owners and operators of data centres located within their territory with a power draw of at least 1 MW to take account of the good practices set out in the latest version of the European Code of Conduct on Data Centre Energy Efficiency.

Lastly, drawing on the data collected, the European Commission must develop legislative proposals to improve energy efficiency and assess the feasibility of achieving “carbon neutrality in the data centre sector”. This could lead to the establishment of a timetable requiring existing data centres to “comply with minimum performance standards”.

Recommendation 2. Use internationally standardised environmental impact assessment methodologies to facilitate comparisons between AI systems

A more effective assessment of AI's environmental impact across the value chain requires methodologies that are robust, reliable (multi-criteria, multi-component and multi-stage) and harmonised in order to enable comparisons between models and services.

Standardised Life Cycle Assessment (LCA) methodologies are already available and should be promoted and widely adopted in order to avoid the proliferation of approaches and to ensure that assessments are carried out on a common basis. These methodologies are set out in the General Framework for Frugal AI, developed by AFNOR in partnership with the Ministry for the Ecological Transition and with the participation of Arcep. The framework can be directly used by stakeholders to conduct comprehensive quantitative and qualitative assessments of AI services. It draws in particular on recognised academic work (AFNOR, 2024; Ligozat et al., 2022) and on existing guidance, including ITU-T Recommendation L.1410.

At the European level, Article 40 of the AI Act (see box below) could provide a basis for the development of a common methodology. In its Apply AI Strategy published in October 2025, the European Commission announced that it would shortly issue a standardisation request to European standards bodies concerning common reporting and documentation processes for the impact of AI systems and general-purpose AI models on energy consumption¹¹³. This is a positive initiative that deserves to be taken forward, as it could establish a harmonised European framework for measuring the environmental impact of AI systems.

At the international level, work has already been published such as ITU-T Recommendation L.1801, which is designed to assess the environmental impact of AI systems. The recommendation provides a coherent international framework for assessing AI's environmental footprint, including guidance on comparing two AI systems or comparing an AI system with a non-AI alternative, and will contribute to ongoing work on the development of an environmental rating or score (ITU, 2025).

¹¹³ European Commission *Apply AI Strategy*, COM(2025) 723 final, 08/10/2025, Brussels, p.9, [Apply AI strategy](#) | Shaping Europe's Digital Future

In particular, there is a need to ensure that the various initiatives aimed at standardising methodologies for assessing AI's environmental impact are properly aligned. These include initiatives led by the ITU, the European Telecommunications Standards Institute (ETSI) working alongside the European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENELEC), the Institute of Electrical and Electronics Engineers (IEEE) and the International Organization for Standardization (ISO)¹¹⁴. In this regard, a roadmap on the standardisation of AI environmental impact was published during the AI Action Summit in Paris. Developed by more than 30 public and private organisations, it sets out guidelines for the environmental assessment of AI and identifies areas for cooperation between the principal international standards organisations, namely ISO, ITU, UNESCO and IEEE. A mapping of the various initiatives relating to the standardisation of environmental impact assessment methodologies for AI is provided in Appendix 6. More broadly, international cooperation remains essential to coordinate efforts towards more sustainable AI, particularly between established stakeholders in the energy sector and those operating across the digital ecosystem¹¹⁵.

¹¹⁴ The French Competition Authority, *Autorité de la Concurrence*, has also highlighted the importance of standardisation in ensuring competition between market participants on the basis of their respective merits. To limit competition risks, however, the Authority emphasises the need to maintain an appropriate balance within standardisation bodies between representatives of private interests and public authorities. It also stresses the importance of transparency towards users regarding the operation of a standard and of independent verification of the information provided by stakeholders during the development of the standard. Certain standards may have limitations where they are biased in favour of particular market participants. Conversely, standardisation may be hindered by the behaviour of actors that fail to provide the information necessary for the development or implementation of a standard.

¹¹⁵ By way of illustration, the International Energy Agency (IEA) has also launched the first global observatory dedicated to energy and AI. By centralising and analysing data, the observatory aims to: (i) improve the forecasting of the energy requirements of data centres and AI models on the basis of a transparent methodology; (ii) optimise energy systems and reduce carbon emissions through the use of AI; and (iii) promote breakthrough innovations in the application of AI to energy.

The AI Act

Regulation (EU) 2024/1689 establishes a harmonised framework for the development, marketing and use of artificial intelligence services. It adopts a risk-based approach and introduces obligations relating in particular to safety, transparency, data governance, technical documentation and human oversight. The Regulation also contains specific provisions for so-called general-purpose AI models (GPAI), particularly in the area of generative AI.

Article 40, paragraph 2 allows the European Commission to issue a standardisation request to the relevant organisations (CEN, CENELEC and/or ETSI) with the aim of reducing energy consumption and other resource use across the life cycle of AI services and improving the energy efficiency of general-purpose models.

Article 95 states that the European Commission (through the AI Office) and the Member States must encourage stakeholders to develop voluntary codes of conduct aimed at minimising the environmental impact of AI services.

Article 112 provides for an assessment of the two preceding provisions: a progress report on standardisation must be published by 2 August 2028 and every four years thereafter, while a report on the impact and effectiveness of voluntary codes of conduct must be published by 2 August 2028 and every three years thereafter.

Article 53(1)(a) requires providers of GPAI models to supply technical documentation that includes the assessed or estimated energy consumption of their models or, where such information cannot be provided, at least the computing resources used.

Axis 2: Promote the ecodesign of AI services as a strategic lever for European competitiveness

Digital ecodesign principles provide a practical means of ensuring that environmental considerations are embedded in AI services from the design stage and throughout their life cycle. Drawing on existing best practices can help minimise the environmental impact of AI services and may become a source of competitive advantage in today's market environment. Ecodesign can also contribute to stronger European competitiveness by improving resource efficiency and encompassing digital sobriety practices that align services more closely with the needs of end users.

Recommendation 3. Integrate the ecodesign of AI service into the European regulation of providers

Arcep aims to support the emergence of standard-setting, incentive-based and regulatory initiatives on the ecodesign of digital services, primarily at the European level but also internationally.

Within the European Union, physical products are covered by extensive regulation, whereas services, including AI services, are not. Soft-law initiatives have emerged at the Member State level to encourage the ecodesign of digital and AI services, notably the RGEN (General Policy Framework for the Ecodesign of Digital Services, Arcep & Arcom, 2024), the General Framework for Frugal AI (Afnor, 2024) and the recently published ISO Digital Services Ecodesign standard. In light of the rapid development of generative AI services, these frameworks will need to be updated regularly to reflect technological advances, with input from experts across the sector.

However, non-binding instruments such as frameworks and codes of conduct may not be enough. Although public authorities can encourage the ecodesign of AI services through these mechanisms, they may provide insufficient incentives where environmental performance is not a significant factor in competition, particularly when providers do not directly bear the negative externalities associated with the use of their products or services. For example, higher electricity consumption by user devices does not generate direct costs for AI service providers that would encourage them to improve the efficiency of their services. It can, however, increase the frequency of battery charging and discharging cycles, thereby shortening the lifespan of the device's battery.

Further action is therefore needed to accompany the rapid growth of AI services and address the environmental impacts associated with their development. Arcep accordingly supports the need to implement a European framework that would integrate AI service ecodesign requirements into the regulation of service providers.

Various initiatives currently being discussed at the European level, including the Digital Networks Act (DNA) and the Digital Fairness Act (DFA), could help advance this agenda. Arcep has, for example, contributed to a number of European Commission public consultations on AI to promote consideration of the environmental issues associated with its development and use (Arcep, 2024; Arcep, 2025d; Arcep, 2025e; Arcep, 2026b). In its response to the public consultation on the forthcoming DFA, the draft of which is expected to be published in 2026, Arcep emphasised that this legislation could provide a useful framework for strengthening consumer choice and improving the environmental performance of digital products and services. More specifically, it advocated the introduction of ecodesign requirements for AI services that would encourage the use of more resource-efficient models, building on measures already identified in frameworks such as the General Policy Framework for the Ecodesign of Digital Services (RGEN – see inset below) and the General Framework for Frugal AI. In addition, the draft Digital Networks Act published in January 2026 expands the objectives of digital network regulation to include traffic efficiency and the sustainability of digital services. In practical terms,

providers of digital services, including AI services, would be required to implement ecodesign measures designed to reduce unnecessary data traffic and the resulting energy consumption.

