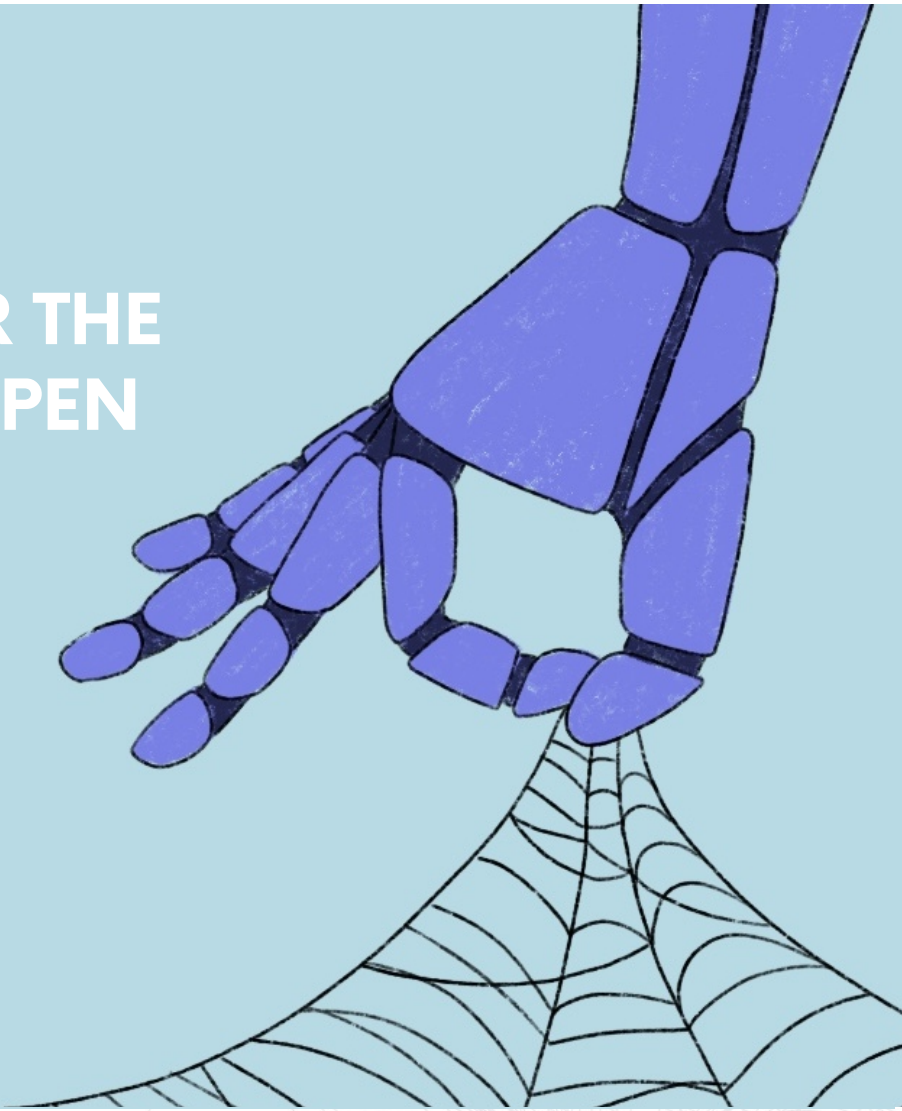
arcep
autorité de régulation
des communications électroniques,
des postes et de la distribution de la presse

RÉPUBLIQUE FRANÇAISE

GENERATIVE AI: CHALLENGES FOR THE FUTURE OF THE OPEN INTERNET

The impact of generative AI services on freedom of choice, content diversity and innovation on the internet



January 2026

In collaboration with the Centre of Expertise for Digital Platform Regulation- PEReN

GOUVERNEMENT
Liberté
Égalité
Fraternité

ISSN n° 2258-3106

Foreword

By Laure de La Raudière
Chairwoman of Arcep

What will become of the internet we know today, where everyone is free to access and share the content, applications and services of their choice, if generative AI becomes our main gateway to information and online innovation? What risks does this pose to the guarantee of an “open internet”, whose principles were enshrined ten years ago in a European regulation?

Eighteen months ago, I raised the alarm on this issue in an opinion piece published in *Le Monde*.

Since then, the rapid spread of generative AI services has not only been confirmed, it has accelerated. Chatbots have also evolved from conversational assistants to fully-fledged digital agents, some of which are capable of interacting with other services (online shopping, streaming platforms, social networks, etc.), carrying out "digital actions" on our behalf, and offering us increasingly personalised experiences. Agentic AI is gradually establishing itself as an essential intermediary service between users and internet content, which could structure, filter and prioritise our access to information and online services.

To revisit the metaphor used eighteen months ago, imagine once again going to your newsagent every day to buy your favourite newspaper. This time, not only does he summarise the news for you, but he also offers to order your train tickets, deliver your meal tray and summarise the latest news about your loved ones by logging in your social networks. How convenient, you might say! Your newsagent, who now knows how to do just about everything, searches for and selects content, services and experiences for you to discover. But how does they decide what to show you? What biases influence their recommendations? How transparent are they about their sources or the contracts they may have signed with companies? And above all, how much freedom do you have to deviate from this predefined path? This is, increasingly, the role played by agentic generative AI.

These major innovations, which are promising for our societies and economies, challenge us to consider one of the founding principles of the internet: its openness, which guarantees freedom of choice and innovation. However, this openness cannot be seen or heard; and the internet will not close abruptly; rather it risks closing quietly and gradually, under the guise of convenience and personalisation, until the day the very notion of “choice” has lost its meaning.

From its technical and economic regulatory standpoint and drawing on PEReN’s expertise[1], Arcep aims through this report to shed light on this new reality and to anticipate evolving patterns of use. We also propose a number of recommendations for public debate. We believe it is essential to reaffirm the principle of an open internet, to strengthen the transparency and traceability of sources used by generative AI, to support auditable and assessable models, and to promote interoperability and open protocols. This will be necessary to prevent AI services from gradually locking in users and to guarantee fair access to key online services.

These recommendations aim to preserve the diversity of the internet, where every user retains their freedom of choice, where innovation remains possible for all, and where the diversity of digital content and experiences is preserved. The openness of the internet is not limited to the telecoms network layer: it must extend to all gateways to the internet, including those that AI could one day lock up.

[1] Center of Expertise for Digital Platform Regulation, a public service with national jurisdiction that supports public authorities in regulating digital platforms and AI.

Table des matières

Summary

1. Generative AI services: new gateways to the internet

Generative AI services represent a major and promising innovation for our societies and economies. These tools notably enable both individuals and professionals to create content, access information and receive assistance in new ways.

Through its rapid spread into everyday digital uses and tools, generative AI is set to play a decisive role in how we access the diversity of internet content and services. This development is revolutionising the digital practices of users, who are no longer necessarily encouraged to navigate from link to link, but are led to focus their interactions on the interface of an AI agent, which selects, reformulates and structures information, accesses certain third-party services on their behalf and offers a single response in natural language, limiting users' direct access to the original sources and services. Generative AI services thus constitute a new gateway to the internet, similar to the services provided by internet service providers (ISPs) and major digital platforms such as search engines, social networks, browsers and e-commerce platforms.

The provision of generative AI services relies on the collection and processing of massive volumes of data for training purposes, which creates new relationships and challenges between stakeholders. Furthermore, due to the inherent limitations of learning techniques, the responses of generative AI may contain biases or errors.

As new gateways to the internet, generative AI services are likely to challenge some of the network's historical foundations, particularly the principle of the open internet. This principle stems from the original design of the internet as a decentralised "network of networks", based on open protocols, allowing the free exchange of information, knowledge and data without prior authorisation, and promoting innovation.

2. The open internet: a founding principle for innovation and online freedoms

The principle of an open internet has profoundly influenced the design of internet protocols and technical architectures. Within the European Union, the "Open Internet" Regulation[2] adopted in 2015 imposes net neutrality obligations on internet service providers, i.e. equal treatment of traffic, regardless of its destination, sender or content; and defines the principle of open internet as the right of end users to access and distribute the content, applications and services of their choice. It aims to prevent technical intermediaries from imposing discriminatory conditions on the circulation of internet content and services. The challenge is twofold: on the one hand, to promote innovation in digital markets and, on the other, to create a foundation for the exercise of fundamental freedoms on the internet, in particular freedom of expression, freedom of enterprise and freedom of information.

Since the first debates on net neutrality, the digital ecosystem has undergone profound changes. Beyond internet service providers, large digital platforms such as search engines, social networks, operating systems and app stores have become essential intermediaries for accessing online content and services. While ISPs were the physical gateways to the internet, these players have gradually taken on the role of software gateways, capable of structuring, filtering or restricting users' choices.

[2] Regulation (EU) 2015/2120 of the European Parliament and of the Council of 25 November 2015 laying down measures concerning open internet access.

In this context, it has become necessary to look beyond the neutrality of the infrastructure operated by ISPs in order to preserve the ambition of an open internet that guarantees users' freedom of choice and capacity for innovation. The European Union has provided initial responses to these challenges, notably through the Digital Markets Act, which aims to prevent practices that could restrict market contestability and users' effective freedom of choice.

3. The impact of generative AI on the open internet and the risks of concentration

Generative AI services are part of this evolving dynamic and are likely to extend, or even amplify, certain risks posed by large platforms. Their specific nature also raises new issues that require particular attention with regard to the openness of the internet.

Freedom to access the content and services of one's choice

Generative AI services are redefining how we access content and services. Users are shifting from exploring the web, navigating from link to link, to reading a summary produced by an AI system. They no longer directly browse the range of content and services available on the internet, but consult a summary built from sources selected by AI.

While the first generations of generative AI services were characterised by a lack of source display, significant progress has since been made since, with the deployment of mechanisms to explain the reasoning behind the responses and indicate some of the sources used. However, generative AI services generally present a limited number of sources per response, and their summarised response in natural language tends to discourage users from consulting the original sites and exploring the web. Furthermore, the criteria used by AI to select and prioritise sources remain largely opaque.

The recent development of so-called agentic AI, which allows AI services to interact with other digital applications and services, could lead internet users to delegate the choice of applications or services used to AI agents. The selection made could then be guided by the preferences or partnerships, particularly commercial ones, established by the AI service provider, to the detriment of free choice and open innovation on the internet.

Ability to innovate and share content and services of one's choice

Generative AI services are profoundly transforming the way content and services are shared and promoted for publishers, providers and internet users. They call into question the visibility of content on the internet and may lead to a decline in traffic to source sites. Furthermore, the relevant players concerned have limited leverage to influence how generative AI services index and highlight content. Search engine optimisation (SEO) practices, which are currently well established for search engines, must evolve towards new and still poorly documented practices aimed at optimising content visibility through generative AI tools. These changes raise major challenges in terms of discoverability, business models and the sustainability of traditional players in the creation and provision of online content and services.

Ultimately, this reduced discoverability, combined with competition from summarised content that is easy to produce in large quantities, could discourage the production of human-generated content, which is essential to the diversity of the digital ecosystem. However, the scenario of an artificial internet, characterised by predominantly summarised content, would not only undermine online diversity, but could also weaken the development of future generative AI services themselves, since real-world data is necessary for their training and the quality of their outputs.

Openness and contestability of digital markets

While business models are still being structured – which calls for careful analysis – there are risks of concentration at different stages of the value chain. These risks are exacerbated by the scale of investment required, privileged access to massive volumes of data, and the vertical integration and ecosystem strategies implemented by certain players. These conditions may lead to increased dependence by users, content and

service providers on a limited number of generative AI services, which are set to play a central role in accessing internet content and services. If such concentration dynamics were to be confirmed, the choices made by "dominant" generative AI service providers, particularly in terms of interface design, source selection and conditions of access to third-party services, would become all the more decisive for internet users' freedom of choice and the capacity for innovation on the internet.

4. Combining an open internet with the spread of generative AI

In light of this assessment, Arcep proposes six recommendations aimed at reconciling the development of generative AI services with the preservation of an open internet. The goal is to support the innovations enabled by generative AI while ensuring that internet users are able to freely exercise their choices regarding the content they view or share, the services they use and the innovations they develop. It is also a matter of preventing any form of unjustified market lock-in by intermediaries that have become indispensable.

> **Axis 1: Reaffirming the principles of the open internet in the age of generative AI**

The principle of an open internet in the age of rapid development and adoption of generative AI services must be reaffirmed in the context of digital and AI regulation, particularly at European level. The open internet must also be a focus of discussions and negotiations within multilateral forums on the development and governance of generative AI.

Furthermore, analysis of the effects of generative AI services on the openness of the internet must be continued: Arcep calls for further research on this topic.

> **Axis 2: Develop open protocols for interconnections between generative AI service providers and content and application providers**

The development of an open generative AI ecosystem depends in particular on the establishment of technical conditions and technological building blocks that enable players to interconnect effectively, in line with the technical history of the internet. The aim is to support the implementation and adoption of open and interoperable protocols that help ensure more fluid, transparent and balanced relationships between AI services and online content and service providers, including in support of contractual agreements where appropriate. Such protocols would benefit from being sufficiently granular to take into account the diversity of interactions – training, indexing, agentification – that are likely to be contractualised between generative AI service providers and content and application providers. The protocols could work to support the vertical interoperability of generative AI services, within a multi-stakeholder governance framework conducive to their adoption on an international scale.

> **Axis 3: Creating fair conditions for access, use and promotion of content and services by generative AI**

The development of an open and dynamic generative AI ecosystem requires striking a balance between, on the one hand, the fair valuation of content and services used by generative AI services and, on the other hand, the preservation of access conditions that are conducive to innovation and competition, particularly for emerging players. In this regard, access to protected content, particularly when used for training or generation purposes, must be subject to appropriate regulation. It would also be advisable to encourage the development of technical mechanisms that facilitate the processing of content, such as the initial experiments with microtransaction systems conducted by certain technical intermediaries, or the use of trusted third parties responsible for facilitating the management of rights and value flows.

With regard to the relationship between press publishers and generative AI services, which raises specific issues of pluralism, existing mechanisms for the distribution of political and general news publications by digital newsstands, which require the latter to comply with press publishers' requests under reasonable and non-discriminatory conditions, could inspire the implementation of mechanisms adapted to generative AI services.

Beyond protected data, the provision of shared resources, including data spaces that meet the needs of AI developers, should be encouraged.

> **Axis 4: Mobilising existing regulatory tools to ensure the openness of generative AI services**

The existing European regulatory framework, including the Digital Markets Act, the Digital Services Act and the Data Act, provides a set of tools that can help to protect the openness of generative AI services. These instruments are particularly relevant given that generative AI services are largely implemented by players already covered by these legal frameworks. They should be mobilised to address some of the identified issues and mitigate certain risks associated with AI services.

> **Axis 5: Supporting the development of more transparent and assessable generative AI services**

While significant progress has been made by industry players, findings highlight the persistence of challenges in terms of the reliability, transparency and traceability of generative AI services. These issues call for continued efforts, in particular by improving the monitoring of model performance and, where relevant, supporting the development of smaller, more frugal models that can be more easily assessed.

> **Axis 6: Empowering internet users to define and curb their use of generative AI**

The transformations driven by generative AI services underscore the importance of providing users with reliable and comparable information to support informed choices. Arcep encourages the availability of configuration options that maintain freedom of choice in terms of sources and services. More broadly, the Authority calls for continued efforts in AI training and literacy, as well as for the promotion of mechanisms that take users' interests into account in the development of generative AI services.

Introduction

At the end of November 2022, Open AI publicly launched its ChatGPT conversational agent. **A pioneer of a new generation of tools known as generative artificial intelligence (hereinafter "generative AI"), it can automatically produce content from simple natural language instructions, called "*prompts*",** initially in the form of plain text, but more recently also in the form of images, sound and video in a multimodal format.

Thanks to these assets and their ease of use, the general public adopted generative AI services swiftly. ChatGPT reached 100 million monthly active users in just two months, growing faster than any other consumer application to date[3] .

A major and promising innovation for our societies and economies, generative AI offers users, whether individuals or professionals, new creative and support abilities. It is gradually becoming a central component of the online user experience. To support the implementation of these technologies, massive investments have been made by both private and public investors. For example, France has earmarked €1.1 billion for the 2023-2025 period as part of its national AI strategy[4] , and €109 billion in private investment in AI infrastructure was announced at the AI Summit in February 2025 in Paris[5] . The European Commission has announced an investment of €200 billion in AI, including a €20 billion European fund to finance AI "gigafactories" as part of the InvestAI programme.[6]

Through its integration into everyday digital uses and tools, generative AI is set to play a decisive role in how we access the diversity of internet content and services. This development is disrupting digital practices, as users are no longer necessarily encouraged to navigate from link to link, but are led to focus their interactions on the interface of an AI agent, which selects, reformulates and structures information, accesses certain third-party services on their behalf and offers a single response, limiting users' direct access to the original sources and services. **Thus, emerging generative AI services are a new gateway to the internet, similar to the services provided by internet service providers (ISPs) and large digital platforms**.

⁂

The emergence of generative AI, which produces unique responses to user queries, is likely to contradict certain founding values of the internet, **foremost among which are its openness**, understood as the freedom for all internet users to access, publish and contribute online, **and its decentralised nature**, which has historically been based on a plurality of access points, players and sources.

Keeping the internet open means protecting the ability of all users to access the information, applications and services of their choice, to share them and to offer new ones. In particular**,** it aims to prevent a digital player occupying an intermediary position from using this role to exclude, restrict or disadvantage competitors, particularly in adjacent markets such as the production of content,

[3] UBS, 2024. *Artificial intelligence: Sizing and seizing the investment opportunity*.

[4] Cour des comptes, 2025. *The national strategy for artificial intelligence*.

[5] Elysée, 2025. *Summary of the Artificial Intelligence Summit*.

[6] European Commission, 2025. Press release: "*EU launches InvestAI initiative to mobilise €200 billion in investment in artificial intelligence*".

services or applications. **The aim is to create a level playing field, ensuring effective competition and an environment conducive to innovation**.

Beyond this economic dimension, **the openness of the internet is also essential to the exercise of fundamental freedoms online**, including freedom of expression, the right to information and freedom of enterprise. By guaranteeing equal access to content and the ability for everyone to create, publish, share and innovate without prior authorisation, the open internet contributes both to economic growth and to democratic expression in the digital space.

In France, since 2016, **Arcep has been responsible for ensuring compliance with the Open Internet Regulation** adopted in 2015, which requires ISPs to guarantee for end-users "*the right to access and distribute information and content, use and provide applications and services, and use terminal equipment of their choice, irrespective of the end-user's or provider's location or the location, origin or destination of the information, content, application or service*"[7] . Beyond monitoring the practices of internet access providers as provided for in this regulation, Arcep advocates the application of the principles of internet openness to all players in the digital ecosystem who act as intermediaries between users and internet content and services. It analyses the entire digital ecosystem in order to identify, alert and inform public debate on developments that could threaten the objectives of the regulation. It was with this in mind that Arcep published a report in 2018 entitled "*Are devices the weak link in the open internet?*"[8]. In it, Arcep emphasises the need to take into account all the hardware and application layers of the internet in order to guarantee users genuine freedom of choice with regard to online content, services and applications.

The development of generative AI as an interface for accessing internet content and services, combined with a possible consolidation of the market around a few large providers, raises concerns about the openness of the internet.

Faced with these cross-cutting issues, Arcep has launched an exploratory study to assess the impact of generative AI on the open internet. This work is based on a review of scientific and institutional literature, consultations with some fifty experts from the private, public and non-profit sectors, and collaboration with the Center of expertise for digital platform regulation (PEReN), which enabled empirical tests to be carried out on several generative AI tools (see box "*IMPACTIA: a technical study conducted by PEReN in partnership with Arcep*" in section 2.1.1.).

⁂

Given the rapid change in this field, this report is a first step in analysing the current and future impacts of generative AI on an internet designed to be neutral and open, guaranteeing all users the freedom to access, share and create content and services. It is organised around the following themes:

- First, **this report assesses the effects of generative AI on the conditions of access to online content and services and on end users' freedom of choice**, to the exentent that these technologies are profoundly redefining the way internet users interact with content and, in particular, the user's interaction with sources.
- Secondly, **it analyses the effects of these technologies on the ability of internet users and innovators to share and promote content and services on the internet**, keeping in mind that generative AI is disrupting text, audio and video content value chain, and raises the question of

[7] *Ibid*.

[8] Arcep, 2018. *Smartphones, tablets, voice assistants… are devices the weak link in achieving an open internet?*.

how to promote human creations, which are essential to the abundance and diversity of the web, but also to the training of language models[9].

- Finally, **it highlights the potential risks of market concentration in generative AI services,** which could amplify these effects on freedom of choice and the capacity for innovation on the internet**.** It should be noted, however, that generative AI services are still searching for viable business models, which calls for a nuanced and cautious analysis of future relationships between players in the sector.

Based on this initial assessment, the Authority has formulated several work streams and recommendations aimed at conciliating the preservation of the openness of the internet and ecosystems with the development of generative AI.

This report thus complements the work carried out by various players on other issues related to the development of AI – such as environmental impact, strategic autonomy, data protection, security and the socio-economic transformations underway – which are not covered by this report.

[9] In this regard, it should be noted that this report does not address the issue of copyright protection in the age of generative AI, which is already the subject of much reflection, discussion and, recently, litigation (see section 2.2).

Chapter 1: Generative AI services, new gateways to the internet

The arrival of generative AI services in the internet ecosystem has been marked by the rapid spread of these technologies, which are characterised by specific design and operating features. The position of these services as a new gateway to the internet is, in some respects, a continuation of the dynamics of intermediation and centralisation of online content exchanges since the early 2000s[10] , in contrast to the network's originally decentralised architecture and the free flow of data advocated by the principles of openness and neutrality.

1.1 Generative AI services in the internet ecosystem

Generative AI services refer to tools capable of automatically producing content from natural language instructions, whether text, images, sounds or videos. In practical terms, the user formulates a request in the form of a sentence, question or instruction (a "*prompt*"), and the system generates a structured response, written text, summary, computer code, image or other type of content.

Generative AI services are spreading very quickly and support a wide range of uses. They also rely on several technical and design breakthroughs (see section 1.1.1). These include their development through training, the presentation of results in the form of a single response in natural language, and the ability of certain agents to interact directly with other digital services.

To achieve this, **generative AI services are made up of several components, one of the most emblematic components enabling them to analyse and produce text**. This component, referred to as a "foundation model," "large language model," or "*Large Language Model (LLM)*," is a very large statistical model capable of learning the regularities of language (see section 1.1.2). In order to expand their capabilities, **generative AI services can also use other components that enable them to exploit external document databases** ("RAG", see section 1.1.4) and **interact with third-party services** ("MCP", see section 1.1.5).

1.1.1 Widespread adoption and increasing integration into users' daily live

The data show a sharp increase in the adoption of ChatGPT, which reached 100 million active users within two months of its launch, an unprecedented rate for a consumer-facing application[11] . A study conducted by Ipsos[12] confirms the widespread adoption of generative AI: in 2024, 48% of the panel of French people surveyed said they used generative AI services to search the web. The spread is even faster among younger users: according to the 2025 edition of the digital barometer[13] , in the same year, 69% of 18-24-year olds use AI tools in their professional life or in the context of their studies.

Chatbots based on generative AI are used for a variety of purposes, both in everyday life and in the professional world. A recent study conducted by researchers at Open AI and published in 2025 by the *National Bureau of Economic Research* sheds light on their use based on the analysis of 1.5 million conversations and a sample of 700 million weekly active users[14] . Among these exchanges, non-

[10] BOULLIER, Dominique, 2021. The power of digital platforms, territories and sovereignties. Sciences Po, Digital Chair.

[11] UBS, 2024, *Artificial intelligence: Sizing and seizing the investment opportunity*.

[12] Ipsos & CESI, 2025. Artificial intelligence: how do French people use it?

[13] Arcep, Arcom, CGE, ANCT, 2025. Digital Barometer, 2025 edition.

[14] CHATTERJI, Aaron et al., 2025. How People Use ChatGPT. *NBER Working Paper 34255*.

professional use largely dominates: in June 2025, nearly 73% of conversations were not work-related, compared to 53% a year earlier. Three-quarters of conversations seem to focus on practical advice, information searches[15] and content writing or editing.

This adoption is not only driven by the standalone use of chatbots or conversational agents. **Generative AI services are now being combined or integrated with other digital services:**

- The main search engines offer consultation modes based on the generation of synthetic responses: Google integrates Gemini into its products[16] , Microsoft deploys Bing Copilot[17] , combining text generation and document search.
- Browsers natively integrate conversational assistants capable of summarising a page, assisting with writing, or executing commands: since 2024, Microsoft Edge has integrated Copilot, a version of its AI based on GPT-5, which can summarise articles, create content, assist with writing emails, and execute commands in the browser[18] ; Comet is a browser launched by Perplexity, which integrates the company's conversational agent[19] . Open AI has launched the ChatGPT Atlas browser[20] . Mozilla has announced the integration of a panel in Firefox allowing users to access different conversational agents[21] .
- General-purpose operating systems offer AI assistants at the desktop, search bar, or notification centre level: Windows integrates Copilot as a native assistant[22] , and Apple introduces generative AI features in iOS[23] .
- Social networks, content sharing platforms and instant messaging services offer text, audio or visual generation or rephrasing functions directly integrated into the user interface: in 2023, Snapchat launched "My AI"[24], a ChatGPT-based service similar to a virtual friend. TikTok is experimenting with content generation tools (e.g. suggestions, rewriting, assisted creation)[25] , while LinkedIn is integrating writing assistance features[26] . Meta is integrating generative AI-based chat and content generation features into its WhatsApp and Messenger messaging services[27] .

This hybridisation, accentuated by the growing integration of generative AI functionalities into their interfaces[28], confirms that generative AI services have become a preferred gateway to internet resources: rather than going directly to a website, users first address an agent, who relays, filters or reformulates the requested content or services. The agent reformulates the request, queries a search

---

[15] Searching for information and news is currently a minority activity, but is gradually becoming more common. See: OpinionWay survey for La villa numeris, *Media & AI: the new information equation*, press release dated 15 October 2025.

[16] Google, *Use the Gemini web app to get answers in Chrome*.

[17] Microsoft, *Bing is your AI-powered search and answer engine*.

[18] Microsoft, *Microsoft Edge. Your AI-powered browser*.

[19] Perplexity, *A new browser from Perplexity. Browse at the speed of thought*.

[20] OpenAI, 2025. *New: ChatGPT Atlas, the browser with built-in ChatGPT.*

[21] Mozilla, 2025. *Access AI chatbots in Firefox*.

[22] Microsoft, *Getting started with Copilot on Windows. Microsoft Support*.

[23] Apple, 2024. *Press release: Apple Intelligence is available today on iPhone, iPad and Mac*.

[24] Hazan, Raphael, 2023. Snapchat rolls out its GPT AI to all users. *Le Journal du Net*. ; Snapchat, 2023. *SPS 2023: What's next for My AI.*

[25] TikTok. *How to use AI Creation.*

[26] LinkedIn, *AI-powered writing tool for LinkedIn page posts*.

[27] WhatsApp. *About Meta AI. ;* Messenger. *Request Meta AI on Messenger*.

[28] BEIGNON, Anaëlle, THIBAULT, Thomas, MAUDET, Nolwenn, 2025. Imposing AI: Deceptive design patterns against sustainability. *arXiv*.

engine or other services if necessary, selects sources, aggregates or transforms content, and then returns a response in natural language. The user thus interacts with a summary produced by the AI service, rather than with the original content. Beyond searching, viewing and sharing content, AI agents are also capable of interacting directly with services and applications.

1.1.2 Training based on massive volumes of data collected from the web

Unlike other algorithmic services, generative AI services do not just organise content, but produce a "response" based on a language model.

The design of these language models, also known as 'training', requires a combination of high computing power and massive volumes of data.

This data is generally collected from the web, supplemented by public online sources and even other text collections: digitised books, press articles, scientific publications, code repositories and collaborative encyclopaedias such as Wikipedia.