Through one or more future legislative initiatives, the European Union could therefore establish instruments that promote the development of AI services aligned with its environmental objectives. Over time, the efficiency gains delivered by such services could become a significant competitive advantage in a context of growing pressure on resources.

Focus on the "Algorithm" section of the General Policy Framework for the Ecodesign of Digital Services (RGESN)

A range of good practices and recommendations for the ecodesign of AI services, particularly generative AI services, is already available. For example, the General Policy Framework for the Ecodesign of Digital Services¹¹⁶ (RGESN, Arcep and Arcom, 2024) sets out a series of good practices that designers can adopt as part of a voluntary ecodesign approach for services incorporating AI systems. The RGESN emphasises the importance of clearly identifying both the service requirements and the needs of target users so that the AI model can be designed accordingly. Once the suitability of the model has been defined, the most resource-efficient options should be selected, such as relying on pre-trained models wherever possible. It also recommends ensuring that algorithm training is proportionate to the essential needs of the service, particularly with regard to the volume, frequency, compression and collection of data. Beyond practices specific to the algorithmic model itself, the use of sustainable hosting solutions and optimising user interfaces are also key to reducing the environmental footprint of AI-based services throughout their lifecycle, from model training through to inference.

In its "Algorithms" section, the RGESN sets out seven criteria aimed at embedding ecodesign and resource-efficiency principles into the training and inference phases of the algorithmic models used in AI systems:

- 9.1 - Has the digital service questioned the need for a training phase to avoid unsubstantiated and unreasonable usage?
- 9.2 - Does the digital service use a training phase with a minimised level of complexity proportionate to the actual use of the service?
- 9.3 - Has the digital service implemented mechanisms for limiting the amount of training needed for it to operate?
- 9.4 - Does the digital service limit the amount of data used for the training phase to what is strictly necessary?
- 9.5 - Does the digital service optimise model updates and retraining according to its own needs and those of target users?
- 9.6 - Does the digital service use compression techniques for the models used during the training phase?
- 9.7 - Does the digital service use an inference strategy that is optimised in terms of resource consumption and target users?

¹¹⁶ The General Policy Framework for the Ecodesign of Digital Services is a technical reference document designed for digital professionals and specialists seeking to implement an ecodesign approach for digital services, including websites, applications, AI systems, software and APIs. It was developed by Arcep and Arcom in partnership with ADEME, DINUM, CNIL and Inria. The latter played a particularly significant role in developing the Algorithms section.

**LES CRITÈRES DU RÉFÉRENTIEL GÉNÉRAL DE L'ÉCOCONCEPTION
DES SERVICES NUMÉRIQUES POUR UNE IA FRUGALE**



La phase d'entraînement est justifiée par rapport aux cibles et besoins du service



La complexité, la quantité et la fréquence de l'entraînement est proportionnée à l'usage effectif du service



L'hébergement utilisé sur toute la chaîne de valeur du service permet de limiter les impacts environnementaux associés



La quantité de données utilisée pour la phase d'apprentissage est limitée au strict nécessaire



La phase d'inférence prend en compte les enjeux environnementaux et correspond aux besoins des cibles utilisatrices



Des techniques de compression sont utilisées pour les modèles de la phase d'entraînement

**THE CRITERIA OF THE GENERAL POLICY FRAMEWORK FOR THE ECODESIGN
OF DIGITAL SERVICES FOR FRUGAL IA**

The training phase is justified in relation to the target users and the needs of the service	The complexity, quantity and frequency of training is proportionate to the actual use of the service	The hosting used throughout the service value chain is designed to minimise the environmental impact
The quantity of data used for training is limited to what is strictly necessary	The inference phase integrates environmental considerations and is tailored to the needs of target users	Compression techniques are used for the models used during the training phase

In addition to the "Algorithms" section, developers of generative AI services can draw on other RGENS criteria to implement best practices, including:

- Assessing the need to use generative AI in order to deploy it only where it is most needed¹¹⁷,
- Streamlining and optimising model training to reduce its frequency and align energy consumption with electrical grid constraints,
- Using pre-trained models and existing datasets, including those originating from research, to avoid systematic retraining involving high energy costs,
- Systematically reporting environmental footprint indicators for both AI model training and inference,
- Choosing sustainable hosting solutions across the entire AI service value chain¹¹⁸,

¹¹⁷ In connection with criterion 9.1 "Has the digital service questioned the need for a training phase to avoid unsubstantiated and unreasonable usage?" and criterion 1.1 "Has the digital service been assessed favourably in terms of utility by taking its impact on the environment into account?"

¹¹⁸ In connection with criterion 8.1 "Does the digital service use a hosting provider that has a strategy to reduce its environmental footprint" and criterion 8.3 "Does the digital service use a hosting provider with a low PUE (Power Usage Effectiveness)?"

- Providing an optimised user journey through AI-based conversational agents, including by limiting attention-grabbing practices and misleading interface design¹¹⁹,
- Informing users when prompts have a significant environmental impact¹²⁰,
- Encouraging users to reduce their environmental impact and allowing them to remove questions and suggestions automatically generated by the AI after responding to a prompt.

Recommendation 4. Strengthen eco-conditionality in innovation support and public procurement rules

National and local public authorities can help encourage more sustainable AI models and services by strengthening eco-conditionality¹²¹ in innovation support schemes¹²² and public procurement rules¹²³. Such measures would align with broader public policy objectives by supporting AI development while keeping environmental impacts under control¹²⁴.

Environmental criteria could be incorporated into AI innovation support programmes at both national and European levels. For example, certain public incentives already exist, particularly regarding eco-conditionality, through the France 2030 innovation support programme. Applicants must use the Green Algorithms tool¹²⁵ (Lannelongue, 2021) to assess the environmental impact of the AI model for which funding is sought. Using a shared assessment framework enables environmental indicators to be considered when evaluating applications and awarding funding. At the EU level, public support could similarly be tied to a standardised environmental assessment based on a harmonised methodology. Such an eco-conditionality mechanism could, for instance, apply to future funding rounds under the GenAI4EU initiative¹²⁶.

Beyond innovation funding schemes, incorporating AI frugality into public procurement requirements would demonstrate leadership by example and help stimulate market demand, given the potentially significant scale of public sector procurement (AI service requirements across different levels of government). One practical approach would be to introduce a dedicated environmental criterion, with meaningful weighting in the overall evaluation of tenders, assessing the resource efficiency of AI services. To assess this environmental criterion, public contracting authorities could draw on established frameworks such as the RGEN or the General Framework for Frugal AI, both of which provide objective good-practice criteria. This requirement could ultimately be enshrined in legislation,

¹¹⁹ In connection with criterion 4.3 “Does the digital service optimise the clickstream for each main feature?” and criterion 1.6 “Does the digital service collect data in a responsible and judicious manner?”

¹²⁰ In connection with criterion 4.12 “Does the digital service flag features with significant environmental impacts to users?” and criterion 4.15 “Does the digital service provide users with a means to monitor their usage, to be able to track and reduce their associated environmental impacts?”

¹²¹ Eco-conditionality involves linking public funding or procurement decisions to compliance with environmental principles and performance criteria

¹²² Innovation support can take various forms, including grant programmes, innovation competitions and equity investments

¹²³ Public procurement encompasses all contracts awarded for payment by public bodies or authorities carrying out public service functions. In France, the applicable rules are set out in the Public Procurement Code.

¹²⁴ For example, the draft third edition of France’s National Low-Carbon Strategy (SNBC) introduces a digital sector chapter for the first time, with the objective of “*stabilising greenhouse gas emissions by 2030–2035 while preserving digital sovereignty*”. The strategy recommends “*encouraging the use of frugal AI*” to support “*the reduction of data centre impacts and improved data lifecycle management*” (Ministry of Ecological Transition, 2025).

¹²⁵ This toolset notably includes an online calculator that estimates energy consumption and carbon emissions using information about the algorithm, such as computing time, the number of processors used and the location of the computation

¹²⁶ European Commission, *GenAI4EU*, accessed in 2026

similar to Article 55 of the Law on Reducing the Environmental Footprint of Digital Technology (REEN), which states that “*public administrations, when making purchases, shall favour software whose design helps limit the energy consumption associated with its use*”.