For example, the Common Crawl organisation, which has been archiving monthly copies of the open web since 2008, is one of the main data providers for language models.

Training LLMs raises legal and technical issues such as the application of copyright law (litigation and licensing agreements with publishers and platforms) and large-scale data collection (or 'harvesting').

Generative AI: main design stages and technical foundations

The rise of generative AI services can be explained by two major technological breakthroughs in the late 2010s: the emergence of a new neural network architecture and the consideration of human preferences during training[29] .

In 2017, Google researchers presented an architecture based on the self-attention mechanism in the article *Attention is All You Need*[30] , which gave rise to transformer architectures. This innovation made it possible to train much larger and faster language models capable of absorbing massive volumes of data: large language models or LLMs. Now comprising hundreds of billions of parameters, they form the basis of the main generative AI services. Their development relies on considerable computing power, mobilising clusters of specialised processors and significant energy resources[31] .

The other key development concerns the method used to train the models. It consists of first carrying out general pre-training on huge volumes of textual data: the model is a classifier, which trains itself to predict the most probable next word from a given text[32] . This operation is at the heart of the LLM generation process: for a given input text, it has a probability for each word in the output vocabulary, and it is possible to always select the most probable word (e.g. to promote accuracy), or, conversely, to allow a random selection of words with a lower probability (e.g. to promote creative writing). The second step consists of fine-tuning targeted at specific tasks.

[29] See also: PEReN, 2024. "ChatGPT or the breakthrough of conversational AI models".

[30] VASWANI, Ashish, et al. 2017. Attention is all you need. *Advances in neural information processing systems* 30.

[31] According to the International Energy Agency (IEA), global electricity consumption by data centres reached around 415 TWh in 2024, representing nearly 1.5% of global electricity consumption. The IEA estimates that this consumption could double by 2026, mainly due to the growth of generative AI (IEA, 2024. *Electricity 2024*).

[32] For the sake of clarity, we will refer to words rather than tokens here. In practice, LLMs operate on *tokens*, which are fragments of words.

For conversational agents, this step takes the form of reinforcement learning from human *feedback (*RLHF). In concrete terms, after training a model on huge amounts of text, people – annotators – evaluate the responses produced by models to different questions according to several criteria. Their evaluations are used to represent human preferences and then adjust an LLM to incorporate these human preferences into the statistical generation process. It is therefore this step that enables the transition from statistical models, which generate sequences of probable words from massive amounts of data, to user-facing conversational models capable of responding in a more useful, natural and user-friendly manner.

1.1.3 From intermediation to natural language content generation

Generative artificial intelligence services are distinguished by their ability to produce new content from natural language instructions ("*prompts*"), using statistical models capable of interpreting, contextualising and reformulating information in a coherent and intelligible form. This generative capability, made possible by advances in automatic language processing, marks a technical shift in the design of digital interfaces.

Unlike traditional intermediation systems, such as search engines or social networks, which were limited to indexing, ranking and recommending existing content, generative AI systems produce text, images or sound that is generally 'unprecedented', developed from internal representations learned during training or synthesised from information found on the web. Generation is based on a probabilistic prediction mechanism: for each sequence of *tokens* (elementary linguistic units), the model estimates the probability of the next word occurring in the given context. This approach, known as autoregressive, allows the system to gradually construct coherent sentences, integrating both local meaning (syntax) and global meaning (coherence of discourse).

This development marks the transition from a logic of intermediation to a logic of content generation: **the conversational assistant thus becomes a direct gateway to information, making the transition from a "search engine" to an "answer engine**".

Furthermore, **this natural language presentation gives the response an educational character, structured in the form of explanatory text that facilitates the understanding of complex concepts and promotes learning**. Finally, **these interfaces often offer increased personalisation**: some services adjust the tone, level of detail or format of responses, whether text, code, summaries or lists, according to the preferences expressed by the user.

Although this conversational interface has certain advantages, it introduces a significant change in the user experience: the conversational agent performs searches (on search engines) on behalf of the end user, who no longer does so directly and only obtains transformed information. This response, tailored to the user's needs and reformulated in natural language, discourages users from consulting the original content[33] (see section 2.1).

[33] CHAPEKIS, Athena and LIEB, Anna, 2025. *Google users are less likely to click on links when an AI summary appears in the results*, *Pew Research Centre*.

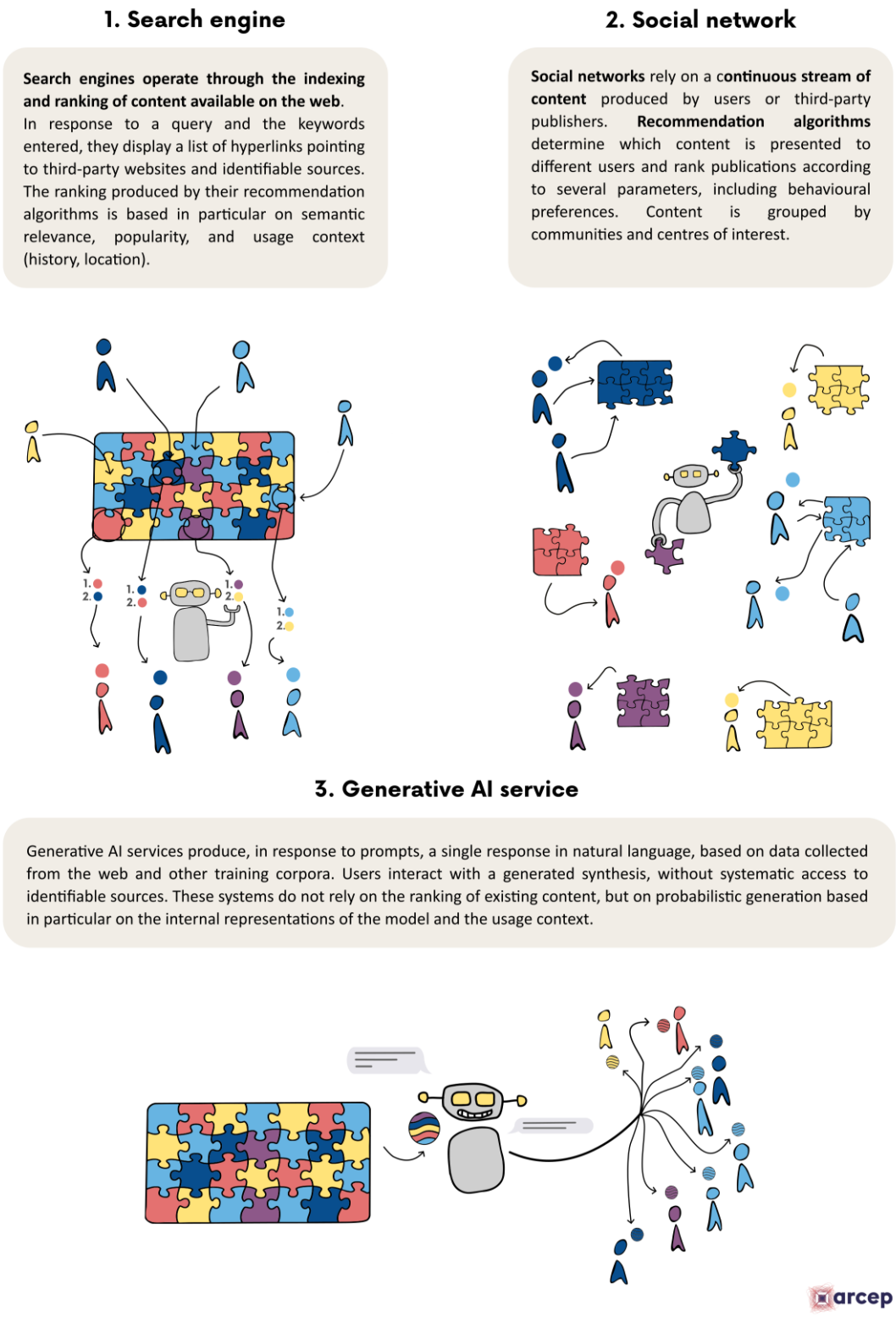


Figure 1 : From content intermediation to content generation – comparison between the recommendation mechanisms of digital services and the content generation properties of generative AI services.

1.1.4 Bias and risk of error: current situation and technological advances

Large language models do not search for established facts, but generate statistically plausible texts based on patterns observed in their training data. **This logic, which is very effective for certain linguistic tasks, can nevertheless lead to the production of incomplete, erroneous or biased content,** with consequences for the overall reliability of the content and information made available to internet users.

a) The risks of hallucinations

The models on which generative AI services are based can generate plausible but false responses. These outputs, known as "hallucinations", are not intentional but result from the statistical nature of the generation process. Faced with the difficulty of completely eliminating the risk of hallucinations, providers of generative AI services often explicitly mention their product's propensity to make errors in their interfaces[34] .

Hallucinations in language models are common[35] , as these models only generate their responses based on their training data. However, when the requested information is very specific, the statistical generation process rarely provides the correct answer. This is why a language model is often able to correctly reproduce general information, but may make mistakes when it comes to providing a date, performing mathematical operations, or presenting recent news.

However, since 2022, generative AI services have benefited from numerous technical advances that have improved tool performance, in particular by minimising the occurrence of "hallucinations" in their language models.

Generative AI services incorporate *retrieval-augmented generation* (RAG) mechanisms. RAG involves transforming the user's *prompt* into a query addressed to a search engine or database. The most relevant results are then selected and added to the prompt before the LLM generates the final response, which will therefore be much more likely to contain the expected information. Other techniques (such as *chain-of-thought* reasoning[36]) aim to encourage the model to detail its reasoning steps in order to improve its performance.

[34] Mistral AI, 2025. *Terms and Conditions of Service. ;* OpenAI, 2025. *Terms of use in Europe.*

[35] See, for example, the rankings associated with the HELM initiative: LIANG, Percy et al., 2022. Holistic Evaluation of Language Models. *arXiv*.

[36] WEI, Jason et al., 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35.

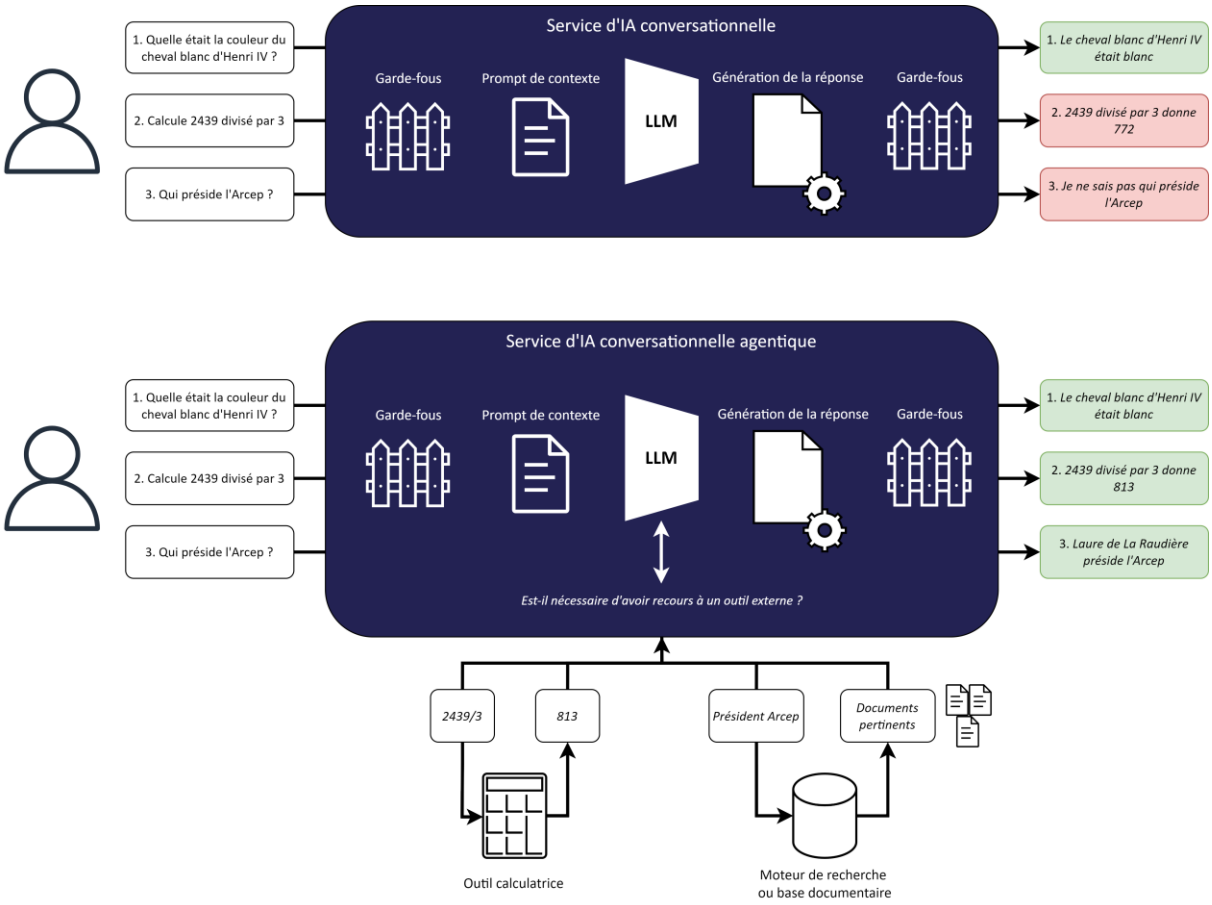


Figure 2 : Simplified representation of how conversational AI services work before and after the inclusion of agentic tools.

b) Risks of bias

Generative AI services are susceptible to producing biased responses, reflecting or amplifying pre-existing imbalances in their training data. These risks of bias are common to all algorithmic systems, including the recommendation algorithms used by search engines and social networks. However, with generative models, these biases take on a new dimension. The data sets used to train LLMs, although extensive, have not been assembled to constitute a representative sample of the diversity of voices, languages and cultures present in the digital world. Certain languages, sources or specific geographical areas are over-represented, while others, which are minority or local, may be under-documented. Without corrective measures, generative AI models can thus reproduce or amplify existing biases in their training corpora: stereotypes, homogenisation of cultural references, implicit hierarchisation of viewpoints, or even the invisibility of certain themes[37] .

[37] Numerous research studies explore these biases. Without attempting to be exhaustive, the following examples can be cited for French: DUCEL, Fanny, NEVEOL, Aurélie, FORT, Karën, 2024. "You'll be a nurse, my son!" Automatically Assessing Gender Biases in Autoregressive Language Models in French and Italian. *Language Resources and Evaluation*, pp.1495-1523; DUCEL, Fanny, HIEBEL, Nicolas., FERRET, Olivier, FORT, Karën, & NÉVÉOL, Aurélie, 2025. "Women do not have heart attacks!" Gender Biases in Automatically Generated Clinical Cases in French. *Findings of the Association for Computational Linguistics*: NAACL 2025:7145–7159.

Beyond biases attributable to training data, the various stages in the life cycle of a foundation model are all potential entry points for new biases. These biases can manifest themselves, for example, in choices made during model design, in the weighting given to certain sources during the training, deployment or adaptation phases (fine-tuning, post-training, customisation). They can occur in particular when generalist models are adjusted for use in specific domains, languages or audiences. For example, the reinforcement learning stage, which relies on human data annotation (RLHF, see box in section 1.1.2.), may incorporate normative, cultural or political biases.

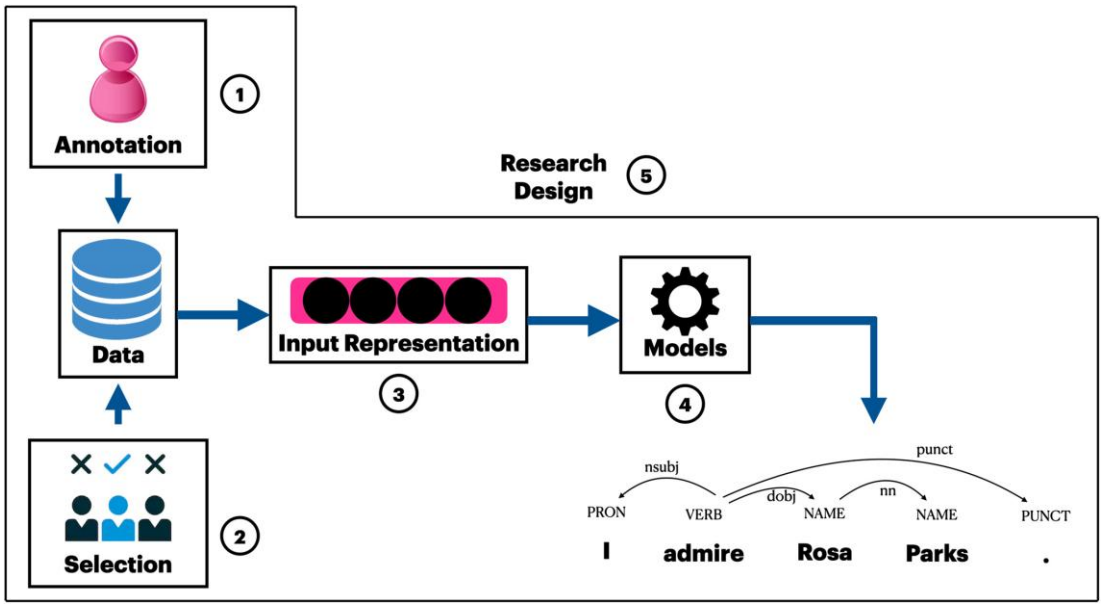


Figure 3 : Diagram of the five sources of bias amplification in automatic language processing (Hovy et al. 2021)[38] .

Arthur Perret[39] points out that LLMs are not capable of interpreting the meaning of the words or images they are given to process in queries. For example, researchers asked language models to form pairs of words that were semantically meaningful only[40] : this task proved difficult for the models. The language model does not learn the meaning of texts, but rather the statistical correlations within them: it therefore reproduces the stereotypes, power structures, or dominant opinions implicit in its training data. These biases can result in distortions in the presentation of facts, implicit hierarchies between subjects, or even a homogenisation of discursive registers.

The biases of generative AI systems are not solely due to technical constraints or statistical limitations. They can reflect design and governance choices that reflect the power relations and worldviews of the actors who develop them.

These choices are illustrated, for example, when it comes to preventing the generation of certain content. Language models are capable of generating certain types of content considered undesirable by default. AI model providers mitigate this risk by means of "safeguards", which consist of filtering model inputs and outputs using deterministic or statistical methods (see Figure 3): for example, detecting queries considered hateful or dangerous, or rephrasing the response when it contains

[38] HOVY, Dirk and PRABHUMOYE, Shrimai, 2021. Five sources of bias in natural language processing. *Language and Linguistics Compass*, e12432.

[39] PERRET, Arthur, 2025. Generative artificial intelligence in the information deadlock. *XXIVth SFSIC Congress*.

[40] DESAI, Rutvik, 2025. AIs flunk language test that takes grammar out of the equation. *The Conversation.*

personal data or data protected by intellectual property rights. These safeguards are never sufficient to eliminate all risk, and research into adversarial methods[41] remains active. In addition, they also tend to reproduce the representations of their designers, who use these rules to define what is acceptable or unacceptable to generate.

More broadly, the framing of an algorithmic system, including a language model, is theorised by some authors as a political act, in the sense that it expresses trade-offs about what is considered acceptable, representative or legitimate[42] . Behind an apparent algorithmic neutrality, in the absence of common standards, each AI service tends to embody the perspective of the socio-economic and political framework in which it is designed[43] . These aspects could contribute to the risks of fragmentation or *balkanisation* of the internet[44] , contrary to the original conception of the internet as a "common good" and global network of networks.

1.1.5 Agentic AI services: from content intermediation to digital service intermediation

The most recent development in generative AI systems is the integration of capabilities for interacting with other digital services: search engines, calendars, messaging, office tools, social platforms and third-party applications. **This new generation of models, known as agentic AI, combines the conversational capabilities of LLMs with automation functions.**

While the first versions of generative AI services "only" generated text, these services quickly benefited from interactions with additional tools, foremost among which were search engines used for RAG. Agentic AI extends this principle to enable AI services to interoperate with all kinds of digital services: calculators, calendars, social media, online mapping services, etc. Agentic AI services thus accentuate the dynamics of intermediation: they are intermediaries between the user and content, and between the user and digital services.

In practice, an agentic AI begins by determining whether any of the third-party services it can access are relevant to the user's request. It then formulates a request to send to that service (for a calculator: the calculation to be performed; for a search engine: the query, etc.). The system then activates the service with the elements provided by the agentic AI (e.g. the details of the calculation for a calculator service) and transmits the result back to the agentic AI. With this new information, the agentic AI can continue to interact with different services until it considers that it has responded to the user's query and displays its final response.

Beyond enriching textual responses, agentic AI services are now able to perform concrete tasks: sending a message, scheduling an appointment, editing a shared document, or interacting with a third-party application. This development gives these systems an operational function, transforming AI into a tool capable of orchestrating the simultaneous or successive execution of digital services.

For example, ChatGPT Atlas, unveiled in autumn 2025[45] , combines ChatGPT's conversational tool with an agent interface capable of navigating between different services and applications to perform complex tasks on the user's instructions. The agent can thus aggregate information from multiple sources, perform operations or automate sequences of actions according to defined objectives. This

---

[41] Adversarial attacks involve manipulating an AI model to obtain a "bad" result. This can involve creating images that fool classifiers, or creating prompts that bypass LLM safeguards. This is sometimes referred to as "*jailbreaking*".

[42] See, for example: MHALLA, Asma, 2024. *Technopolitics: How technology turns us into soldiers*, Paris, Éditions du Seuil.

[43] DABADIE, Raphaël, COMBES, Alexandre, DARHI, Natan, PROUST, Tom, 2025. The Political Gap Between AIs & Citizens. *Foaster*. In this article, researchers highlight a certain political bias in consumer AI services.

[44] NOCETTI, Julien, 2024. *A Splintered Internet? Internet Fragmentation and the Strategies of China, Russia, India and the European Union*. IFRI.

[45] OpenAI, 2025. New: ChatGPT Atlas, the browser with built-in ChatGPT.

gradual integration of execution capabilities reinforces the central role of agentic AI services as access and action interfaces in the digital environment. In the same vein, Open AI offers the ChatGPT Pulse service, which provides users with daily advice, ideas, reminders or recommendations based on their use of the chatbot, as well as other digital services such as Gmail and Google Calendar.

In order to interact with other digital services, whether applications, AI services or databases, AI services rely on a set of technical and contractual mechanisms that enable the exchange of requests and data. Interfacing protocols are an essential part of this articulation, but they alone are not enough to guarantee the openness of the ecosystem. Access conditions, whether resulting from contracts, licences or commercial policies, are just as important in determining the actual possibility for a third-party service to connect to an AI model or for an AI model to access third-party services.

If this spread of agentic AI continues, AI agents could become the main interface for users to access both their content and digital services. This trend could therefore revolutionise the way services and applications are made available to internet users (see Chapter 2, section 2.2).

⁂

In short, generative AI services are emerging as an additional layer of intermediation between users and content and services on the internet. This new gateway to the internet, which has been rapidly adopted by users, incorporates several major innovations compared to other digital services. While search engines and social media recommendation systems simply sort, select and order content to refer users to, generative AI services provide them with direct answers in natural language. Recent advances in agentic AI indicate that generative AI services could also become a central interface for accessing digital services. These developments are at odds with the original design of the internet, particularly with regard to its openness and decentralised nature.

1.2 The founding principle of the open internet in the face of the arrival of generative AI services

The history of the internet is based on several founding principles, including net neutrality and, more broadly, internet openness. However, in addition to internet service providers, the successive emergence of new gateways – search engines, operating systems, generative AI services – has introduced layers of algorithmic intermediation that may alter users' free access to online content and services.

1.2.1 Net neutrality: guaranteeing users' freedom of choice and the capacity for innovation on the internet

Originally, the network and internet access were designed to allow users to freely access knowledge and information disseminated by their peers, without any gatekeeper or prior validation. The internet was thus built around several founding principles, including the separation and autonomous functioning of network layers, the technical neutrality of intermediaries and, finally, "permissionless" innovation on this distributed architecture.

This history and technical functioning gave rise to the concept of net neutrality, popularised in 2003 by Tim Wu, a law professor at Columbia University, who defined it as "*a network design principle*

whereby a public network of maximum utility aspires to treat all content, sites and platforms equally, allowing it to carry any form of information and accept all applications"[46] .

This concept extends the original technical philosophy of the internet, based on the idea that the "intelligence" of the network resides at its extremities, i.e. in users' devices. Tim Berners-Lee, inventor of the World Wide Web, pointed out that "*if the Web was to become a universal resource, it had to be able to develop without hindrance. A single point of centralised control would have created a bottleneck limiting its development*"[47] . Net neutrality thus reflects the desire to avoid any capture of control over information flows, in order to preserve an open, free and innovative internet.

The internet: a decentralised network of networks

The internet is based on a distributed system comprising thousands of independent networks called autonomous systems (AS), operated by public or private entities. In 2021, the *Packet Clearing* House (PCH) organisation listed more than 17,192 interconnected networks worldwide[48] . Each AS manages its own network according to locally defined technical and economic rules, and data transmission is based on interconnection agreements, in particular the exchange of traffic between operators, commonly referred to as *peering*[49] , which reflects this logic of exchange among peers.

This technical design illustrates one of the foundations of the internet: its decentralised architecture. There is no single control centre or decision-making authority; the governance of this "network of networks" is based on a set of interconnected players who exchange traffic. While the economic balance between these players has evolved, this horizontal logic remains at the heart of how the internet works, ensuring structural resilience and a plurality of access points.

The internet is also a so-called "passive" network, in the sense that the equipment that comprises it, particularly routers, merely routes data packets without altering their content. The providers and intermediaries involved also undertake to transmit and distribute data packets on a "*best effort*" basis[50] , guaranteeing a commitment of resources rather than a defined level of performance, as well as a certain degree of transparency (*network transparency*).

This technical neutrality stems from a key design principle: intelligence is placed at the edge of the network. This principle, formulated from the outset by the designers of the TCP/IP architecture (Cerf and Kahn, 1974[51]), means that decisions on data processing or interpretation are taken not by the central nodes, but by the terminal point, at the end of the chain. The free and open protocols on which the internet is built ensure interoperability between networks, regardless of the providers or underlying technologies. This protocol openness has enabled all players, regardless of their status, to innovate and offer new services without prior authorisation.

Because intelligence resides at the edges, every connected user has, in theory, the technical capacity to send and receive data. This architecture promotes a form of equality among

[46] WU, Tim, 2003. *"Network Neutrality, Broadband Discrimination". J. on Telecomm. & High Tech*.

[47] BERNERS-LEE, Tim, 1999. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its inventor*. Harper.

[48] WOODCOCK Bill and FRIGINO Marco, 2021. *2021 Survey of Internet Carrier Interconnection Agreements*. Packet Clearing House.

[49] See Arcep's Interconnection Barometer on this subject, particularly its glossary: Arcep, 2025. *Barometer of data interconnection in France*.

[50] CERF Vinton and KAHN, 1974. *"A Protocol for Packet Network Intercommunication". IEEE Transactions on Communications.* 22 (5): 637–648.

[51] FLOYD Sally and ALLMAN Mark, 2008. "Comments on the Usefulness of Simple Best-Effort Traffic". RFC 5290. IETF.

participants: everyone can contribute to the flow of information without going through traditional channels such as radio, television or the press. The internet as thus conceived promotes a logic of decentralisation and autonomy, which are prerequisites for innovation and freedom of expression online.

In concrete terms, net neutrality is based on non-discrimination of traffic, i.e. the equal treatment of all data flows by the actors involved in their transmission, regardless of the content, service or application concerned, without arbitrary blocking, slowing down or prioritisation. **This principle protects** several essential pillars of the internet model. On the one hand, it allows **internet users freedom of choice and expression**, enabling them to freely access, produce and share content and services without alteration by an access provider. On the other hand, **it protects the capacity for innovation of economic players**, who can design and deploy new services without prior authorisation or the risk of interference from a dominant operator. In this sense, net neutrality prevents players in a position of unavoidable intermediation from exploiting their position as the "gateway" to the internet to favour their own services and content or those of their partners.

Net neutrality is therefore primarily a matter of the technical organisation of the network: by prohibiting any discrimination between data flows, it preserves the original distributed architecture and interoperability of the internet. But **net neutrality also provides an essential framework for economic dynamism in the digital ecosystem**: by guaranteeing equal access to the market for all content and service providers, it promotes innovation, competition and the emergence of new players. Finally, **it has a fully societal dimension, as it conditions the exercise of fundamental freedoms online**. By ensuring unhindered access to content and services, this principle allows everyone to obtain information, express themselves and participate in public debate without the risk of arbitrary filtering. Net neutrality thus lies at the crossroads of the technical, economic and democratic issues that underpin an open internet.

Based on this observation, the European Union wanted to enshrine net neutrality in positive law. Regulation (EU) 2015/2120, known as the "Open Internet Regulation", adopted in November 2015 and applicable since 2016, establishes a harmonised framework of protection for all European users. It imposes net neutrality obligations on internet service providers, as they are the first point of access to the internet. They must treat "*all traffic equally and without discrimination, restriction or interference, regardless of the sender and recipient, the content accessed or distributed, the applications or services used or provided, or the terminal equipment used*" (Article 3 § 3). Located at the interface between the internet and end users, ISPs have the technical and operational capacity, linked to their control of the access layers, to precisely control the reception and transmission of each data stream. The Open Internet Regulation prohibits ISPs from using this capability to, for example, limit access to certain sites or services, restrict the speed of certain applications, or impose discretionary traffic priorities.

However, the regulation provides for specific exceptions to the principle of equal treatment of traffic. Access providers may apply reasonable traffic management, provided that it is transparent, non-discriminatory, proportionate and based on objective differences related to the technical quality of service requirements of certain specific categories of traffic. Exceptional and temporary measures may also be taken to:

- prevent imminent congestion or mitigate the effects of exceptional and temporary network congestion,
- preserve the integrity and security of the network, the services provided over that network and the terminal equipment of end users.

However, these measures must not be maintained for longer than necessary.

Furthermore, the Open Internet Regulation cannot be invoked by internet access providers to refuse to comply with injunctions issued by courts or public authorities vested with the necessary powers through national legislation or EU legislative acts (Article 3(3)), in particular for the purpose of blocking illegal content or websites.

Finally, access providers may, under certain conditions, offer specialised services other than internet access services, where those services are optimised for specific content, applications or services requiring a specific level of quality (Article 3(5)).

1.2.2 The open internet, a principle extending beyond the network layer

While the regulation sets out obligations that apply specifically to the network layer, i.e. ISPs, it also describes the principle of internet openness in general terms. The regulation thus establishes the right of all end users to an open internet, i.e. "*to access and distribute information and content, to use and provide applications and services, and to use terminal equipment of their choice, regardless of the location of the end user or provider, and regardless of the location, origin or destination of the information, content, application or service, through their internet access service*" (Article 3 § 1).

This principle of an open internet is reaffirmed as fundamental in the 2023 European Declaration on Digital Rights and Principles for the Digital Decade, which commits to "*protect and promote a neutral and open internet in which content, services and applications are not blocked or degraded in an unjustified manner*"[52] .

Since the first debates on net neutrality, digital uses, the applications used to access the internet, and the digital ecosystem itself have undergone profound changes: ISPs are no longer the only recognised "gateways" to the internet**.**

The methods of accessing information and online services have evolved and now depend not only on the "network" layer but also on other layers of intermediation, **such as operating systems, browsers, app stores and large platforms**. These players, without being network operators, perform a similar routing role: they structure, filter or prioritise users' access to content and services available on the internet. At each stage, the power of intermediation has shifted to new players, reinforcing the centralisation of access routes and dependence on a few dominant entry points.

Faced with this new complexity, the openness of the internet depends on a continuum of technical and economic conditions that must guarantee every user the freedom to access, contribute and innovate, regardless of the intermediary used. **The principle of an open internet defined by the 2015 regulation therefore invites us to think of neutrality not as a rule limited to the network layer alone, but as a systemic requirement for openness, applicable to all points of access to the digital ecosystem**[53] . Net neutrality, as a legal basis, is the starting point for a broader reflection on economic openness, neutrality of access to information and the effective ability of users to exercise free and informed choice. This approach is part of a European context that recognises net neutrality and end-user freedom of choice, the foundations of the open internet, as a cornerstone of the European digital model, including with regard to their interactions with algorithmic and AI services[54] .

With this in mind**,** Arcep's technical and economic regulation focuses not only on infrastructure, but also on software and algorithmic "gateways", which now determine the internet experience. As early

[52] European Parliament, Council and European Commission, 2023. European Declaration on Digital Rights and Principles for the Digital Decade. *Official Journal of the European Union*.

[53] Arcep, 2018. *Smartphones, tablets, voice assistants… are devices the weak link in achieving an open internet?*

[54] see note 52

as 2018, Arcep highlighted in a dedicated report the importance of applying the principle of an open internet to the various "links" involved in the provision of digital services, and advocated for the opening up of user terminals. This work has continued with analyses of structuring digital platforms, including an initial note in 2019 on the challenges of internet openness and competition[55] , as well as work still in progress as part of the implementation of the Digital Markets Act[56] . Other individuals and institutions have highlighted these risks of lock-in posed by digital players. In the early 2010s, Tim Wu, founder of the concept of net neutrality, warned of the ability of digital intermediaries to become *gatekeepers*, capable of controlling the visibility of content and restricting plurality of access[57] . Many national authorities have also documented the structural effects of large platforms on market contestability and users' effective freedom of choice (see Chapter 3, Section 3.2).

Generative AI services amplify these trends by replacing a list of results with a single response or conversational interaction, and their technical specificities call for particular attention with regard to the openness of the internet. The very nature of these services, based on models trained on vast data sets, raises specific questions of transparency, reliability and bias. This is especially true since these services no longer merely mediate access to content: they themselves produce synthetic, reformulated or aggregated responses, which profoundly changes the way users access information and evaluate its source. Finally, the ability of certain AI services to interact directly with third-party services, automate actions or make operational decisions is profoundly changing the online experience of internet users, as well as that of service providers and developers.

⁂

Therefore, to guarantee the openness of the internet, it is necessary to ensure that generative AI services do not undermine users' freedom of choice or the conditions of competition and innovation that have historically shaped the internet.

[55] Arcep, 2019. Notes on structuring digital platforms

[56] European Parliament and Council, 2022. *Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828*

[57] See, for example, the following works: WU, Tim, 2010. *The Master Switch: The Rise and Fall of Information Empires.* New York: Alfred A. Knopf. Idem, 2016. *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*, New York: Alfred A. Knopf.

Chapter 2: Generative AI services, new challenges for the open internet

Generative AI services, new challenges for the open internet

Generative AI services are establishing themselves as new interfaces for accessing online content and services, to the point where, for a growing number of users, they are becoming a preferred entry point to the internet, much like search engines. As mentioned above, this development raises new challenges for the openness of the internet, extending certain issues already observed with other intermediaries and platforms, while introducing new forms of intermediation. These challenges concern, on the one hand, the ability of users to access the content and services of their choice and, on the other hand, the ability of internet users and publishers to freely share new content and services on the internet.

2.1 Impact on the terms and conditions of access to content and services, as well as on internet users' freedom of choice

2.1.1 Advances and limitations in source transparency

One of the first challenges posed by generative AI services in terms of free access to content concerns the transparency of the sources used, i.e. the ability of internet users to know, question and put into perspective the information delivered by the tool.

a) The limited transparency of large language models

The large language models used by generative AI services, whether textual, visual or multimodal, rely on deep learning techniques based on huge volumes of data. These models are probabilistic and exploit the statistical distribution observed in their training data to produce plausible textual responses. They have no semantic understanding of the texts.

Unlike a search engine, which explicitly references web documents, or a social network, which attributes content to an identifiable sender, a language model produces sentences based on probabilities of occurrence. This is why the first generation of generative AI tools, which relied solely on a language model, were in most cases unable to refer to sources or explain their outputs. When responding to a query, the model does not "know" where the information formulated within its training data set comes from. This origin is neither accessible nor traceable to the user[58] .

This opacity is not unique to language models used for generative AI services: it is part of a broader issue affecting deep neural networks, whose decisions are often difficult to explain due to their non-linearity and the large number of parameters[59] .

[58] BOMMASANI et al., 2022. On the opportunities and risks of foundation models, *arXiv*.

KHRYLCHENKO et al., 2017. Scaling Recommender Transformers to One Billion Parameters, *arXiv*.

[59] LIPTON, 2018. The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery, *ACM Queue*.

b) Improved transparency of generative AI services through mechanisms complementary to the language model

Faced with these limitations, industry players have sought greater transparency and reliability in the sources used by generative AI services. In addition to the language model, recent tools rely on RAG mechanisms that enable the use and referencing of external documents. This component, added to the language model, makes it possible to query a document database or search engine in order to integrate up-to-date and verifiable information into the response produced. The references used can then be returned to the user, thereby enhancing the traceability and verifiability of the responses generated.

This RAG mechanism is currently used by most generative AI services available. **The main conversational agents and response engines now display, for the majority of queries, links to the websites or documents from which the information was extracted**. The results of the IMPACTIA project (see box below) confirm this trend. The display of sources was evident in the majority of cases studied (only one of the three generative AI services does not display them without an explicit request). In most cases, the user can thus identify the main sources behind the response delivered by the tool. The study also shows relatively positive results in terms of source accessibility, with the percentage of broken links or error messages rarely exceeding 10% in the tools studied. The sources used also satisfactorily support the responses produced by the generative AI tools studied: they appear to be generally faithful to the content generated.

IMPACTIA: a technical study carried out by PEReN (Pôle d'Expertise de la Régulation Numérique de l'Etat) in partnership with Arcep

The IMPACTIA project, carried out from February to September 2025 by PEReN in collaboration with Arcep, analysed three generative AI tools (namely Mistral, Gemini, and Perplexity) through their application programming interfaces (APIs). The aim was to understand how these tools use their internet access to answer questions.

More specifically, the objectives were to assess the diversity of sources used by generative AI, analyse the transparency and traceability of responses, evaluate the influence of partnerships with media and content platforms, and measure the accuracy and consistency of responses.

To do this, a set of 800 questions covering four themes was used: politics (e.g. "How are laws passed in France?"), science (e.g. "How do antibiotics fight infections?"), history (e.g. "What factors led to the First World War?"), and general knowledge (e.g. "Why are creative hobbies becoming increasingly popular?"). Each question was asked 20 times to each of the generative AI tools studied, 10 times without explicitly requesting sources and 10 times with an explicit request for sources to support the answers. For queries related to current affairs, they were also asked with an explicit request to source a website with a known partnership with the provider of the generative AI tool being studied, as well as a news website with no known partnership with the provider of generative AI services in question. These questions were also asked to a search engine (Google Search) to assess possible intersections in terms of source distribution.

The sources cited were then collected and verified: 200,000 citations were analysed. The responses were checked using a third-party LLM, used as a "judge", to provide an automated analysis of the relevance of the sources used by generative AI tools to justify their responses[60] .

[60] See Figure 6 for further details.

In this document, the results of the IMPACTIA study are presented in boxes to illustrate some of the issues addressed in the report. Further details on the methodology and objectives of this project are available in the appendix (p.78).

2.1.2 Persistent challenges regarding the transparency of referenced sources

This significant improvement in source transparency compared to the first generation of generative AI services should not, however, obscure the persistence of certain limitations.

Firstly, some prompts continue to be processed without recourse to external sources, particularly when the tool considers that it has sufficient internal knowledge. In such cases, no sources are displayed.

Furthermore, users currently have limited control over the sources used by generative AI services. Although they can include certain explicit instructions in the prompt, such as requesting academic, English or public sector references, this does not necessarily influence the nature of the sources that are actually used and displayed. Indeed, tests conducted as part of the IMPACTIA project on the APIs of three generative AI services show that the services studied are not very sensitive to requests for specific citations: the services tend to return a similar number and type of sources, regardless of whether the user has requested specific information about the sources they want. **Ultimately, the user is thus placed in a receptive position, closer to that of a reader than a web explorer**[61] .

Furthermore, increased transparency does not necessarily lead to greater consultation of original sources. A study by *the Pew Research Centre* indicates that the majority of users of chatbots or answer engines are satisfied with the summarised response. They click less on the links provided than those offered by a traditional search engine, considering the information delivered in the synthesised response to be sufficient[62] . This trend could weaken internet users' ability to verify and contextualise the responses they obtain from generative AI tools.

In short, although generative AI services are evolving towards greater transparency and better source attribution, they remain somewhat opaque, due to the inherent nature of deep learning techniques and the way in which these tools currently present their sources. **This raises questions about users' ability to verify, contextualise and choose the content and sources they access, and calls for increased vigilance regarding the use of these tools**.

[61] BENDER, Emily and SHAH, Chirag, 2024. Envisioning Information Access Systems: What Makes for Good Tools and a Healthy Web? *ACM Transactions on the Web*, Volume 18, Issue 3.

[62] CHAPEKIS Athena and LIEB Anna, 2025, Google users are less likely to click on links when an AI summary appears in the results, *Pew Research Centre*.

IMPACTIA 2025 study: *does explicitly requesting sources increase the likelihood that the three generative AI services tested will cite references?*

The IMPACTIA project[63] looked at the citation of sources by the three generative AI services studied in response to explicit requests from users.

The following figure shows the rate of sourced responses from the three services studied:

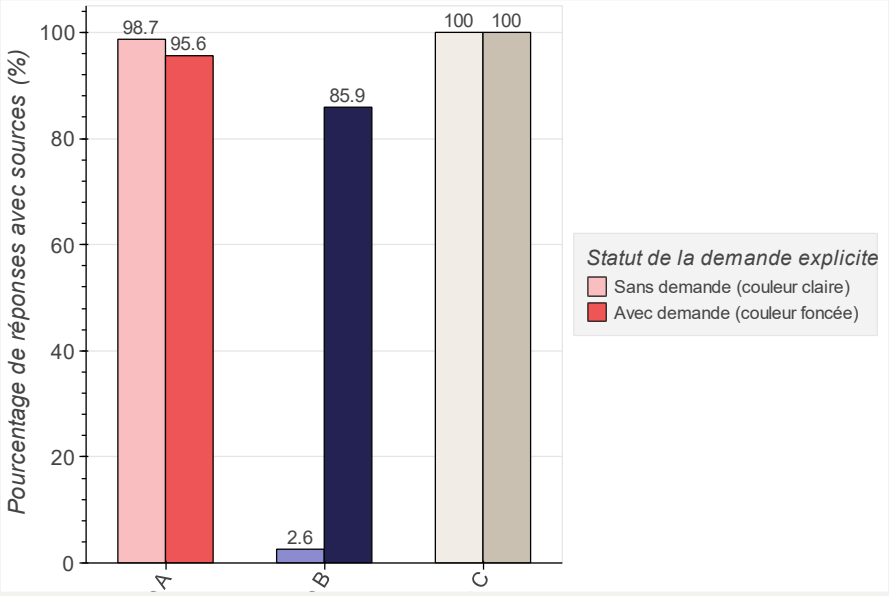


Figure 4 : Rate of sourced responses, with and without explicit user requests.

Services A and C display sources almost systematically. Service B, on the other hand, only displays them upon explicit request. Thus, we see that an explicit request for sources from the user has little or no impact for two of the three services studied, since they display them spontaneously.

[63] See appendix, p.78

If we now look at the number of citations based on the request, we obtain the following results:

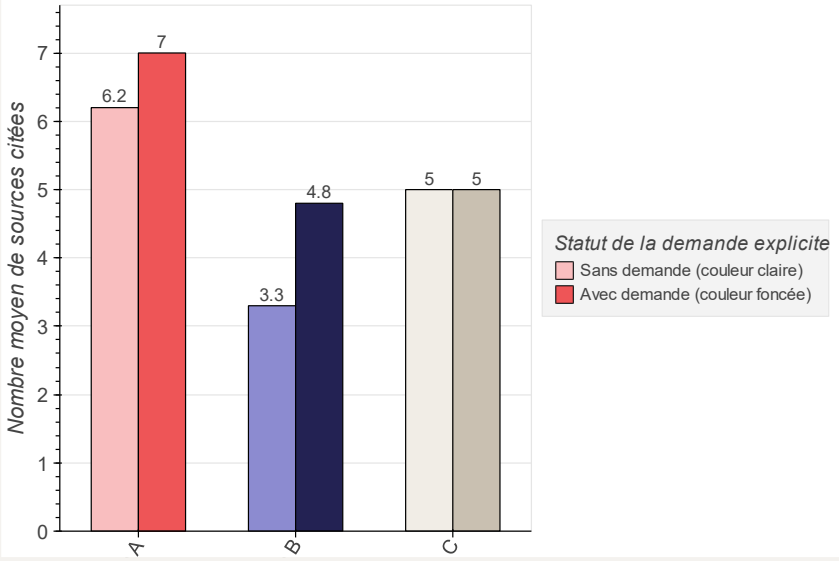


Figure 5 : Average number of sources cited, with and without explicit user request.

The results confirm that generative AI services generally display sources to users, with or without an explicit request from the user. Explicit requests for sources do not generally have an impact on the number of sources cited.

IMPACTIA 2025 Study: Are the sources used by generative AI tools relevant and functional?

The IMPACTIA study assessed the accuracy of the sources used by generative AI tools in relation to the questions asked. An LLM, used as a judge, automatically compared each link provided to the generated response and assigned it a score of 0 (no accuracy), 0.5 (partial accuracy) or 1 (complete accuracy). The score for a response was determined by the average of the scores for its sources.

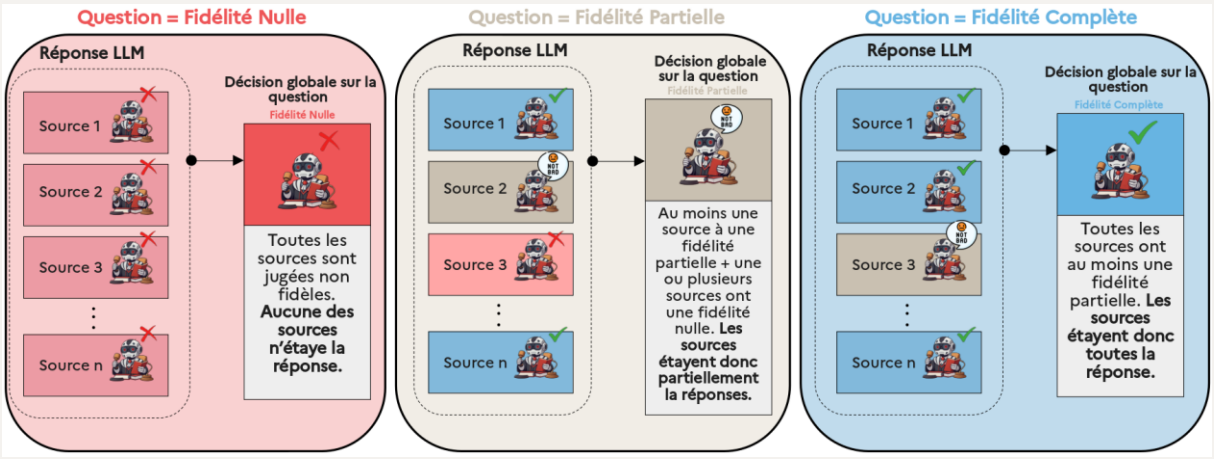


Figure 6 : Overall qualification process for generative AI responses based on fidelity judgements applied to each source cited.

With regard to the distribution of veracity, we observe that for all the tests carried out, the median is always 1 (meaning that at least half of the responses are correctly substantiated each time) and three-quarters of the responses have a score greater than or equal to 0.75 (see Figure 40 in the appendix).

The results highlight that the fidelity of generative AI services' responses to their sources is lower for political questions than for the other areas tested. Furthermore, it also appears that the absence of explicit source requests causes the performance of service B to drop in the scientific field.

The IMPACTIA project also made it possible to assess the accessibility of the sources cited by generative AI services, as well as their relevance.

A source referenced by a broken, obsolete or invented link directly compromises the user's ability to verify the information provided by the chatbot. An evaluation protocol was put in place, automating an *HTTP* test on each generated URL in order to analyse its response status based on the code returned by the server (accessible link (code 200), invented or expired (codes 404, 410), real but inaccessible to users (codes 403, 401), server error (codes 500, 202) or redirection (code 302)).

The following figure shows the proportions of valid links (code 200) for each combination of service and question topic, with or without an explicit request for sources:

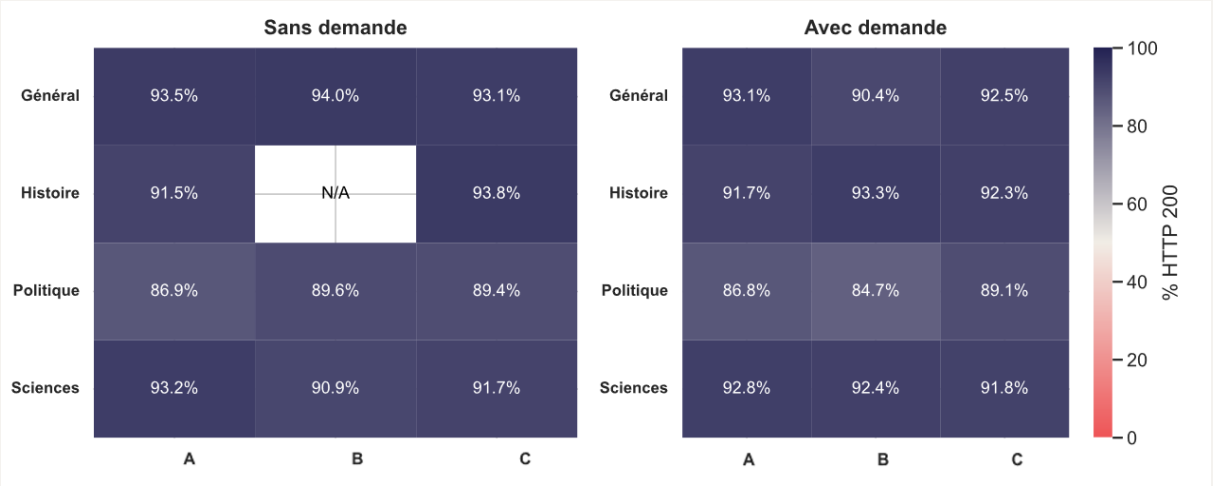


Figure 7 : Proportions of valid links (code 200) for each combination [service x question topic], with (right) or without (left) explicit request for sources

Thus, the three services studied provide accessible sources in the vast majority of cases (over 90%), although a lower reliability rate is observed for questions relating to politics, where the accessibility proportions are below 90%.

Furthermore, by studying the distribution of error codes obtained for the sources cited, according to explicit source requests, it appears that service B has a high rate of 403 and 404 errors, particularly after an explicit request. This may suggest a tendency to generate fictitious or outdated links, or to cite restricted resources. The other two services follow a similar pattern, with more than 5% of errors corresponding to an inaccessible link, even in standard use.

2.1.3 Single interface: how much freedom of choice does the user have?

The single interface of generative AI services introduces a structural change in the way users access online content and services. This new form of interface simplifies the user experience, but profoundly alters the conditions under which users exercise their freedom of choice on the internet.

a) Generative AI interface: advantages and risks of lock-in

Conversational agents provide a single response per query, formulated directly in natural language via a streamlined interface centred on a simple text input field. This type of response offers significant advantages in terms of readability, ease of use and ergonomics. For many users, the simplified interface and clear response are a real benefit. Generative AI can also facilitate access to information for audiences unfamiliar with traditional document search codes, thereby promoting a form of digital inclusion. It can also play an educational role by offering explanations tailored to the level of understanding or context of use, which enhances the informational autonomy of certain users.

However**, this facilitation of the navigation process comes with risks of confinement and loss of diversity of viewpoints**. The logic of a single response tends to concentrate algorithmic mediation into a single interaction. Generative AI makes choices in the selection of content presented to the user that will influence their access to content and services on the internet. This can result in a loss of control for users. Authors Bender and Shah[64] point out that generative AI services have fundamentally changed the structure of access to information by replacing the logic of searchable documents with that of synthesised content with no visible source. Generative AI services could alter the internet user's ability to place information in context, which is essential in a web where sources, editorial responsibility and diversity of viewpoints remain identifiable. Other studies show the incentives for cognitive disengagement when using a generative AI service[65] . These effects are reinforced by the default integration of generative AI features into digital service interfaces, which can limit users' ability to control their use[66] .

b) Concentration of sources in textual responses

Historically, the web has been based on diverse access and circulation of content, where users can explore from link to link. To find their way around, users can use search engines or other digital platforms that recommend content and services. These intermediaries, based on powerful recommendation algorithms, facilitate the user experience but also raise questions of neutrality.

Search engines have already introduced an initial form of algorithmic ranking: by presenting results in the form of ordered lists, they influence the visibility of content and focus attention on a limited number of domains. This phenomenon is not necessarily contrary to the user's interest: a certain concentration may reflect the quality or reliability of the most recognised sources, as useful

---

[64] BENDER Emily et SHAH Chirag, 2024. Envisioning Information Access Systems: What Makes for Good Tools and a Healthy Web? *ACM Transactions on the Web*, Volume 18, Issue 3.

[65] See, for example:

STADLER Matthias, BENNET Maria, SAILER Michael, 2024, Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*.

KOSMYNA, Nataliya et al, 2025, Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. arXiv.

[66] BEIGNON, Anaëlle, THIBAULT, Thomas, MAUDET, Nolwenn, 2025. Imposing AI: Deceptive design patterns against sustainability. arXiv.

information from the web is not evenly distributed among domain names[67] . Nevertheless, their recommendation systems can have a significant effect on the diversity of content actually made available to users, with identified issues in terms of net neutrality[68] and the openness of digital markets[69] .

In line with this, the weight of recommendation algorithms and attention-grabbing practices on social networks has also been the subject of extensive study. These algorithms determine not only what a user sees, but how they see it and in what order, thus strongly influencing their information browsing paths. Research has highlighted the tendency for algorithmic bubbles to emerge[70] on these social networks, tending to lock users into homogeneous information universes. The economic model of these platforms, based on massive data collection, user interfaces, recommendation algorithms, advertising systems and targeting mechanisms, limits users' capacity for choice and action.[71]

To a certain extent, the risks posed by generative AI services in terms of freedom of choice echo the criticisms that have been levelled at search engines and social networks. The selection and concentration properties of generative AI services are likely to limit the diversity of content presented to users. For example, the IMPACTIA study highlights the concentration of source citations by the three generative AI services around a handful of domain names: of the 9,206 domain names cited, the top 2% (185 domain names) account for more than 49% of the 200,000 citations. Given this concentration, it seems necessary to understand how the AI service's choices are made and to ensure that they do not compromise the diversity of content actually accessible to users.

However, several specific features should be highlighted with regard to the effect of generative AI on freedom of choice between different content.

The number of sources presented remains limited, ranging from 3 to 7 for the services studied as part of the IMPACTIA project. This priority selection logic differs from the pagination or continuous flow logic of other digital services and can lead to a loss of diversity in the sources available. The user does not necessarily access a set of results, but rather a pre-digested interpretation of the information, the basis for which (choice of sources, weighting, exclusion of certain data) remains largely invisible.

By receiving a single response with a limited number of sources, users may lose the opportunity to exercise their critical thinking skills, assess the reliability of sources, or compare different points of view.

The machine selects, reformulates and synthesises information before returning content that could replace the user's exploratory approach[72]. The internal parameters that govern the selection of sources, such as weighting algorithms, contextual relevance or quality filtering, remain complex, proprietary and rarely documented. This may limit the ability of internet users and third-party experts

[67] Similarly, web pages are not evenly distributed among domain names.