In practical terms, in both innovation funding programmes and public procurement processes, public authorities could encourage the use of generative AI models that are appropriately sized for their intended purpose. Smaller models can be adapted for specific use cases¹²⁷, such as chatbots, code generation, translation, text summarisation and medical applications, while generally having a lower environmental footprint, especially during training. Public authorities could also favour organisations that follow recognised codes of conduct designed to reduce the environmental impact of AI services through ecodesign practices. One avenue to pursue in establishing a recognised European code of conduct would be to give practical effect to Article 95 of the AI Regulation (see text box at the end of Priority 1), which provides for voluntary codes of conduct aimed at minimising the environmental impact of AI systems and services.

¹²⁷ Hugging Face, John Johnson, [Small Language Models \(SLM\): A Comprehensive Overview](#), 2025

Axis 3: Enable users to choose their generative AI services on the basis of their environmental impact through an appropriate European regulatory framework

The development of generative AI services must be compatible with preserving end users' freedom of choice (Arcep, 2026a). Users must therefore be able to choose their services and functionalities, and more broadly the technologies they use in their daily lives. To this end, providers must give users the means to manage their usage and, consequently, their environmental impact.

Recommendation 5. Require greater environmental transparency from chip manufacturers and the main providers of AI models and services

Greater environmental transparency from actors across the value chain (chip manufacturers, model providers, service providers and others) would provide businesses and users with the information needed to make informed choices. The disclosure of environmental footprint data, based on reliable information and harmonised methodologies, is particularly important for enabling comparisons between generative AI models and services and, consequently, for ensuring the proper functioning of competition in these markets.

Information on the environmental impact of AI services and their associated uses is currently not readily accessible to users, whether they are public users, businesses or developers of AI services. Moreover, a study carried out in 2025 found that the people who use generative AI services most frequently are generally less aware of the associated environmental issues (Moravec et al., 2025). Developers of AI services that rely on third-party models could also benefit from clear and transparent environmental information in order to use the models best suited to their needs and with the lowest impact. Greater environmental transparency would help companies gain a clearer understanding of the impacts associated with generative AI, which may affect their ability to achieve their climate targets (Moravec et al., 2025).

In response to this information gap, public authorities should require greater environmental transparency from chip manufacturers and providers of AI models and services. Many companies are already required to report environmental information¹²⁸ under frameworks such as the EU Corporate Sustainability Reporting Directive (CSRD) or France's Greenhouse Gas Emissions Reporting scheme (BEGES). As part of these obligations, they gather data on issues including direct and indirect greenhouse gas emissions and energy use. Some organisations also undertake Life Cycle Assessments (LCAs) voluntarily and therefore already hold environmental information that could usefully be disclosed.

Public authorities could reinforce transparency requirements through a variety of measures. One option at the European level would be to introduce mandatory environmental labels, similar to existing energy labels¹²⁹, or certification schemes for services that disclose their environmental impacts and, in

¹²⁸ Before the Omnibus I Act, these reporting obligations applied to companies meeting at least two of the following three conditions: a balance sheet total of €25 million, net turnover of €50 million or an average of 250 employees during the financial year. They also applied to companies listed on a regulated European market and to non-European companies that, over the previous two consecutive financial years, generated more than €150 million in net turnover within Europe and operated a branch in France with net turnover above €40 million. The Omnibus I Act narrows the scope to companies with more than 1,000 employees and annual net turnover exceeding €450 million. For companies from third countries, the revised rules will apply only where the parent company generates annual net turnover of more than €450 million within the EU and the subsidiary or branch exceeds €200 million.

¹²⁹ An energy label could present information on the energy consumption of generative AI models and services across both the training and inference stages, including, for example, the total energy used during training and the energy consumed per prompt when using the service

time, for services designed according to ecodesign principles. Such schemes could be based on a multi-criteria methodology covering the entire life cycle of AI systems. The European Commission has already launched a study to develop an assessment framework and a European label for the energy consumption and greenhouse gas emissions of AI services¹³⁰. A public consultation was launched on 7 April to gather input from stakeholders¹³¹. If brought to fruition, this initiative would significantly improve the environmental transparency of AI services.

For AI model providers, in addition to the environmental information mentioned above, it would be valuable to have access to technical details such as model size, where and for how long training took place and the hardware used (particularly the chips) for training and inference. Including this information in the technical documentation required under Article 53(1)(a) of the AI Regulation (see text box at the end of Axis 1) would help identify the main factors driving the energy consumption of AI models.

Chip manufacturers should also make environmental information available, including Product Carbon Footprint data sheets and other environmental indicators. Given the substantial influence of chips on energy consumption during both training and inference (see Chapter 2), this would mirror existing environmental transparency requirements in sectors such as construction and transport. Improved transparency for newer generations of chips would provide a more reliable and detailed picture of energy consumption associated with AI and, more broadly, with data centres.

Recommendation 6. Ensure AI services remain open

As highlighted in an Arcep report¹³² on the challenges that AI poses to the open internet, end users' freedom to choose must also come with practical ways to act when using generative AI services. However, as digital services and generative AI become increasingly hybrid and AI features are built into interfaces, it can be difficult or even impossible for users to switch off generative AI services that are enabled by default, particularly in some applications.

At the same time, the default integration of generative AI within already widely used digital services strengthens the market power of vertically integrated players and increases the risk of harmful practices for healthy market dynamics. It also creates an extra barrier for competitors offering more environmentally responsible solutions.

In this context, Arcep supports using the Digital Markets Act (DMA) to keep generative AI services open. Several actors and services integrating AI features such as search engines, operating systems and browsers¹³³ are already covered by this regulation. However, some obligations need to be adjusted¹³⁴ so that users can freely choose AI services other than those of the gatekeeper when services are bundled¹³⁵ and so that third party AI agents can access the gatekeeper's digital tools and services (such as search engines) to respond to user prompts.

¹³⁰ European Commission, *Artificial Intelligence Act: Call for tenders to measure and foster energy efficient and low emission artificial intelligence in the EU*, 2024

¹³¹ European Commission, *Targeted consultation on measuring energy consumption and emissions of AI models and systems*, 2026

¹³² Arcep, 2026, [Generative AI: Challenges for the Future of the Open Internet](#)

¹³³ European Commission, 2024. [High-Level Group for the Digital Markets Act Public Statement on Artificial Intelligence](#).

¹³⁴ In particular through the adoption of delegated acts

¹³⁵ Article 6(3) of the DMA, which already allows end users to change the default settings of certain core platform services, could be extended to cloud services, search engines and instant messaging services, in order to prevent AI services from being integrated by default, or even exclusively, as is currently the case with Microsoft Copilot and Microsoft 365, and with Meta AI and WhatsApp, among others

Finally, the Data Act could also apply to certain providers of generative AI whose services may be regarded as cloud services, particularly 'AI as a Service' (AlaaS) offerings. Applying the Data Act could help open up generative AI services by making it easier for users to transfer their data and by improving interoperability between systems.

Providing users with options to configure settings, uninstall integrated generative AI services and benefit from interoperable systems would strengthen freedom of choice, give users greater control over the environmental impact of their digital activities and could encourage the development of alternative solutions that may be more sustainable.

Axis 4: Construct a development strategy for data centres in Europe that combines sovereignty and sustainability

Data centres are strategic digital infrastructures that support both European and national sovereignty. Their growth is also becoming increasingly important for economic competitiveness and the continued expansion of digital services. In this context, as the regulator of digital infrastructure, Arcep puts forward a number of proposals aimed at balancing sustainability, competitiveness and sovereignty in Europe's approach to data centre development.

Recommendation 7. Inform public and investment decisions: make transparency and data-driven regulation assets for the development of data centres in Europe

In its forward-looking work¹³⁶, Arcep highlighted the need to find ways of better incorporating social concerns into the deployment of new technologies and the digital infrastructure that supports them. It stressed the importance of giving end users access to reliable and transparent information so they can form their own views on a neutral and objective basis and contribute more effectively to public debate.

As discussed earlier in this chapter, transparency is essential to informed decision-making, whether by public authorities, businesses or individuals. It must therefore apply across the entire AI value chain, including at the upstream infrastructure level, namely with regard to data centre infrastructure.

Achieving this requires the monitoring of relevant data centre indicators, including resource consumption (particularly electricity and water), as well as the level of electrical capacity allocated to these infrastructures. In this respect, the revision of the Energy Efficiency Directive and Commission Delegated Regulation (EU) 2024/1364, which introduced a Europe-wide system for collecting environmental data on data centres, represent an important first step. The insights generated from this data will help support the European Union's work on future energy planning.