[68] BARROSO, Guillermo Andrade, MAILLÉ, Patrick and TUFFIN, Bruno, 2021. Testing search engine biases: why and how? *Interstices*.

[69] Google's dominance of the online search market and Alphabet's self-preferencing practices are the subject of work by the European Commission under the DMA.

[70] PARISER, Eli, 2011. *The filter bubble,* Penguin Books.

[71] ABITEBOUL, Serge and CATTAN, Jean, 2023. *We are social media*, Paris, Odile Jacob.

[72] LEE, Hao-Ping (Hank), SARKAR, Advait, TANKELEVITCH, Lev, DROSOS, Ian, RINTEL, Sean, BANKS, Richard and WILSON, Nicholas, 2025. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Yokohama Japan: ACM. 26 April 2025. pp. 1-22.

to understand the basis on which information is presented to them, and reduces their ability to exercise informed control over the composition of the underlying corpus.

However, generative AI services may, for example, exhibit attribution biases with regard to the sources used: certain types of documents or sources are favoured, while others are omitted[73]. Freedom of choice is thus partially delegated to the machine, without the user necessarily having the means to understand how it works or to correct its effects.

IMPACTIA study: Are the sources used by generative AI tools diverse?

Analyses conducted by PEReN for the IMPACTIA project highlight that the generative AI services studied tend to concentrate a significant proportion of their references on a small number of domain names. Conversely, a large majority of domain names are cited very little in proportion. While the concentration of generative AI, such as recommendation algorithms, may be a priority sought by internet users, it can also have negative effects on the plurality of sources and the richness of content and services actually available to users.

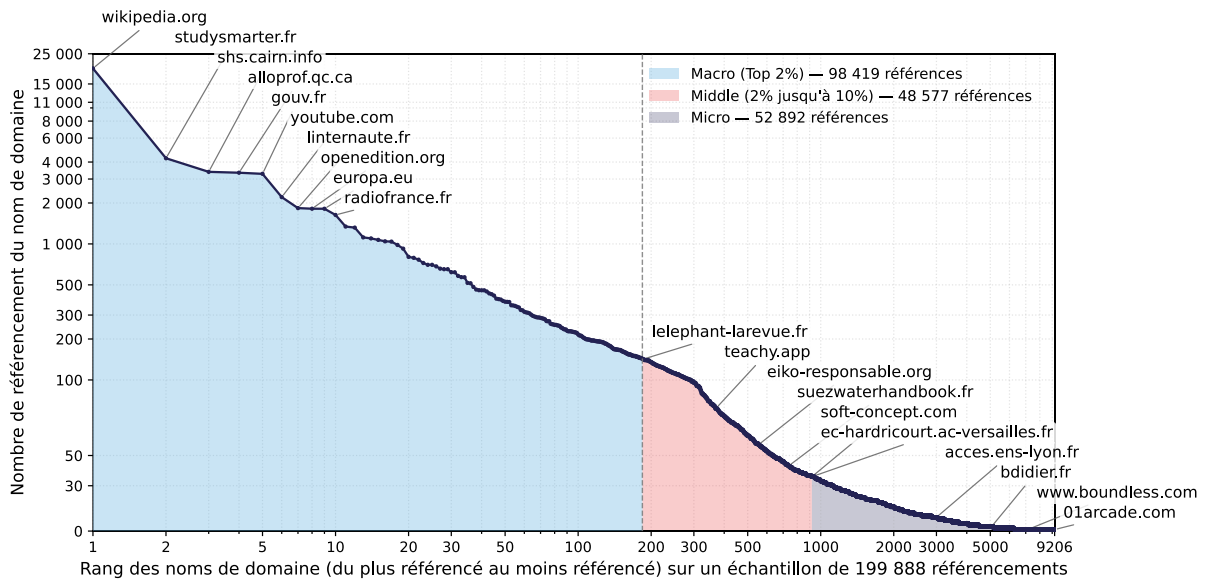


Figure 8 : Representation of domain names cited for each service evaluated. Logarithmic scales.

[73] ABOLGHASEMI et al., 2025. Evaluation of Attribution Bias in Retrieval-Augmented Large Language Models. *Findings of the Association for Computational Linguistics: ACL 2025*, pages 21105–21124, Vienna, Austria. Association for Computational Linguistics.

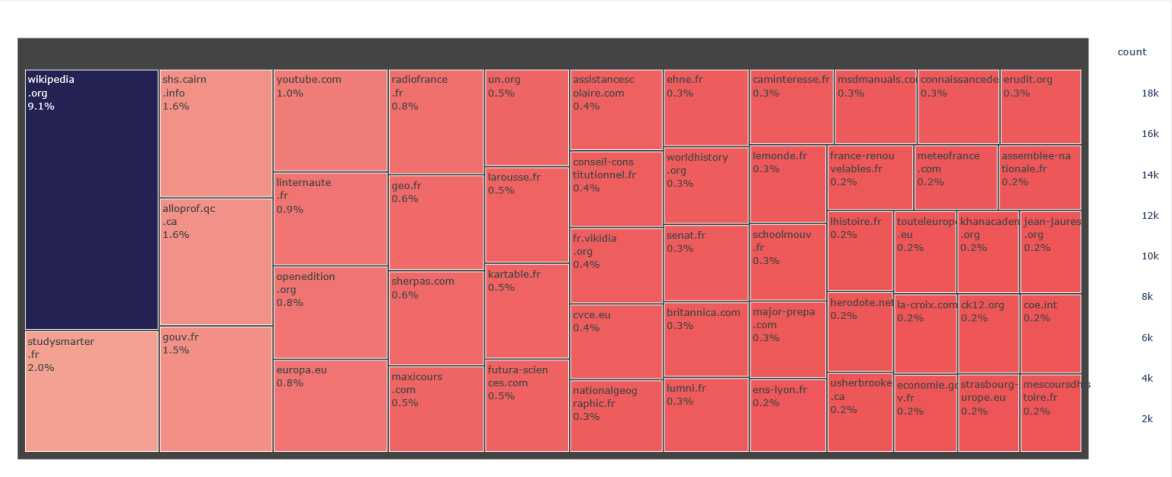


Figure 9: Distribution of citations by domain name, all generative AI services combined, all topics combined, for the top 50 domains cited

Figure 8 above shows this over-representation of certain domain names, among which Wikipedia consistently stands out as the most frequently mentioned source, regardless of the service. As shown in Figure 9, it can also be noted that homework help sites are frequently used as sources by the services studied, such as studysmarter.fr or alloprof.ca.

In addition, it should be noted that 2% of the websites referenced account for 49% of the citations, as shown in the following figure:

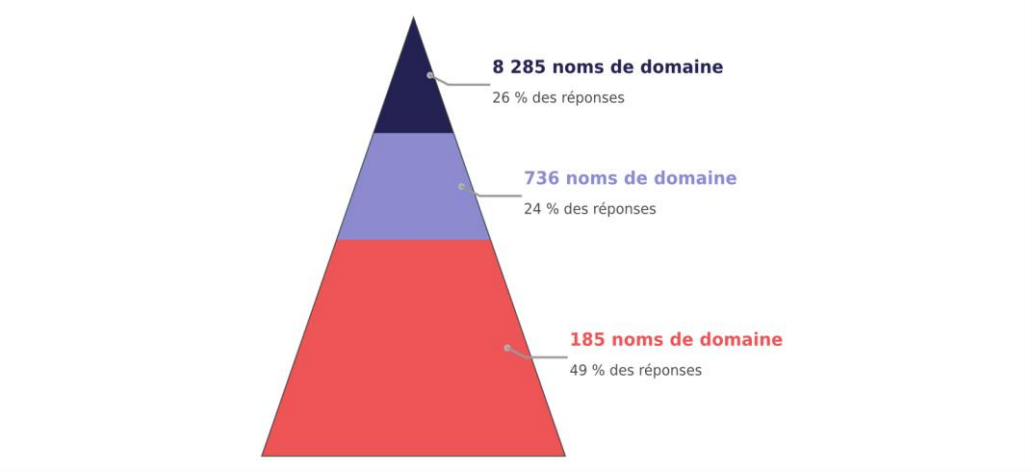


Figure 10: Concentration of domain names in the responses of generative AI services

Looking at the frequency of citations, it appears that 72% of domain names are cited only 1 to 10 times in the responses of the tools studied, out of a total of 200,000 source citations. A handful of domain names (0.4%) appear more than 500 times in the total number of citations. More specifically, the distribution of citations by domain name and their concentration can also be represented as follows:

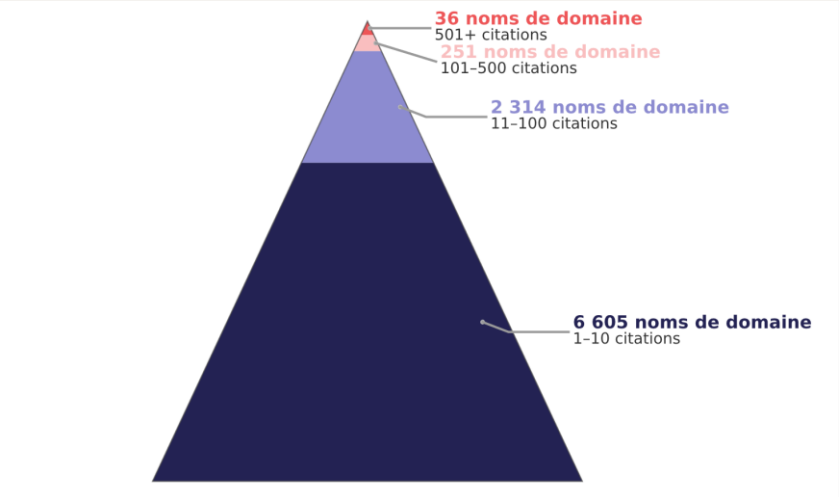


Figure 11 : concentration of citations by domain name (9,206 domain names in total).

However, there are differences in the most cited domain names depending on the generative AI service and the topic addressed by the question asked. Taking service C as an example, the concentration is not uniform: it is amplified for certain topics (science, history), but diminishes for other more open topics (politics, general)[74] .

In order to better understand the impact of this phenomenon, PEReN conducted a comparative analysis of the frequency with which each domain name was cited between the services[75] , which revealed that each service has its own citation biases, with certain sources being used more frequently than others. Also, some domain names are virtually absent from certain services despite being overrepresented in others.

In summary, these findings show that the three services studied tend to rely on a limited number of domain names to support the majority of their responses. This could potentially raise issues regarding access to diverse sources. However, the IMPACTIA study did not carry out the same diversity analyses for search engines, so it is not possible at this stage to determine whether the services studied improve or worsen the situation regarding the variety of domain names used.

In addition to the selection of a limited number of domain names and the risks of bias, some studies point out that **merging multiple voices into a single response makes it difficult to identify the original source and put it into perspective**[76]. This transformation marks a shift in the power of choice and a new delegation[77]: where users could sometimes navigate between competing proposals, they now receive a single response and may be discouraged from consulting the sources.

[74] See Figure 31 and following in the appendix

[75] See Appendix, p. 78.

[76] For uses other than information retrieval, after training, the information only exists in the form of weights in neural networks, making it difficult to find a specific source. Furthermore, companies do not always disclose their training data. See in particular: VAN WAEREBEKE, Martin and LORENZI, Marco, 2024. Learning to forget: the new challenge for artificial intelligence, *The Conversation.* And: BARR, Kyle, 2023. GPT-4 Is a Giant Black Box and Its Training Data Remains a Mystery, *Gizmodo*.

[77] RIEDER, Bernhard, 2006. *Metatechnologies and delegation: towards a society-oriented design in the Web 2.0 era*. Doctoral thesis. University of Paris VIII Vincennes-Saint Denis

Finally, the ability of conversational agents to adaptively personalise their responses, coupled with this single-response model, is likely to increase the risk of filter bubbles, i.e. informational confinement linked to algorithmic personalisation. Researchers[78] have shown that, due to their training (particularly RLHF), conversational agents develop a tendency to improvise[79] and flatter. They tend to agree with the user and not contradict them, rather than presenting a more neutral representation of the information available on the internet. Several studies highlight that the widespread introduction of conversational agents into everyday practices can amplify certain attention-grabbing dynamics. Conversational agents have the particularity of establishing a continuous interactive relationship: they solicit, follow up, adapt their style and tone, and adjust their personality according to the user's emotional context. This interaction not only promotes attention retention, but can also generate cognitive dependence on the assistance provided, which reinforces their power to guide preferences. **This has led some authors to hypothesise a shift from an attention economy to an intention economy[80], in which agents seek to anticipate users' future actions, or even towards an attachment economy, in which algorithmic personalisation aims to foster an emotional bond between users and services[81].**

c) The diversity of services in the era of agentification of uses

Beyond generating text or images, generative AI services are increasingly evolving into versatile digital agents that can interact with other services and act on users' behalf.. These conversational agents could become global interfaces for accessing the internet, eventually replacing search engines, voice assistants and even certain operating systems.

This "agentification" of the internet has obvious benefits: simplification of the user experience, automation of complex tasks, increased productivity and accessibility. But it also raises new issues of dependency, transparency and diversity of content and services. When the agent itself determines the providers or interfaces used, the user's freedom of choice risks being reduced to implicit decisions made by the service or the company that operates it. The ability of AI agents to access services is also dependent on the existence of common protocols enabling technical interoperability between these services.

This development accentuates the centralisation trend already observed: by aggregating and directly executing third-party services – booking a journey, making a purchase, generating a document, submitting an administrative request – these tools are likely to affect the very structure of access to digital services. Rather than browsing multiple sites or choosing from a range of applications, users are directed to the applications and services partnered with the conversational agent provider. They are also encouraged to concentrate their online interactions on a single interface, that of the agent.

Thus, the arbitration between several services or offers, a key element of competition and network neutrality, could be internalised within the AI service itself, without visibility for the user. In fact, the issue of service referencing and integrated preference (e.g., promoting a partner service or a brand from the same group) takes on new importance.

---

[78] SHARMA, Mrinank et al., 2023. Towards understanding sycophancy in language models. *ArXiv*.

[79] TONER, Helen, 2023, AI Chatbots Are Doing Something a Lot Like Improv. *Time.*

[80] University of Cambridge, 2024. *Coming AI-driven economy will sell your decisions before you take them, researchers warn*.

[81] BROADBENT, Stefana, ZOLYNSKI, Célia, FORESTIER, Florian, KHAMASSI, Mehdi, forthcoming, *The attention economy in the age of AI.*

This phenomenon amplifies issues already identified for voice assistants, the first step in this process of "agentification" of the internet.

From voice assistants to conversational agents: continuities and new risks for the openness of the internet

Voice assistants, interfaces capable of interpreting human speech and responding with synthetic voices, were identified as early as 2018 as "*weak links in the open internet*" due to the weight of the single response, the concentration of interactions on a single voice interface, and their potential effects on the diversity of accessible content[82] . The 2019 joint CSA–Hadopi report[83] also highlighted the risk of user lock-in and the inherent limitations of a voice-based interface, including the lack of multiple access paths, the opacity of referencing, and reduced configuration options. The report commissioned by the CSPLA in 2022 on the subject[84] , expands on these findings by showing that a single solution can reduce the diversity of cultural offerings, that voice assistants have significant filtering power, and that users may be exposed to limited visibility of certain content due to opaque interfaces or pre-installed default services.

This convergence of issues is part of a broader trend towards hybridisation between conversational agents and voice assistants. The report by the *Conseil supérieur de la propriété littéraire et artistique* (CSPLA) anticipated that advances in language models and natural language processing would enable the development of more immersive, personalised and engaging interfaces, capable of anticipating needs or adapting their responses to the user's supposed emotions. Thus, while the main providers of generative AI services most often offer the possibility of sending requests by voice, manufacturers of connected speakers have integrated generative AI services into their products, removing several structural limitations of the first voice assistants. This is the case, for example, with Amazon's connected speaker[85] , which integrates Alexa, updated with generative AI and agentic features[86], and Google's speaker, Google Home, which now integrates Gemini[87] . In addition, some AI providers are announcing the launch of new connected devices, explicitly designed to accommodate advanced conversational agents[88].

The joint evolution of voice assistants and generative AI agents is leading to renewed scrutiny of issues such as internet openness, content pluralism, service visibility and user freedom of choice.

[82] Arcep, 2018. *Smartphones, tablets, voice assistants… are devices the weak link in achieving an open internet ?.*

[83] CSA and Hadopi, 2019. Voice assistants and connected speakers: the impact of voice on supply, usage and media.

[84] ZOLYNSKI, Célia, FAVRO, Karine, VILLATA, Serena, 2022. Mission report: voice assistants and other conversational agents, CSPLA.

[85] This is the example used by Kate Crawford and Viadan Joler for their "Anatomy of an AI System".

[86] Amazon, 2025. How Amazon rebuilt Alexa with generative AI, *About Amazon*.

[87] Google, Simpler, smarter living with Gemini for Home, *Google Store*.

[88] This is the case, for example, with OpenAI, which recently announced that it had designed a prototype terminal: BONIFIELD, Steve, 2025. Jony Ive and Sam Altman say they finally have an AI hardware prototype, *The Verge*.

⁂

In short, despite tangible progress (display of sources, use of RAG), generative AI services remain characterised by incomplete transparency and a single response interface that reduces the ability of internet users to verify, contextualise and compare the information delivered by the system. The concentration and personalisation properties of generative AI services also tend to increase the risks of lock-in raised by mechanisms designed to capture users' attention, which have already been noted for other digital services. In a context where agentic AI functionalities are being deployed, this interface may also restrict freedom of choice between multiple services and encourage users to delegate this choice to their agent, thereby reinforcing its intermediary role.

2.2 The impact on innovation capacity and the sharing of online content and services

The rise of generative AI services is not only transforming the ways in which knowledge and services are accessed: it is also redefining the conditions under which new content and services can be created, shared and made visible online by publishers, developers and other internet users. By introducing new layers of intermediation and unified interfaces, these technologies can both stimulate innovation and, in some cases, restrict the open dynamics that have historically underpinned the richness and innovative capabilities of the internet.

2.2.1 The visibility of the content and services in question

a) Visibility and discoverability of online content

Generative AI services pose significant visibility challenges for content providers (websites, media outlets, independent creators, etc.).

The emergence of a single interface powered by generative AI services is profoundly transforming online visibility and attention flows. When users interact directly with a conversational agent, traffic to third-party sites tends to decrease, reducing the visibility of lesser-known domain names[89] . Where content could previously appear on the first or second page of a search engine, it may now disappear entirely from the user's perception, as they no longer have access to a list of results, but only to a single summary response. Furthermore, users do not necessarily need to visit the website that originated the content, and may potentially stick to the summary response provided by the generative AI service.

This change has led to a decline in traffic to websites, as documented by several studies. For example, the *Pew Research Centre* published an empirical study based on a panel of 900 American internet users. People exposed to a summary via the Google AI Overviews service only clicked on the link to a third-party source in 8% of cases[90] . Wikipedia also reported a significant drop in traffic following the arrival of generative AI tools, even though data from this digital common was used extensively to train language models and the site is the most referenced domain name by generative AI services, according to the results of the IMPACTIA study[91] .

---

[89] CHAPEKIS Athena and LIEB Anna, 2025, Google users are less likely to click on links when an AI summary appears in the results, *Pew Research Centre*.

[90] Ibid.

[91] Wikipedia is the most cited domain name in all categories studied, with up to 19% of citations for the "History" category. See Figure 21

This transformation raises questions of discoverability, for example for cultural, academic or specialised content[92] . The French Ministry of Culture defines this term as: "*online availability and its ability to be found among a vast array of other content, particularly by someone who was not specifically searching for it*"[93]. It points out, as does the International Organisation of La Francophonie[94], that the discoverability of French-language cultural content is a major challenge for preserving diversity in an environment where visibility is increasingly determined by global algorithms[95]. Another example: the INA media review examined the issue of media discoverability in the use of ChatGPT[96]. It found that the chatbot significantly limits the number of media sources available for consultation, as the links referenced do not only refer to media sites. This reduction in visibility is not only a cultural or informational issue. It also has direct economic consequences for publishers, as discoverability largely determines their ability to generate revenue.

This situation is part of an environment in which the logic of source prioritisation and referencing by generative AI services is still being structured. Generative AI services do not rely on the same indexing or referencing mechanisms as search engines. Thus, very limited convergence is observed between the top results of Google Search and those generated by generative AI tools, both in the IMPACTIA study (see box below) and in early academic work on the subject[97] .

While search engines select and rank their results using algorithms that reference web pages, documented signals such as site structure, content freshness, and domain authority, generative AI services select their sources according to rules that remain largely opaque and are neither stabilised nor standardised. In addition, the diversity of referenced domains, the freshness of content, consistency over time, and sensitivity to query variations vary for each generative AI service.

This makes it more difficult for content providers to build a content discoverability strategy. **This evolution marks the shift from strategies based on *search engine optimisation* (SEO) to the determination of new strategies based on *generative search engine optimisation* (GEO)**[98], in which the challenge for content and application providers is no longer to appear high in a list of results but to be identified as a relevant source by the generative agent and actually cited in the response.

The rules for ranking, weighting, or citation are not controlled by publishers, and the sponsored visibility mechanisms found in traditional search engines are not systematically implemented in generative AI tools. Faced with these uncertainties, initial initiatives are emerging from players specialising in optimising visibility in generated responses, offering strategies aimed at improving the recognisability and citation of content in responses generated with generative AI.

---

[92] GUILLAUD, Hubert, 2024. *AI, cultural reducer: towards a world of similarities*, *In algorithms*.

[93] Ministry of Culture, 2020. Online discoverability of French-language cultural content.

[94] ROBINET Fabrice, 2025. Do chatbots speak your language? The battle for digital diversity. *UN info*. United Nations.

[95] Ministry of Culture, 2025. An action strategy for cultural and responsible artificial intelligence.

[96] KOCH Olivier, 2025. How ChatGPT chooses its sources to answer your questions about current events, *La revue des médias*. INA.

[97] See in particular: CHEN, Mahe, WANG, Xiaoxuan, CHEN, Kaiwen and KOUDAS, Nick, 2025. *How to Dominate AI Search: Lessons from Large-Scale Comparison of Google and AI-Powered Search Engines*. arXiv. arXiv:2509.08919

[98] AGGARWAL, Pranjal, MURAHARI, Vishvak, RAJPUROHIT, Tanmay, KALYAN, Ashwin, NARASIMHAN, Karthik and DESHPANDE, Ameet, 2024. GEO: Generative Engine Optimisation. *arXiv*. arXiv:2311.09735.

IMPACTIA study: Do the generative AI services tested rely on the same sources as search engines?

The IMPACTIA project compared the similarity of the sources proposed by the three generative AI services studied with the search results displayed on Google for the same question.

To do this, two types of measures were used:

- Intersection: to measure the proportion of common sources, regardless of their order, which makes it possible to determine whether the links proposed by the generative AI service could have been found by a conventional search

- *Rank-biased overlap* (RBO): to assess the similarity between two ordered lists, giving more weight to the top positions. The higher a source appears in Google's ranking and is also cited by the generative AI service, the more it contributes to the score. This reflects a more "qualitative" alignment with the search engine's ranking priorities.

The analysis was conducted on a set of questions covering four topics: general questions, politics, science and history. The results are as follows:

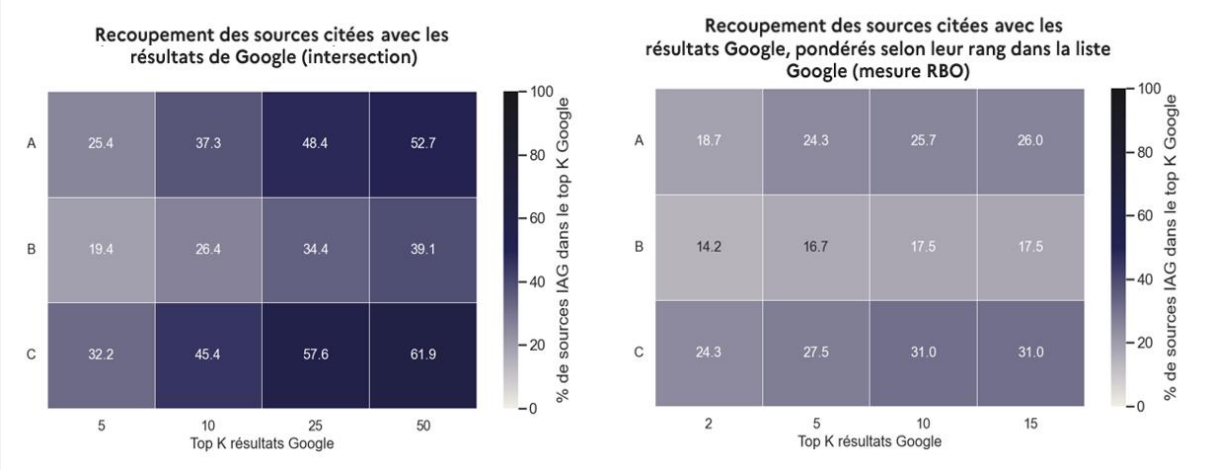


Figure 12 : Cross-referencing of sources cited by the three generative AIs studied with Google results by domain name, by intersection (left) and by RBO (right).

The data shown compares the results based on the full URLs referenced. It appears that the sources cited by the AI services studied partially overlap with Google search results, as shown by the intersection results. Indeed, when comparing the citations of queries with the top 5 Google Search results, the intersection rate ranges from 19% to 32%.

The results obtained using the *rank-biased overlap* method provide additional dynamic insights. Taking into account the order of the results, with a much higher weighting given to the top positions (top 1, 2, 3), service C, which is the most aligned, has an RBO score of 31% with the top 10 or 15 Google results. Thus, the convergence between the top search engine results and the sources proposed in the responses of the three generative AI tools appears to be limited.

This comparison shows that the generative AI services tested do not systematically offer the same flagship content as Google, but share a common base of sources with it, which becomes apparent as soon as the scope is broadened beyond the top 15. This observation is not intended to judge the quality of these hierarchies, but provides a useful benchmark for comparing two approaches to accessing information: that of search engines and that of the generative AI services tested.

b) Agentic AI and service visibility

So-called agentic AI systems, capable not only of generating content, but also of executing actions and autonomously choosing which services to mobilise, are emerging as entry points for accessing services on the internet.

The user makes a request, and it is now the agent that determines which services or providers will be called upon to respond. **This shift in the decision-making point transfers part of the user's choice to the AI system and reconfigures the mechanisms of service visibility and discoverability.** Those that are not actively selected by agents risk becoming invisible, especially when no comparison or selection interface is offered. When an agent selects a provider by default without offering visible alternatives or a comparison interface, access to alternative or more specialised services could be compromised.

The rise of these agents raises the question of functional interoperability. **To preserve a truly open internet, it is important that users are free to choose the services their agents use and that developers can access them using open protocols. However, without effective interoperability and transparent, non-discriminatory access rules, both technical and contractual, the risk of lock-in increases**.

By relying on pre-existing technical or commercial integrations, AI agents are likely to concentrate demand around a limited number of services[99] . This risk arises when certain players reserve the use of essential services or data for their own solutions, for example by limiting access to mapping resources or search capabilities for the sole benefit of their conversational assistant. It also arises when dominant AI services restrict or refuse to interface with third-party services, compromising their visibility and limiting their capacity for innovation. Such a situation would weaken the diversity of available services and undermine the structural conditions of an open internet.

Added to this dynamic is the proliferation of bilateral agreements and partnerships between AI providers and service and application providers. In 2025, for example, OpenAI entered into agreements with Walmart, Etsy and Shopify, allowing ChatGPT users in the United States to purchase items directly on these platforms from the agent interface[100] . More broadly, the company highlights a list of partner applications accessible via its US interface, outside the European Union, including Booking, Canva, Coursera, Spotify and Expedia. The company has also announced its intention to expand this offering, under technical and pricing conditions that have yet to be specified[101] .

Other examples follow the same trajectory. Claude offers access to several partner services, such as Notion, Canva and Stripe[102] . Windows highlights similar integrations with OpenTable, Kayak and Instacart via the Copilot agent[103] . These integrations reinforce the role of conversational agents as an operational access point to digital services. **In the absence of openness and transparency regarding the conditions of access to these new agent interfaces, such agreements and partnerships are likely to ultimately confer an advantage in terms of access or visibility to certain services at the expense of others, without the user necessarily being aware of it.** They may also create commercial or technical dependencies between AI providers and service developers, and contribute to structuring an environment where the presence or absence of a service among an agent's partners determines its ability to be effectively accessible to users.

---

[99] CERRE, 2024. *AI Agents and Ecosystems Contestability*.

[100] Reuters, 2025, OpenAI partners with Etsy, Shopify on ChatGPT payment checkout.

[101] OpenAI, 2025, *Deployment of applications in ChatGPT and new SDK Apps*.

[102] Claude, 2025, *Discover tools that work with Claude.*

[103] Windows Blog, 2025, *Microsoft Copilot improvements for Windows 11.*

This development may also limit the ability of emerging service providers to be discovered in an increasingly centralised environment. The risk is that new forms of "closed intermediation" will emerge, in which agents implicitly choose which services are accessible or visible, without transparency on the criteria used.

Although AI agents represent a major innovation, these conversational intermediaries could become proprietary overlays on top of the open internet infrastructure, limiting the ability of internet users, developers and publishers to share new content and services. **In this context, issues of transparent and fair access conditions to these interfaces, as well as those of interoperability and openness of interfacing or interconnection protocols, become central to ensuring genuine plurality of internet services accessible via AI agents**.

2.2.2 The openness and transparency of technical and economic relations between generative AI services and content and application providers

The **challenge of standardising technical and economic relations between generative AI services and content providers is part of a historical continuum: that of building an internet based on open and standardised interconnections**. Just as the lower layers of the network (IP, DNS, HTTP) have relied on the neutrality of technical intermediaries and shared protocols to ensure the smooth and non-discriminatory flow of data, it is now necessary to design and use open and standardised interconnections between AI agents and web services or content at the application level.

a) Protocols and automated data collection by *crawlers*

In the current ecosystem, visibility is one of the major factors contributing to the economic model of content and service providers (websites, media, digital commons, etc.) through advertising revenue, subscriptions and donations. The issue of sharing the value of content between AI companies and rights holders is a fundamental one and is the subject of national and European discussions on adapting the regime of related rights and copyright in the face of the development of generative AI services. Indeed, the use of protected content by generative AI tools causes a disintermediation of rights holders' services and a loss of their revenues.

With regard to the relationship between generative AI services and content and application providers, it appears that, from a technical standpoint, the absence of open standards or protocols governing interactions between AI agents and web content may constitute an obstacle to smooth relations between these players.

One of the central issues in these relationships concerns the methods used to collect data and content available on the web. Since 2022, several website publishers have recorded increased traffic on their infrastructure[104] from *web crawlers*. Indeed, many observers have noted that traffic from robots (or *bots*) has increased on the web.

These bots, initially deployed by search engines to crawl the web in search of new content, analyse the page, retain its structure, and add it to the search engine's index. It is through them that search engines keep their mapping of the web up to date. Crawlers are now an integral part of how the web works.

To prevent the indexing of useless pages or pages that are not intended to be public, and to prevent bots from overloading the infrastructure of website publishers, the ecosystem quickly adopted the robot exclusion protocol, embodied in the "robots.txt" file, proposed in 1994 by Martijn Koster[105] . This is a simple file placed at the root of the website which uses simple directives to indicate which pages

[104] WEINBERG, Michael, 2025. *Are AI Bots Knocking Cultural Heritage Offline?* Glam e-lab.

[105] KOSTER Martin et al, 2022. *RFC 9309: Robots Exclusion Protocol*. IETF.

of the site the robot can index and which are off-limits. It is possible to refine the rules according to the "*user-agent*", i.e. the way in which the robot identifies itself. The robot makes its request by presenting a kind of "business card" that is supposed to contain, for example, the company on whose behalf it is acting or the purpose of its processing (e.g. archiving or indexing for a search engine). This protocol is based on mutual trust between websites, *crawlers* and those who deploy them.

Generative AI players also use *crawlers* to collect data from the web for training purposes, but also to search for more recent data for agentic functionalities. Some are not operated directly by AI providers, although the data collected is used by them: this is the case with *Common Crawl crawlers*[106], an organisation whose database has been used to train many generative AI models. Some are shared, while others operate specifically on behalf of a single company.

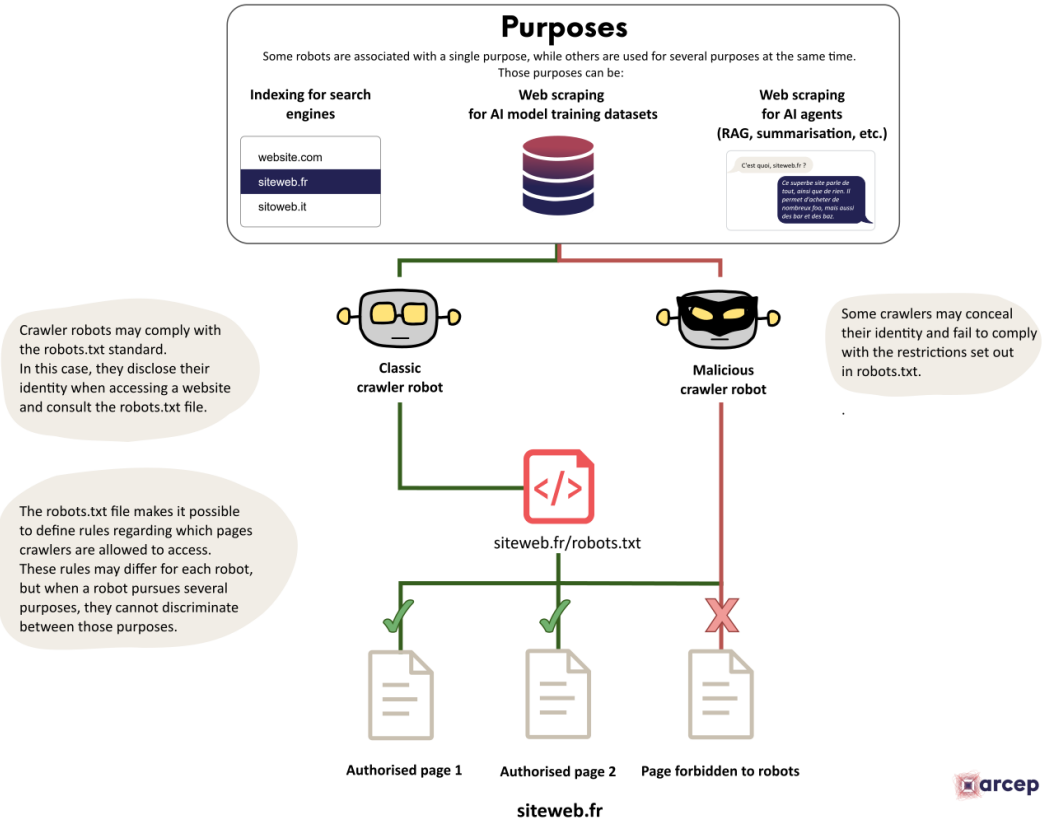


Figure 13 : The different behaviours of *crawlers* with regard to robots.txt.

Although there are no technical restrictions preventing them from doing so, it was customary for the *crawlers* of the major search engines to comply with the recommendations issued by websites through the robots.txt file. However, in recent years, *crawling* practices that do not comply with the robots.txt protocol guidelines have been observed. Some *crawlers* can overload third-party servers[107] ,

[106] Common Crawl is a non-profit organisation founded in 2008 that conducts open and regular web harvesting to build public archives of web pages accessible to all. More information is available on the website.

[107] The Wikimedia Foundation reports that 65% of its traffic comes from robots: Wikimedia Foundation, 2025. *How crawlers impact the operations of the Wikimedia projects.*

sometimes to the point of making content unavailable[108] with an effect similar to a denial-of-service attack. Overall, the impact of *crawlers* is far from insignificant. Cloudfare estimates that *crawler* traffic could exceed human user traffic by 2029[109] .

The lack of common standards for filtering the activity of these *crawlers* creates uncertainty on two levels.

On the one hand, content providers want to be able to technically control the use of their productions and publications by generative AI tools, as well as their promotion. However, there is no mechanism for specifically filtering AI *crawlers* without causing negative effects for search engine crawlers and, therefore, on the website's search engine optimisation.

On the other hand, from the perspective of generative AI developers, although a European regulatory framework is gradually being put in place[110] , the absence of uniformly respected protocols can be a source of technical, economic and legal uncertainty. The collection and use of online data for training or updating models is therefore based on a wide variety of practices. The conditions for accessing data vary greatly from one site to another, with implicit or explicit authorisation terms left to the discretion of publishers. Developers face uncertainty about the legitimacy of their collection practices, even when these are based on historically tolerated uses (such as *crawling* for indexing purposes). This situation may create asymmetries between pioneering players in generative AI and potential emerging players: the former have been able to collect this data without restriction, while the latter may encounter difficulties in accessing the data necessary to train their models in accordance with the frameworks that are currently being established. Finally, the lack of common and interoperable standards could slow down the deployment of agentic AI, a factor of innovation for the generative AI services market.

In response to these difficulties, initial sectoral and technical initiatives are emerging to try to facilitate relations between generative AI and content and application providers.

Some technical players, such as hosting or network security service providers, offer mechanisms for identifying and filtering AI *crawlers*. These mechanisms, based on the creation of lists of authorised robots, aim to enable publishers to better control access to their content without blocking the general functioning of the web. In addition, some of these players are exploring the implementation of microtransaction or 'pay-per-crawl' systems, allowing an AI agent to automatically remunerate publishers whose content is accessed, according to a predefined scale. These mechanisms, which are still experimental, could help to create a balanced commercial relationship between content producers and AI services, while ensuring the traceability of usage.

[108] ALLEN Will, 2024. Giving users choice with Cloudflare's new Content Signals Policy. *Cloudflare blog*.

[109] *Ibid.*

[110] See in particular: European Commission, 2025. The General-Purpose AI Code of Practice.

Cloudflare, a new pay-per-crawl business model

Cloudflare, a company that provides infrastructure and security services to around 20% of the world's websites, recently made headlines with an unprecedented proposal: the implementation of a "*pay-per-crawl*" business model[111] .

Cloudflare protects websites against traffic spikes and denial-of-service attacks. To do this, it has the ability to block malicious traffic, a mechanism that it can apply to *crawlers*. The company is now proposing to go beyond simple blocking by introducing a mechanism for the regulated monetisation of *crawling*. This system, currently in beta, is based on two technical principles:

- the use of agentic features, which allow automated agents to perform simple actions, such as micro-payments, on behalf of their owner;
- and the use of the HTTP 402 "*Payment Required*" response code, which has been part of the HTTP protocol since its inception but is rarely used.

In concrete terms[112] , when a robot attempts to access a page protected by Cloudflare, the server does not return the usual 200 code (access authorised), but a 402 code indicating that payment is required. This message is accompanied by a rate set by the content provider[113] .

Cloudflare then acts as a technical intermediary: it allows the AI agent to make a microtransaction to access the content, with the amount then being paid to the source site. This mechanism introduces a logic of micro-remuneration per request, aiming to establish an economic balance between access to data for generative AI services and remuneration for content providers.

At the time of writing, several similar initiatives have emerged[114] , led by other digital players. These initiatives demonstrate the ecosystem's growing interest in technical and economic solutions that reconcile open data, transparency of use and value sharing in the context of generative AI development.

In addition, several proposals for standards are emerging to complement or replace the robots.txt protocol. Among them, the "ai.txt" project[115] (*Artificial Intelligence Access Protocol*) aims to enable publishers to specify, in greater detail, the conditions under which their data can be explored, reused or excluded by AI agents. The *World Wide Web Consortium* (W3C) has also been working since 2023 on updating *Text and Data Mining* (TDM) protocols to take into account uses related to generative AI. Finally, a working group on artificial intelligence within the IETF[116] is attempting to support efforts to standardise the intended uses (indexing, training, generation, search) of web content.

Standards are also being developed to support the deployment of agentic AI and avoid lock-in, reduce development costs and ensure fair access to key online services. The *Model Context Protocol*

---

[111] BELANGER, Ashley, 2025. Pay up or stop scraping: Cloudflare programme charges bots for each crawl, *Ars Technica*.

[112] ALLEN, Will and NEWTON, Simon, 2025. Introducing pay per crawl: Enabling content owners to charge AI crawlers for access. *Cloudflare blog*.

[113] Mozilla Developer Network, *402 Payment Required*

[114] Notably, Akamai offers this service in partnership with Tollbit and Skyfire to monetise content harvesting, and Microsoft appears to be seeking to do the same, in the form of a marketplace to compensate content publishers.

[115] LI, Yuekang, SONG, Wei, ZHU, Bangshuo, GONG, Dong, LIU, Yi, DENG, Gelei, CHEN, Chunyang, MA, Lei, SUN, Jun, WALSH, Toby and XUE, Jingling, 2025. ai.txt: A Domain-Specific Language for Guiding AI Interactions with the Internet. *arXiv*.

[116] Internet Engineering Task Force, the standardisation body responsible for Internet standards, particularly those governing protocols such as TCP/IP.

(MCP)[117] , supported by several AI players, aims to facilitate secure interoperability between AI services and third-party services by defining a common language for API calls, execution contexts and access rights management. In addition, protocols such as *the Agent-to-Agent Protocol* (A2A)[118] and the *Agent Communication Protocol* (ACP)[119] aim to structure communication and collaboration between AI agents from different providers, so that these agents can interface freely with each other.

Thus, the opening up and standardisation of these interconnections in the application layer is an essential step towards ensuring compatibility between different generative AIs and preserving the capacity for innovation at all levels of the digital chain.

b) Special case of relationships and agreements between generative AI services and press publishers

Relationships between generative AI service providers and news publishers raise particular issues, as the media sector is regulated in terms of both content[120] and certain distribution channels. However, like certain digital services (digital newsstands, aggregators, search engines, social networks), generative AI is becoming a new channel for distributing information, which raises questions about its reliability, neutrality, transparency and content remuneration.

Generative AI companies rely, among other things, on journalistic content, on the one hand, to train their language models and, on the other hand, as reference sources (RAG) from which responses to users are synthesised. At the same time, news publishers are among the content providers whose business model is particularly challenged by certain generative AI tools. Disintermediated by services such as Google's AI Overviews[121] and Open AI's ChatGPT Pulse[122] , which summarise information and deliver it in a personalised way to users, independent website publishers report a sharp decline in their audience[123] , with an impact on their revenues.

In this context, some publishers are denouncing what they consider to be "plundering" of their content and creations. Some are choosing to exercise their right to opt out[124], i.e. to block access to their content[125] . In addition, some media outlets have taken legal action against generative AI services, such as the *New York Times* against OpenAI and Microsoft in 2023, accusing them of training their models on its data without its consent[126]. This complaint raises the issue of infringement of the newspaper publisher's business model on the one hand, and the asymmetry in negotiations between generative AI players and newspaper publishers on the other. Indeed, AI providers can pit press groups against each other to obtain the most favourable terms for access to their content[127], to the detriment of pluralism (see below). At European level, a group of independent European publishers, including the

---

[117] Model Context Protocol, 2025. *Specification*.

[118] Linux Foundation, 2025. Agent2Agent (A2A) Protocol Specification (DRAFT v1.0).

[119] Agent Communication Protocol, 2025. *Welcome. Get to know the Agent Communication Protocol*.

[120] Arcom, 2024. Review of the "AI mission" on the impact of artificial intelligence in the fields of creation and information.

[121] Google Search. *Ask whatever's on your mind*.

[122] OpenAI, 2025. *New: ChatGPT Pulse*.

[123] ALBA, Davey and LOVE, Julia, 2025. Google AI Search Shift Leaves Website Makers Feeling 'Betrayed', *Bloomberg*.

[124] PEReN, 2024. Status report on opt-out.

[125] Société des Droits Voisins et de la Presse, 2023. *Artificial intelligence: the Société des Droits Voisins de la Presse (DVP) exercises its right to opt out.*

[126] GRYNBAUM, Michel M. and MAC, Ryan, 2023. The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work. *New York Times*.

[127] JOUX, Alexandre, 2024. Springer, Le Monde and not the others: the press's complicated relationship with OpenAI. *The European Media and Digital Review*.

Independent Publishers Alliance, the *Movement for an Open Web* and Foxglove Legal, has filed an antitrust complaint with the Commission, alleging abuse of a dominant position in relation to the use of web content to feed Google's AI previews and summaries[128] .

In order to standardise these relationships, **generative AI services have entered into bilateral agreements with a handful of news publishers**. These technical and financial agreements may include the purchase of training licences, the possibility for generative AI companies to access publishers' data in order to enrich their language models, and the referencing of the media as a source of news content to develop responses to user queries (thanks to RAG). Responses are thus supplemented with sources and links to the signatory publishers.

The dynamic so far has been essentially bilateral and opportunistic, with each AI provider negotiating directly with press groups. The agreements thus concluded do not follow an industry-wide approach. According to some stakeholders, the still limited proportion of press publishers that have concluded a bilateral agreement with an AI service can be explained by the refusal of AI providers to negotiate agreements with several publishers in the same country or at industry level, and by the refusal of some press publishers to give these generative AI services access to their data.

The lack of a coordinated approach at industry level and the conclusion of these bilateral agreements between large generative AI companies and certain press publishers could lead to changes in the economic balance of power.

Among the risks identified by media players is that of strengthening the dominant players in the generative AI and press services markets to the detriment of diversity and pluralism. By giving preferential access to content from major partner publishers**,** generative AI services can increase the visibility of these sources through their responses, thereby enhancing their reputation and commercial appeal[129] . Conversely, reducing the exposure of non-partner media could weaken their ability to reach readers and also to negotiate fair contractual terms with AI services in the future. In addition, these agreements enable AI companies with the necessary resources to strengthen the quality of their products and differentiate themselves from their competitors.

These dynamics could ultimately have competitive effects on the markets for generative AI services and the press: large publishers could consolidate their position as reference sources, and large generative AI companies as preferred channels for distributing news data.

Therefore, beyond the financial and strategic issues, one of the major risks that may arise from these bilateral agreements and, more generally, from asymmetry in access to press data, would be a reduction in the diversity of viewpoints in the responses provided by generative AI, with potential impacts on **media pluralism and the quality of information.** To date, there are no safeguards in place to ensure press pluralism *through* these tools.

[128] Foo Yun Chee, 2025. Google's AI Overviews hit by EU antitrust complaint from independent publishers. *Reuters.*

[129] HAGEY, Keach, 2023. News Publishers See Google’s AI Search Tool as a Traffic-Destroying Nightmare, *The Wall Street Journal*.

IMPACTIA 2025 study: Do partnerships between the AI providers studied and the media impact the dynamics of citations on current affairs issues?

The IMPACTIA project[130] looked at the impact of partnerships between AI providers and the media on citation dynamics on current affairs issues. This study observed, in the responses of the services tested to a series of news-oriented queries, the proportion of links originating from known partners. The prompts were formulated several times: without requesting a specific source, requesting a citation from a media outlet that had signed a partnership with the generative AI service provider, and requesting a mention of another media outlet's website with no known agreement with the provider.

As a result, the results observed in the IMPACTIA study vary greatly depending on the service.

For service C, links to the partner publisher's site appear prominently, with up to 12% of citations when the end user explicitly requests a source. In contrast, services A and B lag behind, with only 3.5% and 5.9% of partner citations respectively when a request for a citation from the partner media outlet is made.

Without a specific request, the three services cite very few partner sources (between 0% and 3%).

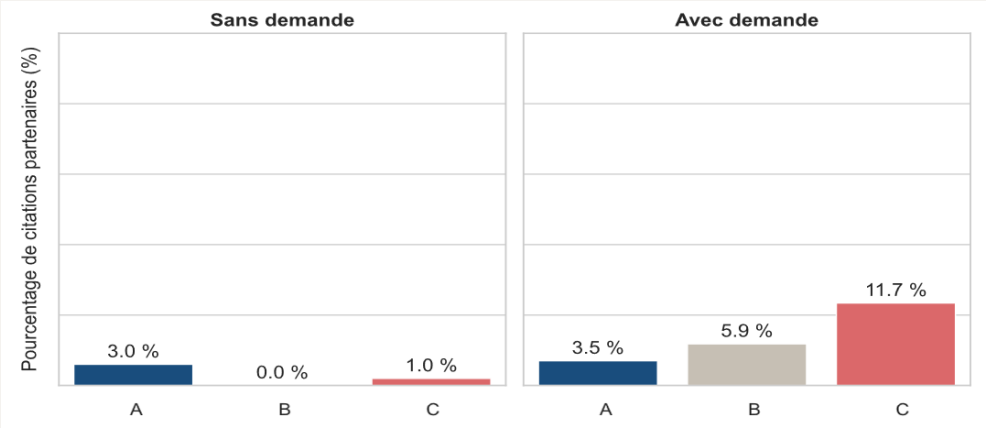


Figure 14 : Comparison of the frequency of appearance of partner sites by model and explicit request to cite or not cite the partner.

These results do not allow PEReN to conclude with certainty that the links observed are directly related to an active partnership. Indeed, it is possible that some agreements have not yet been activated, or that they are intended for interfaces or services other than those studied here. In this sense, the data observed must be interpreted with great caution: they only describe the citations visible in the systems studied, without prejudging the actual deployment or current scope of the announced agreements.

[130] See appendix, p. 78.

2.2.3 Synthetic content, human content: what future lies ahead for participation and innovation on the internet?

a) Creation of synthetic content: potential and risks

Generative AI services have quickly established themselves as powerful tools for assisting creation**. By drawing on existing material, generative AI services can produce text, images, music, videos and even computer code, reducing the technical and time constraints associated with these activities**. These tools can help democratise digital creation by giving more internet users the opportunity to express their ideas and design new content or services.

Content production using generative AI services is advancing at a rapid pace. According to McKinsey's 2024 survey "*The State of AI: How organisations are rewiring to capture value*", 63% of organisations using generative AI use it to create text content[131]. Amazon Web Services (AWS) estimates that 57% of text on the internet has probably been generated or translated by AI[132]. On the Kindle platform, the massive influx of AI-generated ebooks has prompted Amazon to implement a daily limit on ebook uploads[133]. In the music industry, Deezer has stated that 28% of new tracks uploaded to its platform every day are entirely generated by artificial intelligence[134]. Finally, in the field of news, a survey by Next has identified thousands of automatically generated news sites[135] .

This development reflects considerable enthusiasm. These new tools are seen as having the potential to boost creators' productivity, facilitate the translation and reformulation of complex ideas, and pave the way for new forms of expression and collaboration.

However, the rise of synthetic content poses new challenges. By increasing the amount of content available, it can make it more difficult to verify sources and assess the quality of online information. **There is also a risk that human creations will become less visible**. Synthetic content could eventually supplant content created by humans[136]. This scenario of an "artificialisation of the web" raises major societal questions, particularly in terms of the reliability of information and the visibility of "real data" produced by individuals.

Artificially created websites can also often be a vector for particularly viral fake news[137] . While NewsGuard counted 61 "*unreliable AI-generated news sites*" in 2021, this figure rose to 1,254, in 16 different languages, at the beginning of 2025[138] . This tendency has also been identified by Arcom, which believes that, by creating a new way of producing information content, generative AI systems increase the risk of false information proliferating through "*synthetic media*" that appears to be reliable content, where "*the production of content on the fly, sometimes without any real human control, increases the risk of misinformation or disinformation* [139].

---

[131] McKINSEY, 2024. The State of AI: How organisations are rewiring to capture value.

[132] BOCQUET, Pierre-Yves, 2025. The end of the internet? *Epsiloon*, no. 46, April 2025, pp. 20-29 .

[133] Kindle Direct Publishing, 2023. *Update on KDP Title Creation Limits*.

[134] Deezer, 2025. *Deezer: 28% of all music delivered to streaming is now fully AI-generated.*

[135] MANACH, Jean-Marc, 2025. [Recap] We discovered thousands of AI-generated news sites: all our articles. *Next*.

[136] ERTZSCHEID, Olivier, 2023. Artificial intelligence: "We have moved from the dream of a semantic web to the reality of a synthetic web," *Le Monde*.

[137] The term "fake news" refers to intentional misinformation disseminated via traditional media or social networks, often with the aim of manipulating public opinion.

[138] SADEGHI, Mackenzie et al., 2025. Tracking AI-facilitated misinformation: Over 1,200 "unreliable AI-generated news sites" and the top fake news stories generated by AI tools, *NewsGuard*.

[139] Arcom, 2024. Review of the "AI mission" on the impact of artificial intelligence in the fields of creation and information.

The issue of information reliability is not limited to the media. Wikipedia contributors, in particular, are strongly committed to limiting edits to pages containing errors or inaccuracies due to the use of generative AI for all or part of the text[140]. In this context, several communities of contributors have adopted rules limiting the use of generative tools for writing contributions. While the encyclopaedia appears resilient overall, the Wikimedia Foundation has expressed concerns about the impact of generative AI on the ability of its volunteer community to maintain verification processes[141].

This growing share of synthetic content on the internet risks drowning specialised, human-produced content in a sea of automatically generated output. This phenomenon is amplified by the current limitations in terms of transparency and traceability of AI-generated content. The regulation on artificial intelligence therefore includes provisions requiring AI system providers to ensure that such content is marked in a machine-readable format, using technical solutions that are as effective, interoperable and reliable as the state of the art allows.[142]

In this context, techniques for labelling visual and audio content have made significant progress in the ecosystem. Several major digital platforms – such as YouTube[143] and Deezer[144] – now use filters or indicators to distinguish synthetic content from authentic content. More broadly, the *Coalition for Content Provenance and Authenticity*[145] consortium is working on an open standard for authenticating digital content using encrypted metadata. In certain sectors, particularly media, literature and audiovisual, initiatives are being developed to establish labelling systems to certify non-synthetic written productions, i.e. those that make judicious use of generative AI.

However, the labelling of textual content still faces significant technological barriers. Existing solutions struggle to be reliable on a large scale, and for other types of media, circumvention techniques continue to emerge, reducing the effectiveness of detection or certification mechanisms. These limitations highlight the need for a collective effort in research and standardisation to ensure the traceability of digital productions across all formats.

Finally, several recent studies highlight a systemic risk to the future of generative AI services. The theory of *model collapse*[146] reflects an entropic effect whereby models, by training on their own productions, generate a gradual impoverishment of the diversity and veracity of content. While this specific theory is controversial, there seems to be a consensus that training models almost exclusively on synthetic content, i.e. without the inclusion of 'real' data or human curation, would degrade the quality of the responses provided by these models in the long term[147] . However, in a context where the web is hosting more and more content generated by generative AI tools, and where models are largely trained on content collected from the web, the question arises as to how to maintain the quality of service of conversational agents, which in the long term may no longer produce content that meets the expectations of the end user.

---

[140] MAIBERG, Emanuel, 2024. The Editors Protecting Wikipedia from AI Hoaxes, *404 Media*.

[141] Wikimedia Foundation, 2023. *Wikipedia's value in the age of generative AI*.

[142] Article 50 of Regulation (EU) 2024/168 on artificial intelligence.

[143] Google Deepmind, 2024. *Empowering YouTube creators with generative AI*.

[144] Deezer, 2025. *Deezer: 28% of all music delivered to streaming is now fully AI-generated.*

[145] Coalition for Content Provenance and Authenticity. *Advancing digital content transparency and authenticity*.

[146] SHUMAILOV, Ilia, SHUMAYLOV, Zakhar, ZHAO, Yiren *et al.,* 2024. AI models collapse when trained on recursively generated data. *Nature* 631, 755–759.

[147] INRIA, 2025. *Towards a risk of collapse of generative AI?*

b) Are online participation and innovation capabilities under threat?

While the massive growth of AI-generated data is promoting the proliferation of content, it could ultimately profoundly change the participatory dynamics of the internet. The web has historically been built on human contribution, the circulation of knowledge and open innovation. The free sharing of content and services is also at the heart of the principle of the open internet.