This type of mechanism therefore represents an initial step towards carrying out analyses that could support the development of future minimum performance standards for data centres. To unlock its full value as a public policy tool, however, its effective implementation must be ensured through highly complete and reliable data collection. According to the European Union, early feedback suggests that participation among data centres is still incomplete, which currently limits how representative and robust the resulting analyses can be. It is therefore important to support the continued rollout of this mechanism by improving reporting practices and strengthening oversight, with the aim of building a sufficiently comprehensive evidence base to inform future EU regulatory decisions effectively.

Recommendation 8. Strengthen European coordination between digital, energy and infrastructure policies to support the development of data centres

Energy planning is a key factor in the success of European policies aimed at decarbonisation, reindustrialisation and strengthening the continent's energy sovereignty. It is therefore essential to anticipate future investment needs in the electrical grid, as well as potential tensions over access to and use of electricity resources, rather than considering the development of AI – and the data centres that will underpin it in Europe – as a stand-alone issue. Indeed, the expansion of AI and data centres may compete with the electrification of other sectors, such as transport, which will likewise require significant investment in electricity infrastructure. Against this backdrop, it is important to adopt both a national and a European perspective to coordinate the deployment of data centres, an industry that

¹³⁶ Arcep, [What choice for digital technologies: viewing telco networks through the lens of digital practices](#) Future Networks, September 2025.

naturally tends to cluster in geographical hubs. Moreover, given the highly energy-intensive nature of data centres, it is also important to encourage their development in locations where the electricity mix is already largely decarbonised, such as France.

In this respect, work carried out by RTE as part of its Ten-Year Network Development Plan (SDDR), as well as studies on possible changes to the conditions for connecting to the electricity transmission network, can help inform public decision-making. At the European level, the forthcoming publication by the European Commission of the Strategic Roadmap for Digitalisation and Artificial Intelligence in the Energy Sector could take into account the relationship between data centre development and the development of the electricity system. Data centres should be viewed as integral components of that system, taking into consideration the services they could provide to the grid, particularly in terms of flexibility. Europe could also learn from Ireland's experience. Having faced earlier than many other countries the impact of rapid data centre growth on its electricity system – including making it more challenging to meet its 2030 decarbonisation targets – Ireland now imposes demanding conditions on the grid connection of new projects, such as requirements linked to the development of additional renewable energy capacity.

Recommendation 9. Encourage a coordinated territorial approach to data centre deployment

Beyond the strategic directions defined at the European level, the successful deployment of data centres also depends on close coordination with local authorities and communities, as it raises a number of local issues.

Water stress, the effective recovery and use of waste heat, and the land footprint of data centres are all closely linked to the social acceptance of these projects. Opposition to certain data centre developments is beginning to emerge at the local level, including in France. In this regard, the French High Council for Climate has called for projects to be monitored at the local level, noting that conflicts over resource use extend beyond electricity-related concerns. They also involve water resources and the availability of land, both of which are predominantly local issues.

In the context of Europe's strong ambitions for data centre development, as reflected in the AI Continent Action Plan, these more local considerations must be addressed from the outset when shaping local industrial development policies, in order to ensure both the success of these projects and their effective integration into the territories concerned. The objective should not be limited to achieving a high level of computing capacity on European soil; it should also encompass the ability of future projects to demonstrate their value to local communities through genuine sustainability, both economically and environmentally. This includes, for example, local employment creation and tax revenues, early planning to ensure the effective¹³⁷ recovery and use of waste heat, and forward-looking assessments of local water stress.

To this end, issues relating to water resources, land use and the recovery of waste heat from data centres should primarily be addressed through existing legal and regulatory frameworks. These already provide mechanisms for public participation and consultation¹³⁸. Reflecting growing interest in these issues and in the broader challenge of ensuring that data centres are deployed in a sustainable and

¹³⁷ If waste heat does not exceed a certain temperature, it is generally regarded as thermal waste, as it cannot be used for applications such as supplying a district heating network.

¹³⁸ For example, in France, these mechanisms include the public inquiry procedure, which forms part of the approval process for projects likely to have an impact on the environment (not limited to the data centre industry), as well as public consultation procedures conducted upstream by the National Commission for Public Debate (CNDP) for large-scale projects. By way of illustration, a recent data centre project in Châteauroux was the subject of a joint referral to the CNDP by the project developers, namely Google and the electricity transmission system operator (RTE).

coordinated manner, a number of collaborative initiatives have also begun to emerge, offering practical tools to support this approach, such as the Sustainable and Acceptable Data Centre Guide¹³⁹. Such innovative initiatives could also help inform the future operational implementation of the policy directions set out by the European Commission in the AI Continent Action Plan.

¹³⁹ <https://www.ville-demain.com/>

Appendices

1. Bibliography

Abitbol M., Aghion P., Antonin C., Barrage L. (2026), *Comment l'IA peut-elle contribuer aux objectifs d'efficacité énergétique dans l'industrie?*, Institut Veolia et Microsoft.

ADEME (2026), *Prospective d'évolution des consommations des data centers à court, moyen et long terme de 2024 à 2060.*

ADEME (2025a), *ACV de GPU pour l'intelligence artificielle.*

ADEME (2025b), *Évaluation environnementale des effets directs et indirects du numérique pour des cas d'usage, Étape 2 : numérisation de 10 cas d'usage.*

ADEME (2025c), *Évaluation de l'impact environnemental du numérique en France, an update of the ADEME-Arcep study.*

ADEME (2024a), *Évaluation environnementale des effets directs et indirects du numérique pour des cas d'usage, Étape 0 : état de l'art des méthodologies existantes, Étape 1 : analyse de 30 cas d'usages.*

ADEME (2024b), *Étude numérique et métaux.*

ADEME & Arcep (2022 & 2023), *Étude sur l'empreinte environnementale du numérique en France.*

AFNOR, Ministry of Ecological Transition and Territorial Cohesion (2024), *Référentiel général pour l'IA frugale.*

AIE (2025), *AI & Energy.*

AIE (2024), *Electricity 2024.*

Arcep (2026a), *Generative AI: Challenges for the Future of the Open Internet.*

Arcep (2026b), *Contribution à la consultation publique sur le réexamen du programme d'action pour la décennie numérique (PADN).*

Arcep (2026c), *Baromètre du numérique – édition 2026.*

Arcep (2026d), *Annual Achieving Digital Sustainability survey - 2026 (2024 data).*

Arcep (2025a), *Baromètre du numérique - édition 2025.*

Arcep (2025b), *AI and Electronic Communications Networks.*

Arcep (2025c), *Choose your tech: viewing electronic communications networks through the lens of digital practices.*

Arcep (2025d), *Contribution à la consultation publique sur les futures politiques publiques en matière d'IA et de cloud dans l'UE.*

Arcep (2025e), *Contribution à la consultation publique sur la feuille de route stratégique pour la numérisation et l'IA dans le domaine de l'énergie*.

Arcep (2025f), *Annual Achieving Digital Sustainability survey - 2025 (2023 data)*.

Arcep (2024), *Arcep's contribution to the call for contributions on competition in generative AI*.

Arcep & Arcom (2024), *General Policy Framework for the Ecodesign of Digital Services*.

Autorité de la Concurrence (2025), *Competition issues surrounding the energy and environmental impact of artificial intelligence (AI)*.

Beignon, A., Thibault, T., & Maudet, N. (2025), *Imposing AI: Deceptive design patterns against sustainability*. arXiv preprint arXiv:2508.08672.

Belcak, P., Heinrich, G., Diao, S., Fu, Y., Dong, X., Muralidharan, S., ... & Molchanov, P. (2025), *Small language models are the future of agentic AI*. arXiv preprint arXiv:2506.02153.

Bick, A., Blandin, A., & Deming, D. J. (2024), *The Rapid Adoption of Generative AI* (NBER Working Paper No. 32966). National Bureau of Economic Research.

Boavizta (2021), *Numérique et environnement: Comment évaluer l'empreinte de la fabrication d'un serveur, au-delà des émissions de gaz à effet de serre?*

Budenny, S.A., Lazarev, V.D., Zakharenko, N.N. et al. (2022), *Eco2AI: Carbon Emissions Tracking of Machine Learning Models as the First Step Towards Sustainable AI*. Dokl. Math. 106 (Suppl 1), S118–S128 (2022). <https://doi.org/10.1134/S1064562422060230>.