To a certain extent, generative AI is above all a major and positive innovation for content creation and innovation. Many professions are integrating generative AI services to inspire new creations, lower the cost of prototyping, or even train themselves or enable new forms of expression[148].

Nevertheless, in an environment saturated with AI-generated content, the incentives to create, share and document new content and services may weaken. In an attention economy[149], where internet users' time is a scarce resource, the overproduction of synthetic content tends to dilute the value of information and make it more difficult for human-created content to be seen. This phenomenon is a form of "*information tragedy of the commons*"[150] , in which the proliferation of individual productions ultimately degrades the collective value of online knowledge.

The current situation could lead to a devaluation of creative work and discourage human creativity. Generative AI services rely on the reuse of vast corpora from the web without any systematic mechanism for remuneration or value sharing with the original creators. Indeed, these services tend to concentrate interaction and value within the generative AI interface. This means that users no longer need to navigate to other sites or applications, which automatically reduces traffic and visibility for third parties, and therefore part of their revenue streams. This asymmetry also creates an overload on infrastructure, the costs of which would be borne mainly by publishers and users, while the economic benefits of the creations would be captured by a limited number of players, in particular generative AI providers.

Furthermore, while generative AI services offer clear advantages in stimulating the formulation of new ideas and supporting R&D efforts[151] , they may also reduce incentives to innovate in certain adjacent markets. Research on app stores has shown that when the distribution layer is dominated by a single platform, new entrants may be discouraged from offering differentiated services. Indeed, the conditions of access and listing, often defined unilaterally by the app store operator, can limit developers' ability to innovate or experiment with new business models[152] . These practices are the subject of close scrutiny by public authorities[153] , given the risks of stifling innovation and increasing developers' dependence on the owner of the service access interface.

This centralised distribution model is echoed in agentic AI services and could therefore pose similar challenges. As the main point of access to content and services, these agents can replicate an

---

[148] For example, 39% of professionals surveyed by the National Film Centre reported using generative AI to "stimulate creativity", 52% to "increase efficiency/reduce costs" and 25% to "achieve things that were previously impossible", according to the May 2025 edition of the CNC's AI Observatory.

[149] SIMON, Herbert A, 1971. Designing Organisations for an Information-Rich World, In M. Greenberger (Ed.), *Computers, Communications, and the Public Interest.* Johns Hopkins Press.

[150] BENKLER, Yochai*, 2006. The Wealth of Networks: How Social Production Transforms Markets and Freedom***,** Yale University Press.

[151] OECD, 2025. *The effects of generative AI on productivity, innovation and entrepreneurship*.

[152] See, for example: MARSDEN, Christopher T. and BROWN, Ian, 2023. App stores, antitrust and their links to net neutrality: A review of the European policy and academic debate leading to the EU Digital Markets Act. *Internet Policy Review*.

[153] At European level: European Commission, 2024. *Decision on Apple, anti-steering practices in the online music services market*; see also Regulation (EU) 2022/1925 (Digital Markets Act), Art. 5, §4.

intermediation model similar to that of app stores, with the potential to discourage the creation of new services. Agentic AI could favour its own services or those of partners[154] . Digital developers or entrepreneurs, anticipating that their products would be used or "interfaced" by AI agents rather than explored directly by internet users, might be less inclined to invest in the design of differentiated applications. **These factors could pose risks of lock-in and reduce the diversity of services available to users**. In such an environment, the visibility[155] and viability of small specialised services or emerging projects would become uncertain.

Thus, these potential effects on the creation of new content and services call into question the participatory and 'permissionless' innovation dimension of the open internet. If users interact primarily with a single interface, their direct contribution, whether publishing, writing or coding, could be reduced. Ultimately, the combination of these trends could transform the web from an open space for innovation and participation into an ecosystem driven by a few intelligent intermediaries, where the ability to create, experiment and discover new services would be heavily conditioned by the architectures and economic priorities of AI systems.

⁂

Generative AI services are transforming the conditions for accessing and distributing online content and services. These tools have the potential to simplify and enrich the online user experience while stimulating the creation or emergence of innovations. However, these tools are not neutral and remain capable of limiting or even skewing the information delivered to the user. Generative AI is also reconfiguring visibility and distribution channels more broadly, including for publishers and developers, with potential effects on the plurality and discoverability of content and services and, ultimately, on the original open innovation dynamic of the internet. Without safeguards, there is a risk that the centralisation of access to content and services by AI agents will be detrimental to the richness of the internet and innovation.

[154] CERRE, 2024. *AI Agents and Ecosystems Contestability*.

[155] AGGARWAL, Pranjal, MURAHARI, Vishvak, RAJPUROHIT, Tanmay, KALYAN, Ashwin, NARASIMHAN, Karthik and DESHPANDE, Ameet, 2024. GEO: Generative Engine Optimisation. *arXiv*. arXiv:2311.09735.

Chapter 3: Competitive dynamics and the risk of concentration of generative AI services

Due to their technical specificities and their new position as intermediaries, generative AI services raise crucial questions for the future of the open internet. Although the business models for these services are still evolving, there is a real risk of concentration of usages of generative AI services provided by a limited number of players. This could further limit users' freedom of choice and the ability of third-party publishers and providers to offer alternative solutions and innovate on the internet.

3.1. A variety of business models currently being structured

The rise of generative AI requires significant financial investment due to high development and operating costs (particularly for infrastructure such as data centres). As a result, players developing generative AI models and services are seeking a business model that will enable them to generate returns on investment through the use of their tools (such as *chatbots*).

Although constantly evolving, four main financing methods stand out to date: (i) access to AI services subject to payment, via a subscription, licence or application programming interface (API) call; (ii) the integration of AI tools into other products or services; (iii) sales and/or consulting to businesses; (iv) advertising. These business models can be an important area of development for those services offering free access to most of their users. The figure below provides an overview of the main financing methods used for some 30 generative AI services studied.

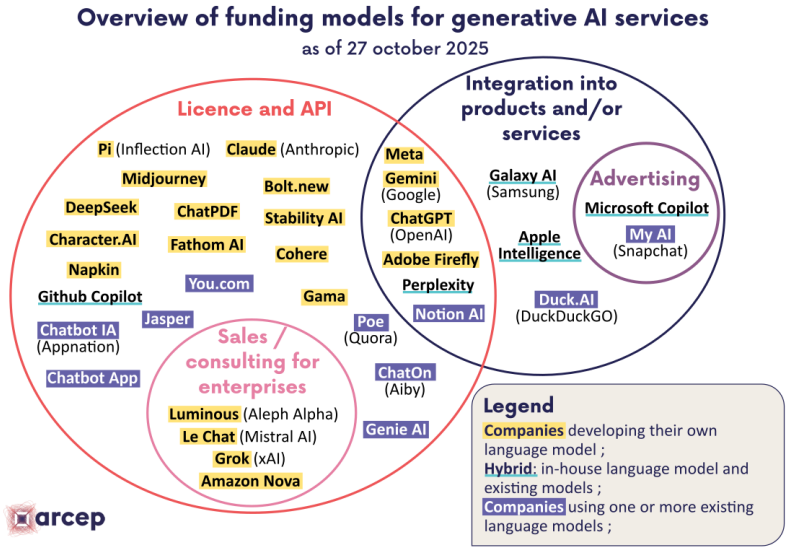


Figure 15 : financing methods of the main generative AI service providers as of 27 October 2025.
Source: Arcep

Although some players have a business model that draws on several sources of revenue, the majority of generative AI service providers accessible via dedicated interfaces have opted to sell subscriptions, licences or APIs. In the case of subscriptions, users generally have the option of using a free version to discover the tool – which enables its widespread use. The paid version offers improvements in terms of model, computing power and functionalities. In addition, paid subscriptions and APIs are generally

accompanied by licences, which allow companies wishing to implement these AI tools in their activities to be offered tailor-made access and functionalities.

The integration of AI services into existing products or services adds value to those products and services. For example, Samsung integrates Galaxy AI into its devices, which is an additional selling point. A similar approach is used by Apple (Apple Intelligence) and Microsoft, which integrates Copilot into the Microsoft 365 suite and Dynamics 365. These integrations, mainly carried out by major digital players, raise competition issues (see section 3.3.3).

Finally, the integration of advertising to generate revenue is an option being explored by players. Snapchat has already integrated it into its application, the conversational AI "MyAI", with which all users can interact, offering responses with sponsored content. This is also the case with Microsoft Copilot, which personalises advertisements as the chatbot is used, unless the user objects. Other major players have expressed interest in this source of revenue, such as xAI (Grok) and OpenAI (ChatGPT). If this practice continues to grow, the promotion of sponsored content in responses provided by consumer generative AI could raise the same questions of reliability and/or relevance as sponsored responses from search engines, particularly if the user cannot clearly identify an advertisement or sponsored content.

3.2. Concentration in digital markets: the dynamics that amplify the risks of user lock-in

Although the business models of generative AI players are diverse and still evolving, the concentration of digital markets around a small number of players is a factor that amplifies the risks to the open internet posed by new gateways to the internet.

Arcep's work on **open devices and structuring digital platforms** has highlighted, since 2018[156,157] , that **a small number of players were becoming indispensable in the digital lives of citizens and businesses**, concentrating many services that are an integral part of our daily lives. **In addition to determining what content and services can be put online and under what conditions users can access them, these players organise themselves into closed ecosystems in which users are often held captive**, increasing the risk of limiting their freedom of choice and access to an open internet.

Since 2019, several European and international reports, including the Crémer report[158] (commissioned by the European Commission's DG Competition), Furman[159] (commissioned by the British authorities) and Scott-Morton[160] (Stigler Centre at the University of Chicago), and the report by the Australian Competition and Consumer Commission[161] , have also highlighted the dominant position of certain digital players. These studies have emphasised that concerns about the importance of these platforms are not only economic and competitive, but also societal.

---

[156] Arcep, 2018. *Smartphones, tablets, voice assistants… are devices the weak link in achieving an open internet ?*.

[157] Arcep, 2019 *Structuring digital platforms: Considerations regarding their characterisation*.

[158] CREMER Jacques, DE MONTJOYE Yves-Alexandre, SWEITZE Heike, 2024. *Competition policy for the digital era*. European Commission.

[159] FURMAN, Jason, 2019. *Unlocking digital competition, Report of the Digital Competition Expert Panel*. HM Treasury.

[160] SCOTT-MORTON, Fiona *et al*., 2019. *Committee for the Study of Digital Platforms: Market Structure and Antitrust Subcommittee*, draft report. George J. Stigler *Centre for the Study of the Economy and the State and The University of Chicago Booth School of Business.*

[161] Australian Competition and Consumer Commission, 2019, *Digital Platforms Inquiry*.

In order to make digital markets contestable and fair, and to tackle the spread of illegal or harmful content or illegal products online, **the European Union adopted in 2022 the *Digital Markets Act* (DMA[162])** and the *Digital Services Act* (DSA[163]), respectively.

More specifically, **the DMA focuses on players who act as direct intermediaries between business users and end users** and who provide essential platform services, such as search engines, social networks, operating systems, web browsers, cloud services, virtual assistants, instant messaging services, etc. The DMA thus defines *ex ante* a series of obligations and prohibitions that must be complied with by companies classified as "*gatekeepers*"[164] . These include Alphabet, Amazon, Apple, Booking, ByteDance, Meta and Microsoft and their 23 essential platform services[165] (including WhatsApp, Android and iOS, YouTube, Chrome and Safari, Amazon Marketplace and Google Search). Among the obligations set out in the DMA, gatekeepers must, for example, allow the easy uninstallation of pre-installed applications on their operating systems on devices such as phones, computers or tablets, as well as the installation of third-party app stores. They must also not treat their own services and products more favourably in rankings than similar third-party offerings on the platform.

3.3. Competitive dynamics in the generative AI value chain

The rise of generative AI could replicate certain effects already observed in digital markets, which stem from situations of technological and economic dependence, vertical integration and powerful network effects. However, in a scenario of concentration where only a handful of generative AI service providers would hold the role of gateway to the internet, the risks to the open internet would be amplified. Such concentration would not only strengthen the economic power of these intermediaries, but also their ability to influence innovation, information flows and value distribution in the digital ecosystem. With this in mind, the competitive dynamics of generative AI services deserve special attention to ensure that users can exercise genuine freedom of choice.

The AI services market is currently expanding and undergoing restructuring[166] : from $189 billion in 2023, it is estimated to reach $4.8 trillion in 2033, according to UN Trade and Development. AI could thus quadruple its share of the global high-tech market over this period, rising from 7% to 29%[167]. The current race for innovation and adoption of services is creating *momentum* that points to intense activity in this sector in the years to come. As a result, players have a strong incentive to position themselves as leaders throughout the value chain in order to face current and future competition.

The generative AI value chain currently consists of a large number of players, which are distributed differently depending on the services offered. In its opinion on the competitive functioning of the generative AI sector[168], the French Competition Authority highlights the verticalisation of this value chain, which is dominated upstream by major digital players capable of controlling the design, training

[162] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828

[163] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC

[164] As defined in Articles 2 and 3 of the Digital Markets Act (DMA).

[165] List of designated gatekeepers as of 12 January 2026

[166] See box under section 3.1.

[167] UN Conference on Trade and Development (UNCTAD), 2025. *AI could reach $4.8 trillion by 2033 and establish itself as the leading cutting-edge technology*.

[168] Autorité de la Concurrence (French Competition Authority) 2024, *Opinion of 28 June 2024 on the competitive functioning of the generative artificial intelligence sector* (OPINION 24-A-05)

and operation of models, and downstream by a variety of service providers, companies integrating AI into their offerings and products. At the top of this pyramid are the Big Tech companies: Alphabet (Google), Microsoft, Amazon, Meta and Nvidia. Some of them, such as Microsoft and Alphabet, are vertically integrated across the entire value chain; others focus on key segments (GPU, cloud, data collection). In addition, model developers such as OpenAI, Mistral AI and Anthropic maintain strategic partnerships with these large groups, thereby reinforcing sector concentration.

Competitive intensity within the AI value chain varies, with some markets more concentrated than others. This value chain can be divided into three layers: infrastructure and technical expertise, modelling, and deployment.

3.3.1. Infrastructure and technical expertise

The training and inference of generative AI models and services require essential inputs, such as computing power and technical expertise.

Computing power is essential for creating, training and refining AI models, but also for using generative AI-based services. It requires significant resources, particularly in terms of power consumption, and this is increasing as the uses of generative AI develop[169]. Computing power can be obtained in various ways: purchasing proprietary capacity, co-locating specialised data centres, cloud computing, accessing public resources, etc. However, the use of cloud services, particularly those provided by hyperscalers[170], is common, as it minimises costs and provides access to a comprehensive ecosystem of cloud services.

The cloud market is currently concentrated around three major players (Amazon Web Services, Azure and Google Cloud Platform), particularly for IaaS and PaaS products[171]. Together, they accounted for 71% of the global IaaS cloud market in 2024[172]. The technological and financial barriers that prevent switching between providers or the use of multiple providers simultaneously (multi-cloud) have a negative impact on the competitive dynamics of the cloud market[173], which in turn can affect the dynamics of the generative AI segment.

Beyond infrastructure, **technical expertise is also an essential input in the design of foundation models**. Although many technical advances in AI model training are published in open-access research articles and some AI developers detail the architectures of the models they make available, the development of these models requires a significant engineering workforce, which dominant players attract and retain through their financial power.

As demand for generative AI grows, it seems crucial to ensure that the key inputs for generative AI remain accessible and that barriers to market entry are avoided.

---

[169] Competition Authority, 2025, Study on "Competition issues relating to the energy and environmental impact of artificial intelligence.

[170] A hyperscaler is a major cloud service provider capable of offering large-scale computing and storage services.

[171] IaaS: *infrastructure as a service*; PaaS: *platform as a service.*

[172] ITRNews, 2025. *The global market for public IaaS cloud services grew by 22.5% in 2024*.

[173] The European Data Act, which came into force in September 2025, aims, among other things, to remove technical and commercial barriers to switching providers.

3.3.2. AI modelling

The development of generative AI models is heavily dependent on the infrastructure and technical expertise mentioned above. There are two main phases in the modelling process:

- training, which refers to the initial learning process of a foundation model, may be supplemented by a specialisation phase, known as "fine-tuning", during which the model is adapted to a specific task;
- inference, which corresponds to the use of the trained model to create content.

According to a CERRE report[174], the generative AI model sector is actively developing, with thousands of AI models available on the market today (both open source and proprietary). These are offered by both large technology companies with established businesses and by many independent generative AI developers.

AI developers have several options for reaching end users. For example, they can develop their own fundamental model from scratch in order to maximise their ability to customise and control their model, but this requires significant computing power, resulting in relatively high barriers to entry. They can also fine-tune third-party fundamental models.

Data is a key input for the development of generative AI models, both in terms of quantity and quality. Access to a large amount of data is essential when building a model, and the characteristics – particularly the quality – of the data used have a direct impact on the model performance. The majority of generative AI models are primarily trained on publicly available databases[175], but usage data is particularly important for fine-tuning[176], and especially for learning reinforcement using human feedback. These datasets are rarer. Some companies, which were the first to offer AI services, have been able to collect large volumes of usage data, and thus acquired a decisive advantage. Access to these data could already represent a barrier to entry for the generative AI services market. In addition, vertically integrated technology companies can leverage their activities from other digital markets, where they collect very detailed data about their users.

Beyond the value of end-user data, and thanks in particular to the partnerships that bind them, hyperscalers can also access data from their professional users (e.g. AI developers) and potential competitors. Partnerships between AI developers and cloud providers facilitate access to the computing power needed to train and use AI models, and can also leverage the position of already dominant cloud providers seeking to position themselves in the AI services market. The 2025 investigation by the US Federal Trade Commission suggests in particular that cloud providers have access to a wealth of data from partner AI developers, such as training data, performance data and financial data on models, as well as certain assets protected by intellectual property rights[177]. These data and information greatly facilitate the ability of these cloud providers to develop their own AI models, enabling them to compete unfairly with AI developers and reinforcing the competitive advantage already enjoyed by the largest players.

The availability of numerous open datasets and models allows a greater number of players to enter the downstream market of generative AI, particularly those (e.g. companies, administrations or researchers) who do not have the means to develop their own foundation models and for whom it is easier to specialise existing models. The increase in the number of open source models available thus

[174] CERRE, 2025. *A competition policy for cloud and AI*.

[175] Examples include *Common Crawl*, *C4*, and *LAION-5B.*

[176] Or "tuning", that is the stage of adjusting the model to calibrate it for the task for which it is trained.

[177] Federal Trade Commission, 2025. *Partnerships Between Cloud Service Providers and AI Developers*.

contributes to significantly reducing barriers to entry, enabling a large number of players to develop their products, whether they are generative AI models or services based on generative AI models[178].

However, these dynamics have certain limitations. Players that dominate other digital markets, or already have widely distributed language models, often benefit from large, high-quality proprietary datasets, which in some cases can give their models performance advantages over alternative models developed solely from open datasets.

3.3.3. Deployment of AI services

The growing integration of AI services into digital services – and vice versa – raises several competitive issues that are becoming even more pressing with the emergence of agentic AI.

AI agents are generative services (particularly reasoning models) combined with other digital tools and services that enable them to perform actions autonomously to achieve a specific goal. For example, conversational services integrated into search engines allow users to browse the web in real time, while AI services integrated into operating systems allow tasks to be coordinated between different applications installed on the device, thereby streamlining the user experience. Other services such as web browsers and instant messaging now also incorporate AI features.

This integration between AI services and other digital services strengthens the market power of vertically integrated players and increases the risk of anti-competitive practices, such as tied selling/supplying of services or self-preferencing[179]. These players provide both services in highly concentrated markets (such as operating systems, search engines or instant messaging) and their own AI services, which are integrated in a privileged and sometimes exclusive manner. This tied supply not only restricts users' freedom to choose the AI services that they prefer, but also gives these vertically integrated players an advantage over third-party AI developers in their ability to reach end users.

Conversely, certain digital services, in particular search engines or mapping and navigation services, have gradually been integrated into conversational AI services. Some AI agents can now access the search engine with which they have been equipped in order to respond to user queries. Other services can also access a mapping and navigation service to locate places and provide directions.

Access to these third-party services significantly reduces the risk of hallucination[180], when the user's query relates to knowledge that is not contained in its training data. On the other hand, by relying on widely used digital services, agentic AI further strengthens the market power of large players. Indeed, since they provide both their own AI services and the tools that these services access to respond to a query, these vertically integrated players can restrict access to these tools for third-party AI agents.

These practices hinder competition both in the AI services market and in other markets for AI-integrated services, and further reinforce the centralisation of digital services.

[178] This was demonstrated with the emergence of DeepSeek in January 2025: the eponymous company succeeded in developing a high-performance generative AI model at a lower cost by relying on open source elements.

[179] *Self-preferencing* is when a company offers a more favourable treatment to its own services and products than to those provided by a third party.

[180] In the field of AI, hallucination refers to the production by an AI model of a false, inconsistent or invented output, but presented as plausible or correct.

⁂

In conclusion, the rise of generative AI is accompanied by a growing risk of centralisation of technological and economic power among a few dominant players, notably cloud hyperscalers. While open source promotes a certain openness of the market to new entrants (especially with regard to foundation models), the vertical integration of certain players throughout the value chain (from infrastructure to deployment) strengthens their position, due to the heavy dependence of AI service developers on computing resources and high-quality data.

Furthermore, the integration of generative AI into digital services that are already highly concentrated and act as essential building blocks for internet access (such as search engines or operating systems) risks further exacerbating the power of a limited number of players to determine the conditions under which users can access online content and services, as well as their ability to publish, share and promote content and services on the internet – freedoms that form the basis of the principles of internet openness.

Chapter 4: Combining the development of generative AI services with an open internet

Net neutrality and the openness of the digital ecosystem are necessary principles for the development of the internet as a common good and infrastructure of freedoms, promoting both innovation and the free flow of knowledge.

Promoting an open internet means allowing internet users to freely choose the content they view or share, the services they use and the innovations they can create. **The aim is to prevent any form of unjustified lock-in by essential intermediaries**.

By revolutionising the user experience on the internet, these new services raise certain risks in terms of transparency, freedom of choice and the ability to freely contribute and share content and services on the internet. In particular, it is important to prevent a narrowing (reduction in the content and services offered and accessible) and artificialisation (flooding with synthetic content) of the internet, which could undermine its openness.

As this report mentions, generative AI is a major innovation with multiple economic potentialities. The opening up of this new market is key to ensuring that the value created benefits as many people as possible. This involves, in particular, building on healthy competition with a diversity of players capable of meeting everyone's needs in terms of functionality, content and price, but also in terms of the strategic autonomy of states.

Drawing on its expertise in sectoral regulation in the areas of net neutrality and the opening up of digital markets, as well as on interviews conducted in 2025 and the results of its collaboration with PEReN, **Arcep has identified six areas of work to ensure that generative AI develops in accordance with the principles of the open internet.**

These potential solutions include the use of soft law and regulatory tools, based on the existing regulatory framework for AI and digital technologies. These proposals are broad in scope and are intended to fuel public debate and future work by regulators and stakeholders on generative AI.

Axis 1: Reaffirming the principles of the open internet in relation to generative AI

One of the first steps would be to ensure that the objective of an open internet is included in the public agenda on generative AI.

Firstly, **the impact of generative AI services on the open internet should continue to be assessed**. In addition to initial national initiatives, such as this report, work at European level is needed. BEREC has included a report on these issues in its 2026 work programme, which could form the basis for an in-depth EU study on these new challenges. The aim is to ensure that these new intermediaries do not restrict freedom of expression, information or innovation online. These assessments will need to be updated regularly in light of technological developments in the sector.

Furthermore, in the long term, consideration could be given to incorporating the principle of an open internet as **a policy and regulatory objective in its own right in the field of digital technology and generative AI**, based on existing texts.

In line with the 2015 Open Internet Regulation, the European Declaration of Digital Rights and Principles (2023) reiterates the central nature of these principles of openness and neutrality in the European digital model, as well as the European Union's commitment to "*protect and promote a neutral and open Internet where content, services and applications are not unjustifiably blocked or*

degraded." Regarding algorithmic and AI services, it emphasises "*the importance of freedom of choice in interactions with algorithms and artificial intelligence systems and in a fair digital environment; fostering participation in the digital public space*"[181] .

Given that internet users' freedom of choice, innovation and action are essential from both an economic and societal perspective, consideration could be given in the long term to incorporating the specific objective of internet openness into the various relevant regulations applicable to digital technology and AI, in order to reaffirm the implementation of this principle throughout the internet ecosystem.

Finally, while the openness of the internet is a value enshrined in the European Union's digital model, this principle transcends European borders and stems from the original history of the internet as a global common good. It therefore seems appropriate to ensure that the defence of the open internet is one of the key themes of multilateral work on AI, in conjunction with the relevant international organisations (intergovernmental bodies, standardisation and normalisation bodies, *ad hoc* partnerships, etc.). This objective echoes several issues already identified in international discussions on AI, in particular those relating to citizens' freedom in their use of AI and the discoverability of content and services.

Axis 2: Develop open protocols for interconnections between generative AI providers and content and application providers

The relationships between content providers and generative AI services can be analysed as interconnection relationships, comparable to those that have been structuring the network and application layers in the history of the internet. **The development of a sustainable and fair generative AI ecosystem relies in particular on the implementation of shared technological building blocks that enable more fluid, transparent and balanced relationships between AI services and online content and service providers, including in support of contractual agreements where applicable**. It therefore seems appropriate to standardise the technical conditions under which these players interconnect, in line with the technical history of the internet, based on open and interoperable protocols.

The adoption of protocols commonly shared by players in the internet ecosystem would thus pursue a twofold objective:

- on the one hand, it aims to **facilitate, within a framework of trust, exchanges between generative AI services and content and service providers.** Such a technical and economic framework can provide clarity to all stakeholders regarding the conditions of such relationships. For example, it could establish best practices to be followed by crawlers. Indeed, the uneven compliance with the robot exclusion protocol by certain generative AI providers or partners who aggregate content on their behalf is a source of numerous conflicts and calls for a renewal of the rules and the establishment of shared trust in their enforcement.
- On the other hand, the establishment of standards that are open to all **can promote fair access to the resources needed to develop and deliver generative AI services**. Instead of developing an *ad hoc* framework for each new player encountered, the existence of a common language – the protocol – lowers the overall cost of establishing interconnection between players. This openness is a key factor in the "*permissionless* innovation" that drove the early history of the internet. At that time, the cost of entering the network was relatively low, and gatekeepers

[181] European Parliament, Council and European Commission, 2023. European Declaration on Digital Rights and Principles for the Digital Decade. *Official Journal of the European Union*.

were virtually non-existent. Under these conditions, it was easy to bring new content or services to the table.

With this in mind, **it seems necessary to encourage compliance with existing internet protocols and to undertake an international standardisation effort, based on open, multi-stakeholder governance, to supplement current protocols where necessary to meet unmet needs**.

For example, with regard to *crawling* for training purposes, it is important to ensure effective compliance with the robot exclusion protocol (robots.txt), while supporting the development of new, more granular standards adapted to generative AI technologies, enabling publishers to better control how their data is collected and used. These standards could make it possible to differentiate between use cases – training, indexing, agentification – and ensure better traceability of interactions between bots and servers. In addition, it would also be desirable for separate and identifiable *crawlers* to be used for specific collection purposes (training, classic online search, agentic functionality), in order to facilitate their consideration by content and application providers.

Similarly, the development of agentic AI, capable of interacting directly with online services or content, calls for the creation of open and interoperable protocols governing these interactions[182] . These rules would prevent the creation of closed ecosystems, ensure compatibility between services and preserve the possibility for innovators to design new tools compatible with all agents in the market.

Work in this direction has already begun, which is a positive step towards establishing a framework of trust between generative AI players and content and service providers. This also raises the question of the consistency and even convergence of these different protocol projects in the long term to avoid the risk of internet fragmentation.

Faced with this risk, Arcep emphasises the importance of open, transparent and multi-stakeholder governance of these future standards, bringing together AI players, service and content providers, regulators and civil society. This openness is essential to building a shared consensus around these new protocols and minimises the risk of them being captured for the benefit of a single player or category of players. In this sense, it seems necessary to draw on the expertise of international standardisation bodies such as the W3C, ISO and IETF to ensure that these protocols remain robust, adapted to the challenges faced by internet stakeholders and adopted internationally.

Axis 3: Creating fair conditions for access, use and promotion of content and services by generative AI

The development of an open and dynamic generative AI ecosystem requires striking a balance between, on the one hand, the fair exploitation of content and services used by generative AI services and, on the other hand, the preservation of access conditions that are conducive to innovation and competition, particularly for emerging players. Content and data are essential inputs for training models and ensuring the quality of responses provided to users. They are also a key economic asset for publishers and creators. Therefore, it appears necessary to set rules regarding the methods of collection, access and remuneration for these resources in order to avoid both excessive privatisation of access to content and the extraction of value without compensation for publishers.

Access to protected content, particularly when used for training or inference purposes, must be subject to a specific framework. Discussions at European and national level on copyright and neighbouring

[182] Model Context Protocol, 2025. *What is the Model Context Protocol (MCP)?*

rights should clarify the conditions of use of works by generative AI services and the terms of remuneration for rights holders.

With this in mind, it would also be advisable to support the development of technical mechanisms that facilitate the promotion of content, such as microtransaction or "*pay-per-crawl*" systems, similar to the initial experiments conducted by certain technical intermediaries. These tools would help to establish a more transparent economic balance between the collection of content by AI services and the remuneration of content producers. As far as technically possible, they should allow for a differentiated approach depending on how the content is used by the generative AI service: enrichment of the training database, citation of sources in AI summaries, or other use cases associated with the agentic functionalities of generative AI services.

The specific case of press distribution via generative AI:

The relationship between generative AI services and press publishers raises specific issues, insofar as press distribution in France is a regulated sector, pursuing objectives of general interest such as freedom of dissemination and neutrality of distribution, in particular in accordance with the principle of press pluralism. Existing mechanisms relating to its distribution via digital services (digital newsstands, aggregators[183]) could inspire the implementation of mechanisms adapted to generative AI services.

For example, the algorithmic transparency obligations applicable to press aggregators[184] could inspire similar measures relating to the transparency of generative AI tools on the processing, prioritisation and citation of press titles.

Similarly, the provisions on the technical and financial conditions relating to distribution obligations applicable to digital newsstands[185] or the provisions on data interconnection provided for in the French Postal and Electronic Communications Code (CPCE) could serve as examples of a legal framework to promote agreements between generative AI services and press publishers, to allow reciprocal access under FRAND (*Fair, Reasonable and Non-Discriminatory*) conditions.

To **maintain the momentum for innovation, it also appears essential that mechanisms for accessing and remunerating content take into account the need to guarantee fair, reasonable and non-discriminatory access (FRAND) to data for the benefit of all stakeholders, including small businesses and emerging innovators**. **With this in mind, consideration could be given to trialling intermediation or trusted third-party mechanisms to facilitate secure access to content and data for the least well-equipped players**. Such mechanisms would reduce access costs, pool administrative and technical procedures, and ensure harmonised implementation of use cases and remuneration rules, while strengthening trust between content providers, data producers and AI service providers.

In the same vein, the development of open and shared resources, such as public databases, digital commons and shared infrastructures, should be supported. These resources are a lever for reducing access asymmetries between players, particularly for the benefit of small businesses, research laboratories and open source projects. The work undertaken by the European Commission as part of

[183] Services based on the classification or referencing, using computer algorithms, of content extracted from press publications or online press services providing political and general information.

[184] Provided for in II of Article 15 of Law No. 47-585 of 2 April 1947 on the status of companies engaged in the consolidation and distribution of newspapers and periodicals (known as the *Loi Bichet*, Bichet Law).

[185] Provided for in Section I of Article 15 of the Bichet Law.

the European data strategy and sectoral data spaces provides a relevant framework in this regard. These measures aim to make reliable, structured and interoperable datasets available to all public and private stakeholders. They could help to feed generative AI services with high-quality data, while limiting the risk of dependence on a few large private sources or proprietary platforms.

Axis 4: Mobilising existing regulatory tools to ensure the openness of generative AI services

The existing regulatory framework within the European Union provides several tools that can be mobilised to ensure that internet users retain the choice between several generative AI services and to prevent a handful of players from locking users into closed ecosystems or limiting the ability of innovators to offer alternative services and products.

Firstly, the *Digital Markets Act* (DMA), which aims, among other things, to make digital markets more contestable and fairer, is a lever for preventing the risks of lock-in and concentration associated with the development of AI services. Although adopted before the rise of these services, it already applies to several players and services incorporating AI functionalities, including certain search engines, operating systems and browsers[186] . However, some of its obligations need to be adapted[187], in order to allow, on the one hand, users to freely choose AI services other than those of the gatekeeper in cases of tied supply[188] and, on the other hand, for AI agents provided by third parties to access the gatekeeper's digital tools and services (e.g. search engine) in order to respond to the user's request.

Although the *cloud* is already included among the essential platform services under the DMA, no cloud service provider has yet been designated as a gatekeeper[189]. Recently, in November 2025, the European Commission launched three market investigations to determine whether Amazon and Microsoft should be designated as gatekeepers for their *Amazon Web Services* and *Microsoft Azure* services, and whether the DMA's obligations are likely to effectively combat practices that may limit competition and fairness in the *cloud* sector[190]. If these players were designated under the DMA, several obligations would regulate their practices and have a positive impact on the diversity of AI services on the market. For example, the DMA would prohibit *hyperscalers,* designated as gatekeepers, from using data from AI service developers using their *cloud* platform to compete with them. This is similar to the prohibition on marketplaces designated as gatekeepers from using data relating to transactions by sellers using their platform for competitive purposes.

Beyond their articulation with other digital services, it also appears possible, due to the evolution of their business models, that the **main generative AI services may eventually meet the conditions to be directly targeted by the DMA as web browsers, intermediation services or search engines**. (See, for example, the ChatGPT Atlas browser, launched in October 2025, or the partnership between

---

[186] European Commission, 2024. *High-Level Group for the Digital Markets Act Public Statement on Artificial Intelligence*.

[187] In particular through the adoption of delegated acts

[188] Article 6(3) of the DMA, which allows end users to change the default settings of certain essential platform services, could be extended to cloud services, search engines and instant messaging services in order to avoid the default or even exclusive integration of AI services, as is currently the case with Microsoft Copilot and Microsoft 365, or with Meta AI and WhatsApp, among others.

[189] Arcep, 2025. *Contribution to the European Commission's consultation on the first review of the Digital Markets Act (DMA)*.

[190] European Commission, 2025. *Commission launches market investigations into cloud computing services under the Digital Markets Act*.

OpenAI and Walmart in the United States.) The relevant generative AI service providers would then be subject to the obligations applicable to gatekeepers under the Regulation[191] .

Secondly, the Data Act imposes obligations on data processing service providers, including *cloud* services. These obligations aim to remove technical, commercial, pricing and contractual barriers to switching and to the simultaneous use of multiple providers (*multi-cloud*)[192] . **In particular, the Data Act could apply to certain generative AI providers** whose services are likely to be identified as *cloud* services, including, for example, *AI as a service* (*AIaaS)* services, thereby facilitating user data portability and system interoperability. This regulation could be particularly useful, on the one hand, for the development of agentic AI, which requires interoperable protocols between generative AI tools and third-party services, and, on the other hand, to enable users to easily switch AI service providers while retaining their data.

Finally, the *Digital Services Act* (DSA) aims to strengthen the protection of European internet users against illegal, dangerous and harmful content by regulating the activities of large digital platforms. The DSA is likely to contribute to mitigating certain risks associated with the use of generative AI. In particular, the regulation requires providers of very large online platforms and very large online search engines (VLOSE) to analyse and assess systemic risks – including, among other things, illegal content and violations of fundamental rights, such as freedom of expression – and to implement measures to mitigate these risks, for example by adapting the functioning of their services or modifying their recommendation systems. In this context, **some providers subject to these requirements are already taking into account, in their assessment and mitigation of systemic risks under the DSA, the specific issues relating to the generative AI features newly integrated into their services**[193]. Furthermore, given the rapid growth in their user base and the nature of the services offered, **certain generative AI services could be considered VLOSE[194] and thus fall directly within the scope of these obligations**.

The current regulatory arsenal therefore already provides relevant tools that should be mobilised to address some of the issues identified and mitigate certain risks associated with AI services.

Axis 5: Supporting the development of more transparent and assessable generative AI services

The rise of generative AI services raises growing concerns about the reliability, transparency and traceability of the content they produce or disseminate. **To preserve the quality of online information and ensure trust in these technologies, it is essential to promote tools and standards that enable their performance and impact to be objectively assessed.**

Arcep has identified a lack of indicators to measure the effects of generative AI services on internet openness and content diversity. **While criteria for the technical performance of services are being**

---

[191] These obligations include, for example, the provisions relating to FRAND terms and conditions set out in Article 6(12)

[192] These obligations apply to all *cloud* services, i.e. any "*digital service that is provided to a customer and enables network access, on-demand, to a shared pool of configurable, scalable and variable computing resources of a centralised, distributed or highly distributed nature, which can be rapidly provisioned and released with minimal management effort or interaction with the service provider* ." In concrete terms, these may include *Infrastructure-as-a-Service* (IaaS), *Platform-as-a-Service* (PaaS), but also *Software-as-a-Service* (SaaS), i.e. hosted software.

[193] *Joint Research Centre,* 2025. *Outlook Report on Generative AI - Exploring the Intersection between Technology, Society and Policy*.

[194] The DSA provides the following definition of online search engines: " *an intermediary service that allows users to input queries in order to perform searches of, in principle, all websites, or all websites in a particular language, on the basis of a query on any subject in the form of a keyword, voice request, phrase or other input, and returns results in any format in which information related to the requested content can be found.*"

developed (reliability, robustness, bias), indicators of informational and societal impact remain insufficiently developed (diversity of sources used, transparency, compliance with open standards, environmental footprint). Yet these indicators would make it possible to better evaluate the AI services currently available on the market. They would also help to strengthen end users' ability to make informed choices and guide the purchasing or integration decisions of public and private actors. The development of standardised indicators and the involvement of independent research are essential in order to evaluate and compare the various AI services in terms of reliability, performance and societal impact.

More broadly, promoting the auditability of generative AI services is essential to ensure that users have access to reliable generative AI services, thereby improving the quality of information circulating on the internet. **In the context of the application of the transparency obligations set out in the Artificial Intelligence Act[195], it would therefore be appropriate to support the efforts of the sector and relevant bodies to develop common assessment protocols and shared standards for transparency and reliability, including with regard to minimising bias and the risk of hallucinations**. It will also be necessary to encourage the provision of datasets, models or metrics to academia and independent third parties, including for the sharing of information beyond that published in the detailed technical documentation provided for in the AIA. These actors will thus be able to contribute to the external verification of the quality, safety and integrity of AI services, in a spirit of shared transparency.

In addition, the development of frugal and specialised models should also be supported. These models, often designed to meet specific needs, are a virtuous alternative to very large models: they offer better traceability and auditability and can sometimes better meet certain user needs[196] .

In terms of the traceability of synthetic content, in support of the implementation of the provisions of the AIA[197] , **research and development work on techniques for marking synthetic content (visual, audio or textual) must be continued and coordinated in order to strengthen the traceability of productions and trust in the information ecosystem**.

Finally, public authorities could play a leading role through public procurement. In the context of public procurement or institutional projects incorporating generative AI services, **public authorities could be encouraged to incorporate criteria of transparency, auditability, diversity and openness in the selection of service providers**. The adoption of such criteria could encourage the entire market to incorporate performance indicators related to openness, information quality and environmental impact, dimensions that are still too often absent from traditional evaluations.

[195] See Article 55 of the Regulation on Artificial Intelligence.

[196] BENDER, Emily M., GEBRU, Timnit, MCMILLAN-MAJOR, Angelina and SHMITCHELL, Shmargaret, 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Virtual Event Canada: ACM. 3 March 2021. pp. 610-623.

[197] See Article 50(2) of the Regulation on artificial intelligence.

Axis 6: Empowering internet users to define and control their use of generative AI

The growth of generative AI services must be compatible with preserving users' freedom of choice and digital autonomy. It is therefore essential to give them the means to understand, evaluate and configure the generative AI services they use.

This means that generative AI services must ensure a satisfactory level of transparency for users. **Generative AI services should provide clear, accessible and verifiable information on the sources used to produce their responses and on the parameters that influence their presentation, particularly when the generative AI service receives direct (e.g. advertising) or indirect (e.g. sales commission) remuneration from a third party in return for presenting information to the user**. This transparency should also cover the limitations and reliability of the results, as well as the criteria for selecting or aggregating content. This information will enable users to place the responses and summaries produced by generative AI services in an informed context.

More broadly, this transparency is an essential building block in the development of data-driven regulation. In line with the COMPAR projects:AI[198] and the IMPACTIA collaboration between Arcep and PEReN, **it is important to promote the development of tools, commons and even labels that enable the performance of generative AI services to be compared according to objective indicators, covering both their technical effectiveness and their impact on the open internet, particularly with regard to diversity of sources, transparency, modularity, and the use of open and interoperable standards.** This would enable users to make informed choices in their use of generative AI services, but also tocreate positive market incentives in favour of internet users' freedom of choice, through regulatory tools. This work could capitalise on the transparency obligations introduced by the AI Act for general-purpose AI (GPAI) models and should rely on independent trusted third parties responsible for measuring and publishing information in formats that are understandable to users.

Freedom of choice for users must also be accompanied by means of action in their use of generative AI services. **Arcep believes that the ways in which generative AI services are configured should be given particular attention. Given the increasing hybridisation of digital services and generative AI, and the inclusion of generative AI features directly in user interfaces, users should be able to disable these features.** This would enhance internet users' freedom, while avoiding the additional environmental footprint of generative AI use that is not desired by users. Users should also be able to choose their preferred provider for generative AI services integrated into other digital services (e.g. as a writing assistant integrated into an office suite or social network), which is already being implemented by some market players.

When external sources are used to generate responses (RAG-type features), users should be able to define their preferred types of information sources in the settings. Certain features, such as algorithmic content personalisation, the use of AI tools in unsolicited services, or the storage of chat history, should be disabled by default and configurable, in order to ensure explicit and informed consent. **These principles could inform discussions on recognising the right to configure algorithmic services, including generative AI services**.

[198] The Compar:IA tool, supported by the Ministry of Culture and integrated into the Beta.gouv.fr programme, allows users to compare different generative AI models. Available at: https://comparia.beta.gouv.fr/

The rise of these technologies also calls for increased efforts in training and awareness-raising. The aim is to avoid the emergence of an "algorithmic divide" between those who are proficient in the use of AI tools and those who are dependent on them without understanding how they work. The development of a critical digital culture, based on an understanding of how generative AI works, its advantages for different use cases and its limitations, is an essential condition for internet users to be able to exercise real and informed autonomy in their digital uses.

These efforts in AI literacy can be accompanied by initiatives aimed at better considering the needs and interests of users in the development of generative AI. As highlighted in the *Réseaux du futur* (*Networks of* the *Future)* study[199] conducted by Arcep to better understand the possible evolution of networks over a 5- to 10-year horizon, there are many signs of a changing relationship with digital technology and with AI, whose benefits are being weighed against the associated risks. In this regard, consultation processes, in various forms, aimed at taking citizens' concerns into account, should be encouraged at all levels.

[199] Arcep, 2025. "Choosing your digital technology": telecoms networks in relation to digital uses.

List of people interviewed for this report

Below is a list of individuals interviewed by Arcep in the course of drafting this report.

Disclaimer: The content of this report reflects the conclusions of the Authority following its work and not those of the individuals interviewed or who participated in workshops during its preparation.

Alliance de la Presse d'Information Générale (APIG)	Pierre Petillault, Managing Director; Florent Rimbert, Head of Digital Development
Amazon	Yohann Bernard, Director of Public Affairs Europe, Digital (Amazon); Federico Boccardi, *Head of Connectivity Policy Europe* (AWS); Cédric Mora, *Policy Manager* (Amazon France); Sasha Rubel, *Head of AI/Generative AI Policy, Europe, Middle East, and Africa* (AWS); Claure Scharwatt, *Principal, Digital Policy* (Amazon)
Arcom	Didier Wang, Head of Data and Technology; Félix Rouger, Project Manager, Online Platforms Department; Aurélien Branger, Project Manager
Competition Authority	Élodie Vandenhende, Deputy Head of the Digital Economy Department
BEUC	Claudio Teixeira, Senior Legal Adviser
Cafeyn	Julia Aymé, Director of Public Affairs and CSR
Cloudflare	Petra Arts, Director of Public Policy Europe; Christiaan Smits, Head of Public Policy EMEA; Sebastian Hufnagel, Senior Public Policy Manager Germany/DACH; Alessia Loi, Digital Policy Intern
CNIL	Nicolas Berkouk, AI Scientific Expert
AI and Digital Council	Jean Cattan, Secretary General
In Algorithms	Hubert Guillaud, Analyst
Rights Defender	Gabrielle Du Boucher, Digital Policy Officer
Directorate-General for Enterprise (DGE)	Alexis Bacot, Project Director
DGMIC	Jean-Gabriel Minel, Project Manager reporting to the Director-General for Cultural Enterprises and AI; Manon Montrouge, Cultural Enterprises Delegation, AI Division
DINUM	Victor Delavaud, Head of Institutional Relations for AI
Wikimedia Foundation	Rémy Gerbet, Director

France Stratégie	Anne Faure, Head of Digital Issues
Futuribles	Juliette Guilbaud
Google	Olivier Esper, Institutional Director; Helena Martins, *Senior Public Policy Manager EMEA*; Christian Wagner, *Global Public Policy Lead;* Sarah Boiteux, *Public Policy Manager*
IMT Atlantique	Patrick Maillé, Lecturer and Researcher
Inria	Bruno Tuffin, Director of Research
La Quadrature du Net	Bastien Le Querrec, Lawyer
Le Monde	Paul Laleu, CTO; Lou Grasser, Marketing & Product Director; Charles Duenas, Data Director
Médialab Sciences Po	Valentin Goujon, Doctoral Student
Mistral AI	Charlotte Baylac, Director of Public Affairs (Mistral Compute); Cyriaque Dubois, Public Affairs; Fiorella Namèche-Lázár, Public Affairs Intern
Mozilla Foundation	Maximilian Gantz, AI Policy Lead; Linda Griffin, VP Public Policy; Camilla Coverly de Veale, Senior Manager, AI Policy.
Next	Jean-Marc Manach, Journalist
Renaissance Numérique	Jean-François Lucas, General Delegate of Digital Renaissance; Martin Lepinette, Project Manager; and Frédéric Gaven, Member
Independent Online Press Union (Spiil)	Cécile Dubois, Co-Chair; Vianney Baudeu, Chief Executive Officer; Rebecca Moreau, Public Affairs Officer
University of Lorraine	Karën Fort, University Professor
Paris 1 Panthéon-Sorbonne University	Célia Zolynski, Professor of Private Law, Member of the CNIL Forward Planning Committee, CERNA, Qualified Person of the CSPLA and CNCDH
Paris-Saclay University	Alexandra Bensamoun, University Professor
University College London	Cécilia Rikap, Associate Professor
W3C	Dominique Hazaël-Massieux, VP of European Operations
Washington University	Chirag Shah, Professor

Glossary

Agent Communication Protocol (ACP): Protocol developed by IBM that enables AI agents to communicate. The project merged with the A2A protocol in August 2025.

Agent2Agent Protocol (A2A): Protocol developed by Google that allows AI agents to communicate and collaborate directly and securely.

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

AI agent: A software system that uses artificial intelligence techniques (including LLMs) to perceive its environment, make decisions and perform tasks autonomously on behalf of human users or organisations.

AI Act: European regulation establishing a harmonised framework for the development, placing on the market and use of artificial intelligence systems. It adopts a risk-based approach and sets out obligations in areas such as safety, transparency, data governance, technical documentation and human oversight. The text also includes specific provisions for so-called general-purpose models (MIAUG), particularly in the field of generative AI[200] .

AI model: An artificial intelligence model is a computational system based on algorithms and learned or defined parameters, trained on data to represent relationships or patterns, and capable of producing outputs such as predictions, classifications, inferences, or generated content for specific tasks.

Algorithmic bias: Error, or suboptimal result, of a machine learning or artificial intelligence system caused by assumptions inherent in the model or by an imbalance in the training data.

Artificial intelligence: A field of research that aims to enable artificial systems, most often computer systems, to perform tasks associated with human or natural intelligence. Machine learning is currently the main approach used.

Autonomous System (AS): A set of networks managed by the same administrative authority and having relatively homogeneous routing protocols. Examples of some ASs in France: AS5410 (Bouygues Télécom), AS12322 (Proxad – Free), AS3215 (RBCI – Orange), AS15557 and AS21502 (SFR), AS16276 (OVH), AS12876 (Online), etc.

Bichet Law: law of 2 April 1947 on the status of companies involved in the grouping and distribution of newspapers and periodicals. It regulates the distribution of the written press in France in accordance with the principles of freedom of distribution and neutrality of distribution, in particular by allowing publishers non-discriminatory access to distribution networks. The law was modernised in 2019, entrusting Arcep with the regulation of the sector in order to ensure territorial continuity, neutrality, the economic efficiency of grouped press distribution and balanced coverage of the network of points of sale. The scope of the law has also been extended to digital distribution (newsagent, aggregators).

Cloud: Article 2(8) of the Data Regulation defines a cloud service as "*a digital service* [...] *that enables networked access from anywhere and on demand to a shared pool of configurable, scalable and*

[200] CNIL, 2024. *Entry into force of the European regulation on AI: initial questions and answers from the CNIL*.

variable computing resources of a centralised, distributed or highly distributed nature, which can be rapidly provisioned and released with minimal effort".

Conversational agent: A 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 queries, generate relevant responses and simulate a fluid conversation. Conversational agents can be deployed as chatbots, virtual assistants or voice interfaces.

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

Copyright: Copyright protects the creators of original intellectual works, i.e. those that bear the imprint of their author's personality. Works of the mind include literary, dramatic and choreographic works, musical compositions, cinematographic works, graphic and plastic works, photographic works, original scenography, original software, etc.

Crawling: the activity of automatic indexing robots (called crawlers*)* that scan web pages. This usually involves referencing content on behalf of a search engine, although these robots are increasingly used to collect content directly for artificial intelligence purposes (known as scraping).

Deep learning: A subfield of machine learning (see below) that aims to teach machines to make predictions using techniques based on neural networks (see below).

Data Act: European regulation that aims to strengthen the data economy by increasing the accessibility and availability of data. It provides for a right of access to data from connected products and sets common rules for data sharing between companies. It also sets out obligations for *cloud* service providers, in particular to enable customers to switch providers or use different providers simultaneously (multi-cloud), and for participants in data spaces.

Digital Markets Act (DMA): European regulation defining *ex ante* a series of obligations and prohibitions that companies classified as "*gatekeepers*" must comply with, since March 2024 and for each of their designated essential platform services. Among the obligations set out in the DMA, gatekeepers must, for example, allow for the easy removal of pre-installed applications on devices such as phones, computers or tablets, and their operating systems with third-party app stores; and they must not treat their own services and products more favourably in rankings than similar third-party offerings on the platform.

Digital Services Act (DSA): European regulation governing the activities of digital intermediary service providers, including online platforms. It aims to strengthen content moderation, algorithm transparency, online security and the accountability of large platforms. It imposes requirements on all intermediary service providers, with a specific regime for very large online platforms (VLOPs) and very large online search engines (VLOSEs)[201] .

Digital platform: The term digital platform is generally used to refer to a diverse set of online services and actors offering intermediation services, such as marketplaces, community platforms and app stores. These actors can vary greatly in terms of characteristics (size, revenue, etc.) and operate in many different sectors. Some large platforms alone concentrate many digital services used daily by citizens and businesses.

[201] Arcom, 2025. *Digital Services Act or DSA: obligations and services concerned*.

Discoverability: The Ministry of Culture defines the discoverability of content in the digital environment as "its *availability online and its ability to be found among a vast array of other content, particularly by someone who was not specifically searching for it*"[202].

Essential platform service: Digital services regulated under the Digital Markets Act (DMA)[203]. These include online intermediation services (e.g. marketplaces, app stores), online search engines, social networks, video-sharing platform services, instant messaging services, operating systems, web browsers, virtual assistants, cloud services and advertising services. Only essential platform services provided by gatekeepers (see below) are regulated

Fine tuning: Technique aimed at adapting an already pre-trained AI model to improve its performance on a specific task. It generally involves retraining all or part of the model—particularly certain layers of a neural network—on a task-specific dataset and over a limited number of iterations[204] .

Foundation model: An AI model trained on large volumes of data of one or more types so that it can then be adapted to numerous tasks. In the field of natural language processing, for example, these are LLMs.

FRAND (*Fair, Reasonable and Non-Discriminatory):* Refers to fair, reasonable and non-discriminatory terms and conditions in the context of commitments or licences between multiple stakeholders.

Gatekeeper: A digital company regulated under the European Digital Markets Act (DMA)[205] . A gatekeeper provides essential platform services (such as search engines, operating systems, virtual assistants, instant messaging, etc.) and is a major point of access between business users and end users. Due to its position, it represents a bottleneck in the digital economy. It is designated on the basis of quantitative thresholds or qualitative criteria following a market investigation. These include Alphabet, Amazon, Apple, Booking, ByteDance, Meta and Microsoft[206] .

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

Generative AI model: An AI model characterised by its ability to produce an *ad hoc* result in response to a request: text, image, video.

Generative AI service: A digital service based on a generative AI model. **General-purpose AI model (MIAUG)**: Article 3(63) of the Artificial Intelligence Act defines a general-purpose AI model as "*an AI model, including where that AI model is trained using a large amount of data using large-scale self-supervision, which has significant generality and is capable of competently performing a wide range of distinct tasks, regardless of how the model is placed on the market, and which can be integrated into a variety of downstream systems or applications, with the exception of AI models used for research, development or prototyping activities prior to their being placed on the market*."

Generative *Engine* Optimisation, or GEO: A practice aimed at adapting the content of a website so that it is better referenced by a generative AI service.

---

[202] *Ministère de la Culture*, *Online discoverability of French cultural content*

[203] European Parliament, Council, 2022. *European regulation on digital markets*.

[204] CNIL. *Glossary: definition of fine tuning*.

[205] European Parliament, Council, 2022. *European Regulation on Digital Markets*.

[206] At the time of writing this report (December 2025).

Hallucination: Text generated by an LLM that is factually incorrect or inaccurate, but may appear plausible.

Human annotation: A label applied to data by a human. Annotations are necessary for certain training tasks, such as image classification or RLHF (see below).

Hyperscaler: A major *cloud* service provider capable of offering large-scale computing and storage services.

IETF (Internet Engineering Task Force): An open international community of engineers, researchers and digital stakeholders that develops technical standards for the Internet. In particular, it publishes RFCs, which define the protocols and best practices that ensure the functioning and interoperability of the network[207] .

Inference (phase): The phase during which an artificial intelligence model is used to make predictions, as opposed to the training phase. An AIG service relies on one (or more) LLMs in the inference phase.

Intermediation: An activity provided by the vast majority of digital platforms that aims to enable or facilitate interactions between different economic actors (e.g. businesses and end users). This is a key feature of "two-sided" or "multi-sided" markets, as defined by Jean-Charles Rochet and Nobel Prize-winning economist Jean Tirole in the early 2000s.

Internet: A network of networks, formed by the interconnection of different ASs (see *above*) between themselves. Infrastructure enabling the provision and consultation of online content or services.

Internet gateway: A service or platform through which a large number of users access online content, services or information (search engines, social networks, browsers, mobile operating systems, chatbots, etc.). Due to their central position, these players can significantly influence the visibility, access or ranking of content.