Carbone 4, Vasselin Z., Maquet P., Benedini C., Paulmier P. (2024), *Net Zero Initiative for IT: Understanding the role of digital solutions in the global net zero effort (NZI for IT)*.

Chen, Z., Lin, W., Xie, X., Hu, Y., Li, C., Tong, Q., ... & Li, S. (2024, December), *An empirical study on the power consumption of LLMs with different GPU platforms*. In 2024 IEEE International Conference on Big Data (BigData) (pp. 8640-8642). IEEE.

Coalition for Sustainable AI (2025), *Standardization for AI Environmental Sustainability*.

Commission de l'intelligence artificielle, Aghion P., Bouverot A. (2025), *IA: Notre ambition pour la France*.

Data for Good (2023), *Les grands défis de l'IA générative*.

Delavande, J., Pierrard, R., & Luccioni, S. (2026), *Understanding Efficiency: Quantization, Batching, and Serving Strategies in LLM Energy Use*. arXiv preprint arXiv:2601.22362.

Deloitte (2024), *Powering artificial intelligence*.

Desroches, C., Chauvin, M., Ladan, L., Vateau, C., Gosset, S., & Cordier, P. (2025), *Exploring the sustainable scaling of AI dilemma: A projective study of corporations' AI environmental impacts*. arXiv preprint arXiv:2501.14334.

De Vries, A. (2023). *The growing energy footprint of artificial intelligence*. Joule, 7(10), 2191-2194.

Diguet, C., Lopez, F., & Lefèvre, L. (2019), *L'impact spatial et énergétique des data centers sur les territoires* (Doctoral dissertation, ADEME, Directorate for Sustainable Cities and Territories).

European Green Digital Coalition (2024), *Net Carbon Impact Assessment Methodology for ICT Solutions*.

Falk, S., Ekchajzer, D., Pirson, T., Lees-Perasso, E., Wattiez, A., Biber-Freudenberger, L., ... & van Wynsberghe, A. (2025), *More than Carbon: Cradle-to-Grave environmental impacts of GenAI training on the Nvidia A100 GPU*. arXiv preprint arXiv:2509.00093.

Falk, S., Corrêa, N. K., Luccioni, S., Biber-Freudenberger, L., & van Wynsberghe, A. (2025), *From FLOPs to Footprints: The Resource Cost of Artificial Intelligence*. arXiv preprint arXiv:2512.04142.

France Stratégie (2025), *La demande en eau, Prospective territorialisée à l'horizon 2050*.

German Federal Ministry for the Environment, Climate Action, Nature Conservation and Nuclear Safety (2026), *Towards a Sustainable and Competitive AI Economy, Recommendations to the German Federal Environment Ministry*.

Google (2025), *Measuring the environmental impact of delivering AI at Google Scale*.

Green IT (2025), *Baromètre de l'Écoconception Digitale*.

Green IT (2025) *Impacts environnementaux et sanitaires de l'intelligence artificielle*.

Gupta, U., Kim, Y. G., Lee, S., Tse, J., Lee, H. H. S., Wei, G. Y., ... & Wu, C. J. (2021, February), *Chasing carbon: The elusive environmental footprint of computing*. In 2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA) (pp. 854-867). IEEE.

Hintemann, R., & Hinterholzer, S. (2019), *Energy consumption of data centers worldwide*. Business, Computer Science (ICT4S).

I Care (2025), *Intégrer la responsabilité environnementale dans le contexte de déploiement rapide de l'IA en entreprise*.

ITU (2025), *Measuring what matters: How to assess AI's environmental impact*.

Jegham, N., Abdelatti, M., Koh, C. Y., Elmoubarki, L., & Hendawi, A. (2025). *How hungry is AI? Benchmarking energy, water, and carbon footprint of LLM inference*. arXiv preprint arXiv:2505.09598.

Kaack, L.H., Donti, P.L., Strubell, E. et al., *Aligning artificial intelligence with climate change mitigation*. Nat. Clim. Chang. 12, 518–527 (2022).

Kachris C. (2025), *A Survey on Hardware Accelerators for Large Language Models*, Appl. Sci. 15 (2), 586.

Kim M., Y. T-A., Chung J. (2025), *Toward Sustainable Generative AI: A Scoping Review of Carbon Footprint and Environmental Impacts Across Training and Inference Stages*, Pre-Print, DOI: 10.48550/arXiv.2511.17179.

Lannelongue, L., Grealey, J., & Inouye, M. (2021), *Green algorithms: quantifying the carbon footprint of computation*. Advanced Science, 8(12), 2100707.

Ligozat, A. L., Lefevre, J., Bugeau, A., & Combaz, J. (2022), *Unraveling the hidden environmental impacts of AI solutions for environment life cycle assessment of AI solutions*. Sustainability, 14(9), 5172.

Li, P., Yang, J., Islam, M. A., & Ren, S. (2025), *Making AI less' thirsty'*. Communications of the ACM, 68(7), 54-61.

Limites Numériques (2025), *Comment les entreprises de la tech nous forcent à utiliser l'IA*, 2025.

Lopez, F., & Diguët, C. (2023), *Sous le feu numérique: Spatialités et énergies des data centers*.

Luccioni, A. S., Strubell, E., & Crawford, K. (2025, June), *From efficiency gains to rebound effects: The problem of Jevons' paradox in AI's polarized environmental debate*. In Proceedings of the 2025 ACM conference on fairness, accountability, and transparency (pp. 76-88).

Luccioni, S., Gamazaychikov, B., Da Costa, T. A., & Strubell, E. (2025), *Misinformation by omission: The need for more environmental transparency in AI*. arXiv preprint arXiv:2506.15572.

Luccioni, S., Jernite, Y., & Strubell, E. (2024, June), *Power hungry processing: Watts driving the cost of AI deployment?* In Proceedings of the 2024 ACM conference on fairness, accountability, and transparency (pp. 85-99).

Luccioni, A. S., Viguier, S., & Ligozat, A. L. (2023), *Estimating the carbon footprint of BLOOM, a 176b parameter language model*. Journal of Machine Learning Research, 24(253), 1-15.

Ministry of Ecological Transition and Territorial Cohesion (2025), *Projet de Stratégie nationale bas-carbone n°3*.

Mistral (2025), *Our contribution to a global environmental standard for AI*.

Moravec, V., Gavurova, B., & Kovac, V. (2025), *Environmental footprint of GenAI—Changing technological future or planet climate?* Journal of Innovation & Knowledge, 10(3), 100691.

Patel, A., Mahalingam, N., & Patel, R. (2025), *The Environmental Impact of AI Servers and Sustainable Solutions*. arXiv preprint arXiv:2601.06063.

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., ... & Dean, J. (2021), *Carbon emissions and large neural network training*. arXiv preprint arXiv:2104.10350.

Qiu, X., Parcollet, T., Fernandez-Marques, J., Gusmao, P. P., Gao, Y., Beutel, D. J., Topal, T., Mathur, A., & Lane, N. D. (2023), *A first look into the carbon footprint of federated learning*. Journal of Machine Learning Research, 24(129), 1-23.

Rincé S., and Banse A., Ecologits, *Evaluating the environmental impacts of generative AI*. Journal of Open Source Software, 10(111):7471, 2025.

Roussilhe G. (2022), *Les effets indirects de la numérisation*.

Roussilhe G. (2025), *La phase G: les GPU et les IA génératives comme nouvelle phase de l'histoire environnementale de la numérisation? (Part 1)*.

Roussilhe G. (2021), *Eau et puces électroniques : l'avenir climatique et industriel de Taiwan*.

Samsi S., Zhao D., McDonald J., Baolin Li, Michaleas A., Jones M., Bergeron W., Kepner J., Tiwari D., and Gadepally V. (2023), *From words to watts: Benchmarking the energy costs of large language model inference*. In 2023 IEEE High Performance Extreme Computing Conference (HPEC), pages 1–9. IEEE, 2023.

Schneider, I., Xu, H., Benecke, S., Patterson, D., Huang, K., Ranganathan, P., & Elsworth, C. (2025), *Life-cycle emissions of AI hardware: A cradle-to-grave approach and generational trends*. arXiv preprint arXiv:2502.01671.

Sopra Steria (2025), *IA & environnement : sortir du brouillard informationnel*.

Stern, N., Romani, M., Pierfederici, R., Braun, M., Barraclough, D., Lingeswaran, S., ... & Niemann, N. (2025), *Green and intelligent: the role of AI in the climate transition*. npj Climate Action, 4(1), 56.

Strubell, E., Ganesh, A., & McCallum, A. (2019, July), *Energy and policy considerations for deep learning in NLP*. In Proceedings of the 57th annual meeting of the Association for Computational Linguistics (pp. 3645-3650).

The Shift Project (2025), *Intelligence artificielle, données, calculs. Quelles infrastructures dans un monde décarboné?*

Unesco (2025), *Smarter, Smaller, Stronger: Resource-Efficient Generative AI & the Future of Digital Transformation*.

Vanderbauwhede, W. (2024), *Estimating the Increase in Emissions caused by AI-augmented Search*. arXiv preprint arXiv:2407.16894.

Varoquaux, G., Luccioni, S., & Whittaker, M. (2025, June), *Hype, Sustainability, and the Price of the Bigger-is-Better Paradigm in AI*. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (pp. 61-75).

World Bank, International Telecommunication Union (2025), *Measuring National ICT Sector Environmental Impact, Arcep Case Study – France*.

You J., *How much energy does ChatGPT use?* (2025). URL <https://epoch.ai/gradient-updates/how-much-energy-does-chatgpt-use>.

Zhang, M., Carbajales-Dale, M., Ma, X., Guo, L., & Fan, C. (2025), *Cleaner Grid or Smarter Cooling? Environmental Impact Trade-offs of a Data Center Using the Life Cycle Assessment Method*. Cleaner Energy Systems, 100223.

2. Interviews

The following were interviewed for this report:

- Théo Alves Da Costa (Data for Good)
- Cyriaque Dubois (Mistral AI)
- Juliette Fropier (Ecolab)
- Marie-Liane Lekpeli, Véronique Bouvelle, Emma Le Boulicaut (French Ministry of the Economy and Finance – Directorate General for Enterprise)
- François Philippe (RTE)
- Samuel Rincé (Ecologits)
- Gaël Varoquaux (Inria)
- Eloïse Villani (French Ministry of the Economy and Finance – Directorate General of the Treasury)

These interviews contributed to Arcep's reflections; however, the report does not necessarily reflect the views of the individuals consulted.

Arcep would also like to thank the experts who were consulted on other occasions:

- Two internal academic seminars were held on this topic, with contributions from Anne-Laure Ligozat (LISN) and Loïc Lannelongue (University of Cambridge).
- Inria was consulted during the drafting of the General Policy Framework for the Ecodesign of Digital Services (RGESN), in particular Denis Trystram and Laurent Lefèvre (Inria).
- During the preparation of the report "Choose your tech: viewing electronic communications networks through the lens of digital practices", consultations were held with industry stakeholders (Ericsson, Télécoop, Google, Meta and Orange), academics (Christian Licoppe, Clément Marquet, Jean-Samuel Beuscart, Nicolas Jullien, Soazig Lalancette and Laura Draetta), local authorities (Nantes Métropole and Rennes Métropole) and associations of local authorities (the Rural Mayors' Association of France and Les Interconnectés).

The Authority also wishes to thank the experts who took part in bilateral and multilateral discussions, including Sasha Luccioni (Hugging Face), Aurélie Bugeau (LaBRI) and Marlène de Bank (The Shift Project). It also extends its thanks to partners with whom it has engaged on these issues over time, including ADEME.

Arcep thanks the students (Maela Guillaume-Le Gall, Rafaël Mourouvin and Julian Rojas) and professors (Franck Malherbet and Pierre Boyer) from the Data and Economics for Public Policy (DEPP) Master's programme of École Polytechnique, ENSAE and Télécom Paris for their work on the environmental impacts of AI.

Finally, the Authority would like to express its particular gratitude to PEReN for its collaboration on the project examining the energy consumption of AI models during the inference phase.

3. Glossary

Agentic AI: Field of techniques relating to AI agents (see below).

AI Act: European Regulation establishing a harmonised framework for the development, marketing and use of artificial intelligence systems. It adopts a risk-based approach and includes obligations relating in particular to safety, transparency, data governance, technical documentation and human oversight. The Regulation also includes specific provisions for general-purpose AI models (GPAI models), notably in the field of generative AI¹⁴⁰.

AI agent: Software system that employs artificial intelligence techniques (in particular LLMs) to perceive its environment, make decisions and perform tasks autonomously on behalf of human users or organisations.

AI model: An algorithm whose operation is determined by a set of attributes, and which is designed to perform various tasks, including prediction, classification, inference or generation, depending on the context (based on the CNIL glossary definition).

AI system: According to ISO/IEC 22989, a system designed to generate outputs such as content, predictions, recommendations or decisions for a given set of human-defined objectives.

Artificial intelligence: Field of research that seeks to enable artificial systems, most often computer systems, to perform tasks associated with human or natural intelligence. Machine learning is currently its principal approach.

Chatbot: Computer system designed to interact with a human user in natural language. It uses artificial intelligence technologies, such as natural language processing and machine learning, to understand prompts, generate relevant responses and simulate a fluid conversation. Chatbots may be deployed as conversational agents, virtual assistants or voice interfaces.

Cloud service: Article 2(8) of the Data Act defines a cloud service as "*a digital service [...] that enables network access everywhere and on demand to a shared pool of configurable, scalable and elastic computing resources of a centralised, distributed or highly distributed nature, which can be rapidly provisioned and released with minimal management effort*".

Content and application providers: Actors involved in the provision of content (web pages, blogs, videos) and/or applications (search engines, mobile applications, route-planning services) on the internet.

Deep learning: Subfield of machine learning (see below) that seeks to enable machines to make predictions using techniques based on neural networks (see below).

Digital platform: A wide variety of online services and actors providing intermediation services, such as online marketplaces, community platforms and app stores. These actors may vary considerably in size, revenues and business models and operate across numerous sectors of the economy.

Expert system: According to ISO/IEC 22989, an AI system that accumulates, combines or encapsulates knowledge provided by one or more human experts in a specific field in order to provide solutions to problems.

Fine-tuning: Technique used to specialise a pre-trained AI model for a specific task. This generally involves training the model as a whole, or only certain layers of a neural network, for a limited number of iterations on a dataset corresponding to the target task.

¹⁴⁰ CNIL, 2024. [Entrée en vigueur du règlement européen sur l'IA : les premières questions-réponses de la CNIL.](#)

Foundation model: AI model trained on large volumes of one or more types of data in order to be adapted subsequently to a wide range of tasks. In the field of natural language processing, LLMs are an example of foundation models.

General-Purpose AI Model (GPAI): Article 3(63) of the AI Act defines this as "*an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market*".

Generative AI: Field of techniques that use artificial intelligence models to generate new content such as text, computer code, images or music.

Generative AI model: AI model characterised by its ability to produce, in response to a request, bespoke outputs such as text, images or video.

Generative AI service: Digital service based on a generative AI model.

Hyperscaler: Large cloud service provider capable of delivering computing and storage services at very large scale.

Inference: Phase during which an artificial intelligence model is used to make predictions, as opposed to the training phase. A generative AI service relies on one or more LLMs operating in the inference phase.

ISO (International Organization for Standardization): Independent, non-governmental international standards organisation. ISO standards ensure the reliability, safety and quality of products and services¹⁴¹.

Large Language Model (LLM): Natural language processing AI model containing a large number of parameters.

Machine learning: Field of research relating to mathematical and statistical techniques that enable machines to learn to make generalisable predictions from data (known as training data).

Neural network: An organised set of interconnected artificial neurons that learn, through optimisation techniques derived from machine learning, to detect patterns in training data in order to solve problems in complex fields such as computer vision and natural language processing (based on the CNIL glossary definition).

Prompt: Text-based request submitted by a user to a model.

Token: According to ISO/IEC 22989/Amd.1, a unit of content that an AI model considers to have semantic meaning.

Traffic: The volume of data transmitted at a given moment between two machines, for example across a data interconnection. When traffic exceeds the capacity of the link, the connection becomes saturated.

Training: The process by which a system uses data and algorithmic models to perform calculations in order to provide functionalities, improve performance or acquire the ability to perform certain tasks autonomously.

¹⁴¹ ISO. [ISO website](#)

Training data: Dataset (e.g. text, audio, images) used to train a machine learning or artificial intelligence model, that is, to adjust its parameters so that it learns to perform a given task.

Virtual Assistant: Article 2(12) of the Digital Markets Act defines a virtual assistant as "*software that can process requests, tasks or questions, including those based on audio, visual, written, gesture or movement inputs, and that, on the basis of those requests, tasks or questions, provides access to other services or controls connected physical devices*".

Vocal assistant: Virtual assistant (see above) that operates primarily through audio inputs.

4. List of acronyms

ADEME: French Agency for Ecological Transition (formerly the *Agence de l'Environnement et de la Maîtrise de l'Energie*)

ADSL: Asymmetric Digital Subscriber Line

AFNOR: *Association Française de Normalisation* (French Standardisation Association)

AI: Artificial Intelligence

A-LCA: Attributional Life Cycle Assessment

API: Application Programming Interface

Arcep: *Autorité de régulation des communications électroniques, des postes et de la distribution de la presse* (French Regulatory Authority for Electronic Communications, Postal Services and Press Distribution)

ARN: *Autorité de Régulation du Numérique* (French Regulatory Authority for Digital Technologies)

BEGES: *Bilan d'Emissions de Gaz à Effet de Serre* (French Greenhouse Gas Emissions Assessment)

BLOOM: BigScience Large Open-science Open-access Multilingual Language Mode

CEEMS: Compute Energy & Emissions Monitoring Stack

CEN/CENELEC: European Committee for Standardization / European Committee for Electrotechnical Standardization

CGDD: *Commissariat Général au Développement Durable* (French Commission for Sustainable Development)

C-LCA: Consequential Life Cycle Assessment

CNRS: *Centre National de la Recherche Scientifique* (French National Centre for Scientific Research)

CPU: Central Processing Unit

CSIA: *Comité de Surveillance des Investissements d'Avenir* (French Oversight Committee on Future Investments)

CSRD: Corporate Sustainability Reporting Directive

DFA: Digital Fairness Act

DGE: *Direction Générale des Entreprises* (French Directorate General for Enterprise)

DLR: Dynamic Line Rating

DNA: Digital Networks Act

ETSI: European Telecommunications Standards Institute

EU: European Union

GHG: Greenhouse Gases

GPT: Generative Pre-trained Transformer

GPU: Graphics Processing Unit

IEEE : Institute of Electrical and Electronics Engineers

Inria: *Institut national de recherche en sciences et technologies du numérique* (French Institute for Research in Computer Science and Automation)

ISO: International Organization for Standardization

ITU: International Telecommunication Union

LCA: Life Cycle Assessment

LLM: Large Language Model

LoRA: Low-Rank Adaptation

MoE: Mixture of Experts

PEReN: *Pôle d'Expertise de la Régulation Numérique* (French Centre of Expertise for Digital Platform Regulation)

PUE: Power Usage Effectiveness

RAG: Retrieval Augmented Generation

REEN (act): *Réduire l'Empreinte Environnementale du Numérique* (Reduce the Environmental Footprint of Digital Technology)

RGESN: *Référentiel Général de l'Ecoconception des Services Numériques* (French General Policy Framework for the Ecodesign of Digital Services)

RTE: *Réseau de Transport d'Electricité* (French Electricity Transmission System Operator)

SLM: Small Language Model

TPU: Tensor Processing Unit

UN: United Nations

UNESCO: United Nations Educational, Scientific and Cultural Organization

WUE: Water Usage Effectiveness

5. Orders of magnitude and equivalences relating to energy consumption and greenhouse gas emissions

Scope	Greenhouse gas emissions	Source
European Union – 2023	3,222,000,000 tCO ₂ e	EU, Edgar , 2025
France – 2024	369,000,000 tCO ₂ e	MTE, SDES , 2025
ICT sector – France – 2022	29,500,000 tCO ₂ e	ADEME , 2025c
User devices – France – 2022	14,700,000 tCO ₂ e	ADEME , 2025c
Data centres – France (including use of data centres located abroad) – 2022	13,500,000 tCO ₂ e	ADEME , 2025c
Networks – France – 2022	1,200,000 tCO ₂ e	ADEME , 2025c
Data centres – France (Arcep data collection scope) – 2023	137,000 tCO ₂ e	Arcep , 2025
Training of the Large 2 model (18 months)	20,400 tCO ₂ e	Mistral, 2025
Training of the LLaMA 3 405B model	9,000 tCO ₂ e	Meta, 2025
Training of the BLOOM model (118 days) – data centres	124 tCO ₂ e	Luccioni <i>et al.</i> , 2023
Average annual carbon footprint of a French resident – 2023	8.5 tCO ₂ e	Citepa , ABC , 2023
Manufacture of an NVIDIA A100 SXM 40 GB GPU	0.150 tCO ₂ e	Falk et al. , 2025

Table 3 – Orders of magnitude and equivalences relating to greenhouse gas emissions

Scope	Final energy consumption	Source
European Union – Electricity – 2023	2 326 TWh	EU, Eurostat , 2025
France – Electricity – 2024	449 TWh	RTE , 2025
Data centres – Worldwide – 2024	415 TWh	AIE , 2025
Data centres – France – Arcep data collection scope – 2023	2.4 TWh	Arcep , 2025
City of approximately 100,000 inhabitants – Electricity	200,000 MWh	CRE , 2025
Training of GPT-3 – Servers	1 287 MWh	Luccioni <i>et al.</i> , 2023
Training of the BLOOM model – Servers	433 MWh	Luccioni <i>et al.</i> , 2023
Average household – France – Estimate produced in 2025 for 2023–2024 data	4.3 MWh	CRE , 2025

Table 4 - Orders of magnitude and equivalences relating to electricity consumption

6. Comparison of the training of the GPT-3 and BLOOM models based on the literature review

The table below compares data characterising the resources required to train the **GPT-3** and **BLOOM** AI models (Patterson et al., 2022; Luccioni et al., 2022; Data for Good, 2023).

Breakdown	GPT-3 (OpenAI)	BLOOM (BigScience)
Number of parameters	175 billion	176 billion
Volume of training data	45 TB	1.6 TB
Architecture	Transformer (decoder-based)	Transformer (decoder-based)
Training duration	15 days	118 days
Resources used	10,000 GPUs (NVIDIA V100)	384 GPUs (NVIDIA A100)
Electricity consumption of training servers	1,287 MWh	433 MWh
Carbon intensity of the electricity consumed for training	427 gCO ₂ e/kWh	57 gCO ₂ e/kWh
Carbon footprint associated with servers' electricity consumption	552 tCO₂e	30 tCO₂e
Electricity consumption of servers and ancillary equipment (cooling, switches, routers, etc.)	N/A	690 MWh
Carbon footprint associated with the data centres used (excluding equipment manufacturing)	N/A	39.3 tCO₂e
Carbon footprint associated with the data centres used (including equipment manufacturing)	N/A	50.5 tCO₂e
Carbon footprint associated with the data centres used (including equipment manufacturing) and including testing, evaluation and debugging	N/A (estimated at approximately 2,200 tCO₂e based on the ratio observed for BLOOM)	123.8 tCO₂e

7. Performance of AI models according to training time or memory footprint for different tasks

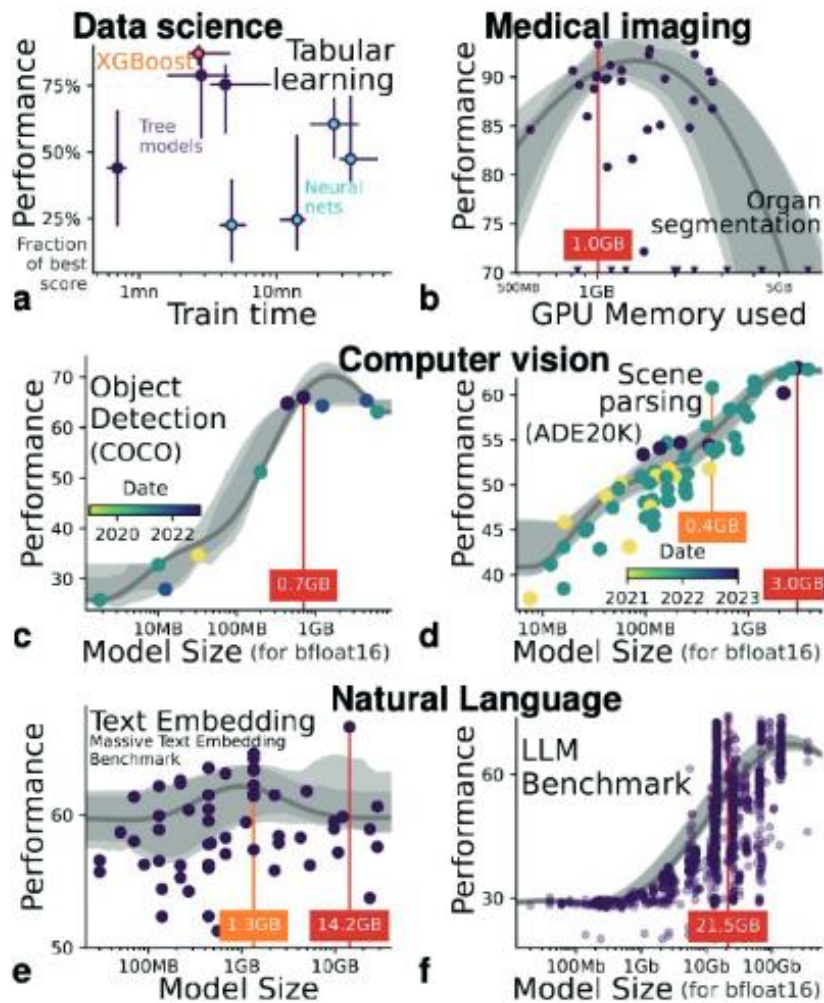


Figure 16: Performance of AI models according to training time or memory footprint for different tasks¹⁴²

¹⁴² Source: Varoquaux et al., 2025. Tasks covered: tabular learning, medical image segmentation, computer vision object detection, scene parsing, text embedding and text comprehension.

8. Mapping of standardisation initiatives for methods to assess the environmental impact of AI

The Coalition for Sustainable AI, launched as part of the AI Action Summit in February 2025, initiated a common roadmap across the various standardisation initiatives, bringing together experts from the ISO, ITU and IEEE, in partnership with the OECD and UNESCO.

This roadmap has helped initiate coordination among the various standardisation bodies and optimise the dedicated resources.

The diagram below presents the existing standardisation initiatives and their links with other standards currently in force relating to the broader field of assessing the environmental impact of digital technologies and AI systems.

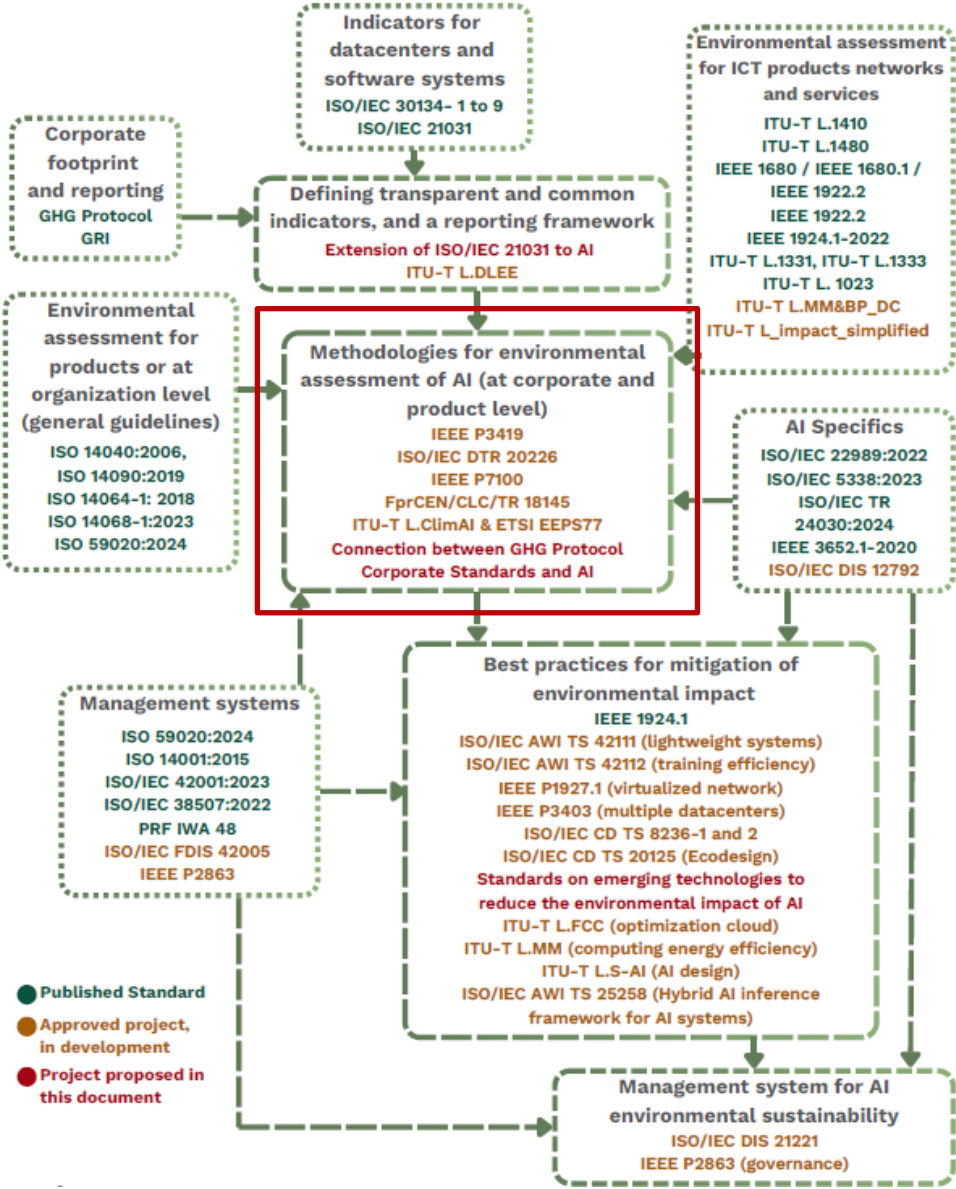


Figure 17 - Published and ongoing standards for assessing the environmental impact of AI systems. Source: Coalition for Sustainable AI, February 2026.

This document was produced by Arcep

Olivier Corolleur, Director General
Rodolphe Le Ruyet, Adviser to the Chair

ECONOMIC, MARKET AND DIGITAL AFFAIRS DEPARTMENT

Anne Yvrande-Billon, Director
Blaise Soury-Lavergne, Deputy Director

Economic Analysis and Digital Intelligence Unit

Marion Panfilj, Head of Unit
Chiara Caccinelli, Deputy Head of Unit
Charles Joudon-Watteau, Policy Officer
Tom Nico, Policy Officer

INTERNET, DATA, PRESS, POSTAL AFFAIRS AND USERS DEPARTMENT

Olivier Delclos, Director

Open Internet Unit

Sandrine Elmi Hersi, Head of Unit
Oriane Piquer-Louis, Policy Officer
Vivien Guéant, Expert

Data Services and Cloud Unit

Léo Quentin, Head of Unit
Philéas Samir, Policy Officer

MOBILE AND INNOVATION DEPARTMENT

Franck TARRIER, Director
Ahmed Haddad, Technical Adviser

EUROPE AND INTERNATIONAL DEPARTMENT

Anne Lenfant, Director

Europe Unit

Emmanuel Leroux, Head of Unit
Malo Le Cointe, Policy Officer

COMMUNICATION AND PARTNERSHIPS DEPARTMENT

Clémentine Beaumont, Director

Victor Schmitt, Policy Officer

LEGAL AFFAIRS DEPARTMENT

Élisabeth Suel, Director

Mobile Markets and Scarce Resources Unit

Léa Ployaert, Head of Unit

Annabel Gandar, Deputy Head of Unit

Many thanks to the PEReN team for its contribution!

The Digital Regulation Expertise Centre (PEReN) supports government bodies and independent authorities responsible for regulating digital platforms and artificial intelligence. As a technical centre of expertise, it brings together leading capabilities in data science, algorithm auditing and analysis, AI evaluation and software development to deliver digital tools, studies and expert advice. It also carries out exploratory and scientific public research projects. For more information: www.peren.gouv.fr

Publication

Arcep

14, rue Gerty-Archimède

75012 Paris

Communication and Partnerships Department: comm@arcep.fr

12 June 2026

ISSN No. 2258-3106



This content is licensed under the [Creative Commons Attribution–ShareAlike 4.0 International Licence \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/).