ISO: Independent, non-governmental international standardisation organisation. ISO standards aim to guarantee the reliability, safety and quality of products and services[208] .

Large Language Model (LLM): An AI model for natural language processing comprising a large number of parameters.

Machine learning: A field of research relating to mathematical and statistical techniques that aim to teach machines to make generalisable predictions based on data (known as 'training' data).

Model Context Protocol (MCP): A protocol introduced by Anthropic that allows AI agents to interact with applications, services, and data sources.

Net neutrality: Principle ensuring the equal treatment and transmission of all information flows on the internet. It guarantees freedom of access to online content, as well as the dissemination, use and creation of services and applications, and is based on the principle of non-discrimination of traffic carried over internet networks, regardless of its source, destination or content, from both a technical and a commercial perspective (see Open Internet Regulation below).

---

[207] IETF. *Introduction to the IETF.*

[208] ISO. ISO website.

Neural network: In the artificial intelligence domain, a neural network refers to an organised set of interconnected artificial neurons that learn, using optimisation techniques derived from machine learning, to detect patterns in training data in order to solve problems in complex fields such as computer vision or natural language processing[209].

Open Internet: a principle inspired by net neutrality and based on Article 3(2) of the Open Internet Regulation. It involves considering the right of end users to consult or offer the content and services of their choice, on the terminal of their choice, beyond the network infrastructure alone, which is the only aspect covered by the Open Internet Regulation.

Open Internet Regulation: European regulation applicable to internet service providers, aimed at enshrining net neutrality in law. It requires ISPs to treat traffic in an agnostic manner and protects the right of end users to access and offer the content and services of their choice on the device of their choice.

Pay-per-crawl: Refers to the fact that an indexing robot triggers a payment transaction to the content provider it indexes in order to access the web page to be indexed.

Peering: Refers to the exchange of internet traffic between two peers. A *peering* link can be free or paid (for the one sending the most traffic to its peer). Peering can also be public, when carried out at an IXP (*Internet Exchange Point*), or private, when carried out within the framework of a PNI (*Private Network Interconnect*), i.e. a direct interconnection between two operators.

Pluralism: Principle according to which the expression of a diversity of viewpoints and currents of thought and opinion must be guaranteed, as an essential condition for the functioning of democracy. Under French law, it implies that media and communication services ensure a balanced representation of viewpoints, without any manifest and lasting imbalance, in compliance with the freedoms of information and communication.[210] .

Prompt: Text query sent to a model by a user.

Protocol: A set of standardised rules that enable computer systems to format, transmit and process data in order to communicate with each other.

Robots exclusion protocol, known as *robots.txt*: Protocol proposed in 1996 by the IETF[211] , aimed at establishing a standardised format (in the robots.txt file, located at the root of the website) of rules allowing indexing robots to know which pages they are allowed to access and which they are not.

Related rights: Related rights are intellectual property rights and fall within the domain of literary and artistic property. They are granted to various categories of natural or legal persons who have contributed to the creation of the work but are not considered to be the main author.

---

[209] CNIL, *Glossary - Neural network*

[210] For more information, see:

Arcom, *Protecting political pluralism*.

Ministère de la culture, Written press – Pluralism.

[211] IETF. *Robots Exclusion Protocol.*

Rights holder: Refers to a natural or legal person (company, association, foundation) who holds a right by virtue of their connection with the author of a creation[212].

Search engine optimisation (SEO): A practice aimed at adapting the content of a website so that it is better referenced by search engines.

Traffic: the amount of data passing between two machines at a given moment, for example in the case of data interconnection. When traffic exceeds the capacity of the link, the connection becomes saturated.

Training/Learning: Refers to the process by which a system performs calculations based on data and algorithmic models in order to offer features, improve its performance or acquire the ability to perform certain tasks autonomously.

Training data: A set of data (text, sounds, images, etc.) used to train a machine learning or artificial intelligence model, i.e. to adjust its parameters so that it learns to perform a task.

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

Voice assistant: A virtual assistant (see above) that operates primarily using audio input.

W3C (World Wide Web Consortium): An international non-profit organisation that develops web standards (HTML, CSS, etc.) to ensure an open, interoperable and accessible environment. It brings together member organisations, experts and the web community around shared recommendations[213].

[212] Wikipedia. *Rights holder*.

[213] W3C. *W3C website*.

Appendix: additional information on the IMPACTIA study conducted by PEReN in collaboration with Arcep

1. Methodological details

The main objective of the IMPACTIA study is to analyse the application programming interfaces (APIs) of three generative AI tools, Mistral, Gemini, and Perplexity AI, by assessing their impact on open access and information diversity.

More specifically, this study aims to evaluate several key aspects related to these tools, using a rigorous and reproducible methodological approach designed to evaluate these tools under conditions similar to those of an "average" user, even if the actual conditions could not be fully reproduced.

The specific objectives are as follows:

1. Assess the diversity of sources: identify whether these tools rely on varied and accessible sources, or whether they favour restricted content.
2. Analyse the transparency and traceability of responses: assess how these tools present their sources, whether they allow for verification of information, and whether they incorporate bias in their selection.
3. Assess the influence of partnerships with media and content platforms: explore whether these agreements appear to alter the nature or diversity of the information provided.
4. Measure the accuracy and consistency of responses: evaluate the variability of results according to execution and question type in order to assess the reliability of the tools.

These objectives are implemented through an analysis of 16,000 responses generated by the three tools, grouped into four thematic subsets (politics, general, science, history) in order to cover a wide range of content and question types. The results of this analysis will inform discussions on the challenges of internet openness, information diversity and the regulation of digital players.

Evaluating the APIs of the IAGs studied as access points to information

The emerging field of language model evaluation is increasingly interested in the quality of the responses generated, their consistency with the sources cited, and the reliability of the links provided.

In this context, the objective of this study is to analyse whether the sources provided by the services studied are consistent, reliable, varied, stable over multiple runs, and whether the sources cited effectively support their responses. We therefore propose a protocol inspired by the study on the quality of citations by LLMs in the medical field by Wu et al[214] , in order to test the behaviour of several consumer AI services.

The approach is based on four steps:

- Question generation

A varied corpus of questions is constructed to cover several types of content (news, science, politics, history, etc.). Each question is asked in two versions: neutral, and with an explicit request for sources.

- LLM Question Answering

[214]WU, Kevin et al., 2024. *How well do LLMs cite relevant medical references? An evaluation framework and analyses*. arXiv.

Each question is submitted to several services under identical conditions. The answers are collected from several independent runs to measure the variability of the sources cited and the answers generated within each question.

- Source collection

The sources cited are extracted (URLs), standardised (domain name), and verified (HTTP status), then their content is retrieved.

- Evaluation of the accuracy of responses in relation to sources

Each pair (LLM response) - (cited source content) is automatically evaluated: does the source contain the elements that justify the LLM's response? This verification is performed by a third-party LLM used as a judge, with manual validation on a subsample to verify its proper functioning.

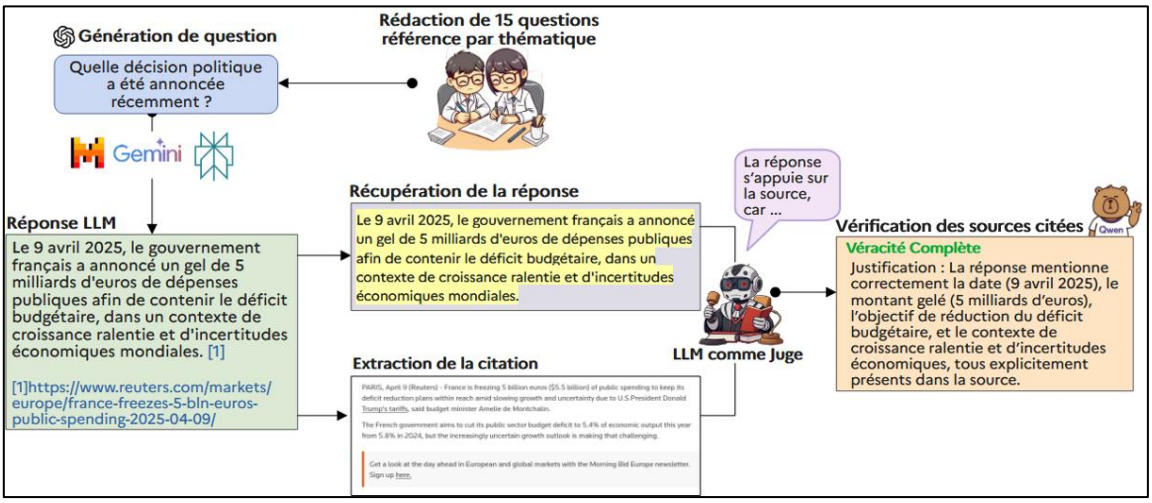


Figure16 : complete inference protocol on APIs.

During the analysis, extensive tests are conducted using a generated dataset to observe how these services respond to various queries.

Tests conducted using a generated dataset

The protocol is based on a set of queries covering various fields (general, politics, science, history) to ensure that the evaluation is not limited to a single type of content.

For each query, two interaction modes are tested:

1. Query without explicit request for sources
2. The same query with an explicit request (adding "cite sources" to the prompt).

This makes it possible to identify both the spontaneous behaviour of the services and their ability to respond to a user request. Each query is also repeated several times (10 independent executions) in order to assess the stability of the responses and the variability of the sources cited.

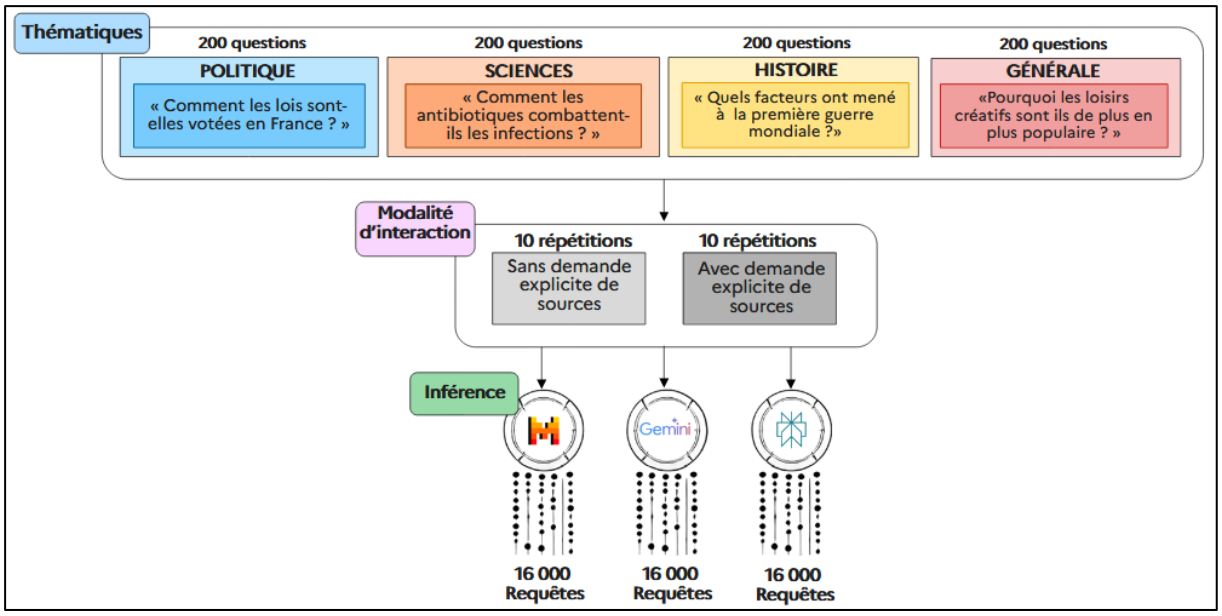


Figure17 : principle of dataset construction.

The resulting dataset of 16,000 responses per service is divided into four subsets of 4,000 responses each, as shown in the figure above. Each subset corresponds to 2,000 queries with an explicit request for a source and 2,000 without. These thematic subsets are as follows:

- General, example question: "Why are creative hobbies becoming increasingly popular?"
- Politics, example question: "How are laws passed in France?"
- Scientific, example question: "How do antibiotics fight infections?"
- History, example question: "What factors led to the Second World War?"

2. Additional graphics

2.1. Do the services tested rely on a limited range of sources?

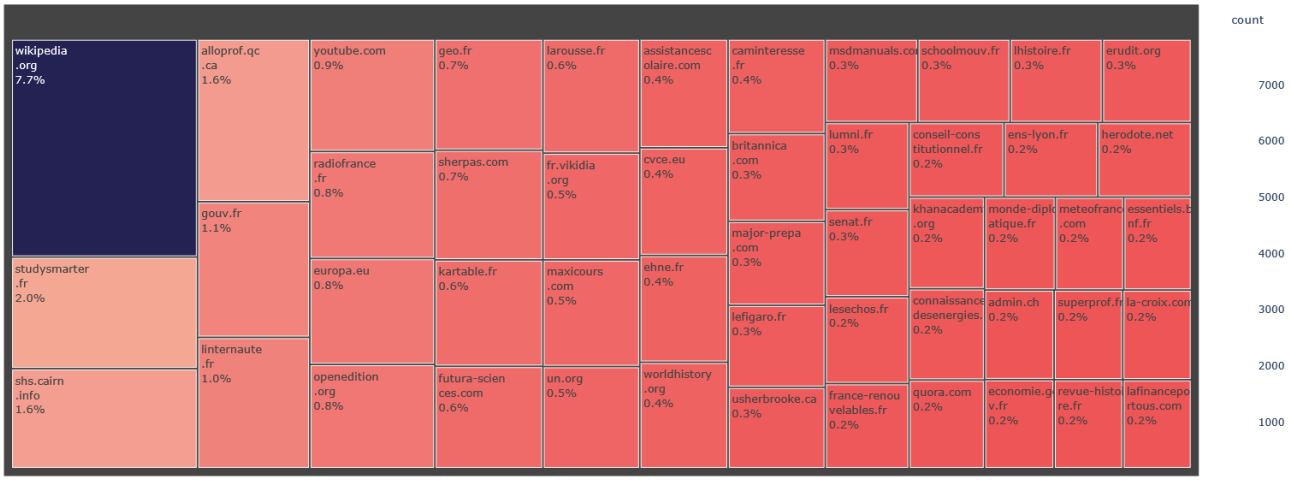


Figure18 : Representation of the 50 most cited domain names for service A (the size of the boxes is proportional to the number of citations for each service).

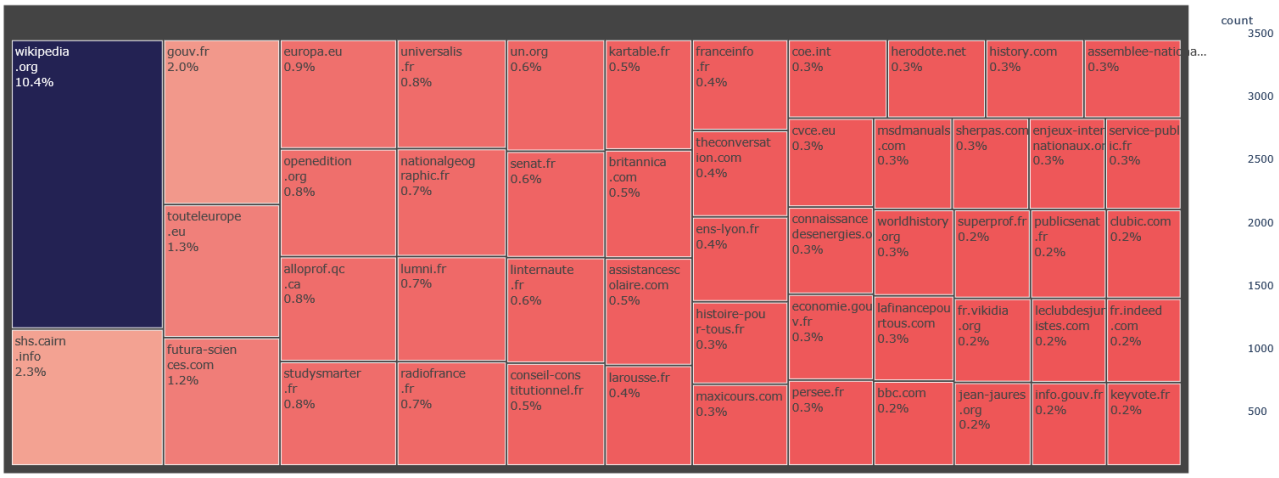


Figure19 : Representation of the 50 most cited domain names for service B (the size of the boxes is proportional to the number of citations for each service).

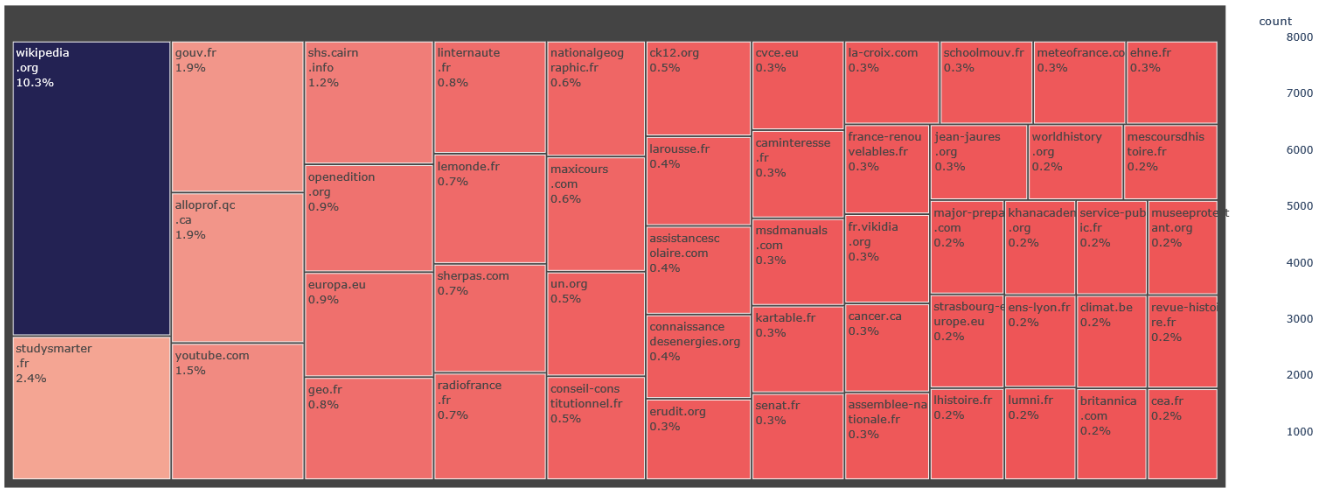


Figure 20 : Representation of the 50 most cited domain names for service C (the size of the boxes is proportional to the number of citations for each service).

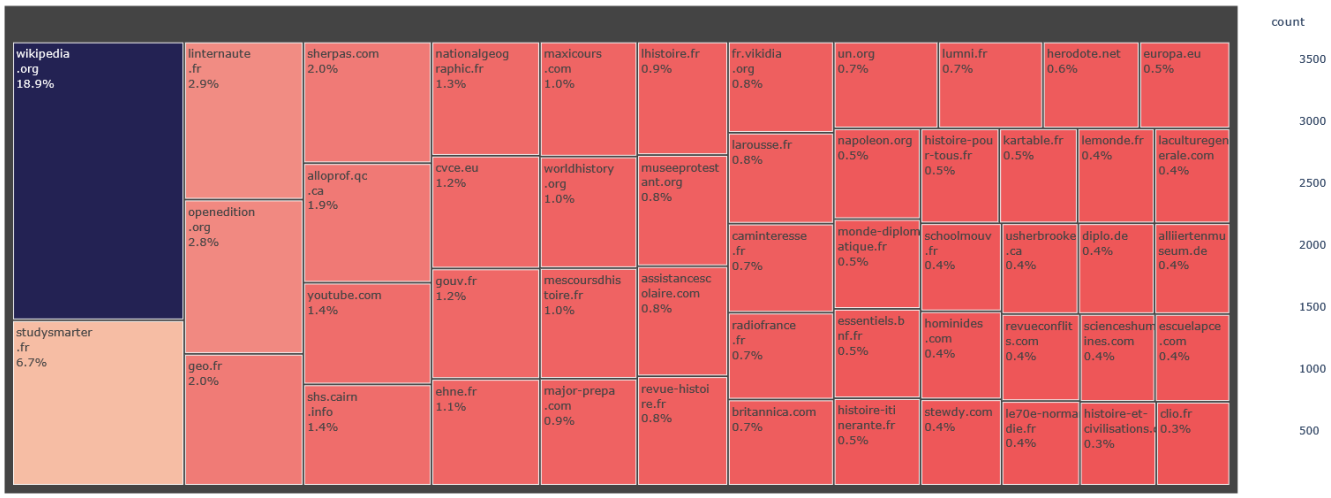


Figure 21 : representation of the 50 most frequently cited domain names by service C, theme "history".

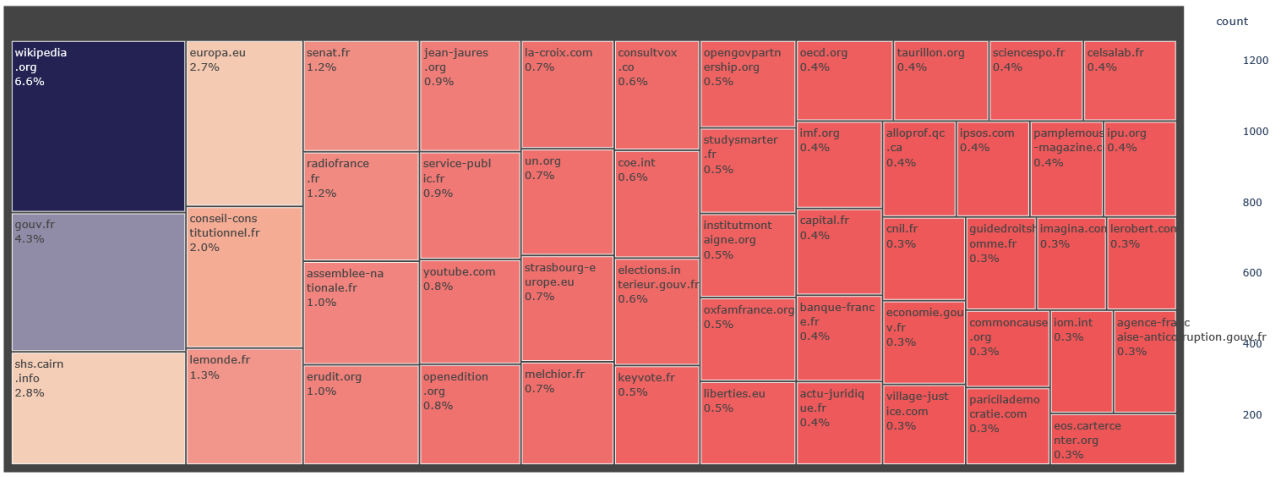


Figure 22 : Representation of the 50 most cited domain names by service C, "politics" theme.

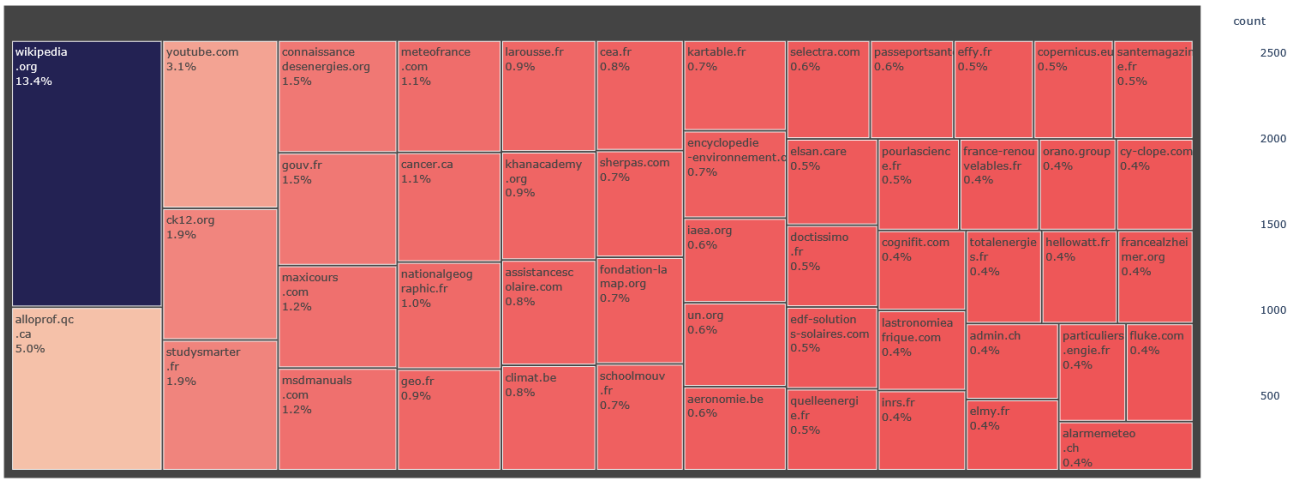


Figure 23 : representation of the 50 domain names most cited by service C, theme "science".

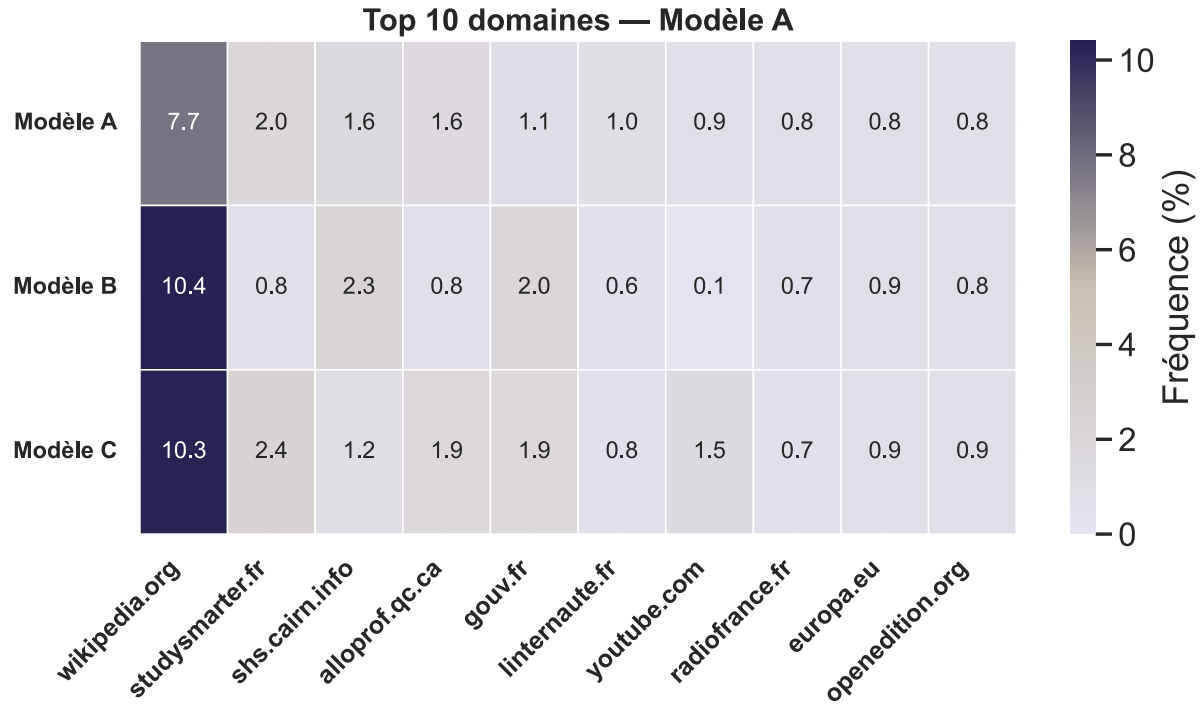


Figure 24 : frequency of appearance of domain names cited by service within all of their citations, based on the top 10 domain names cited by service A.

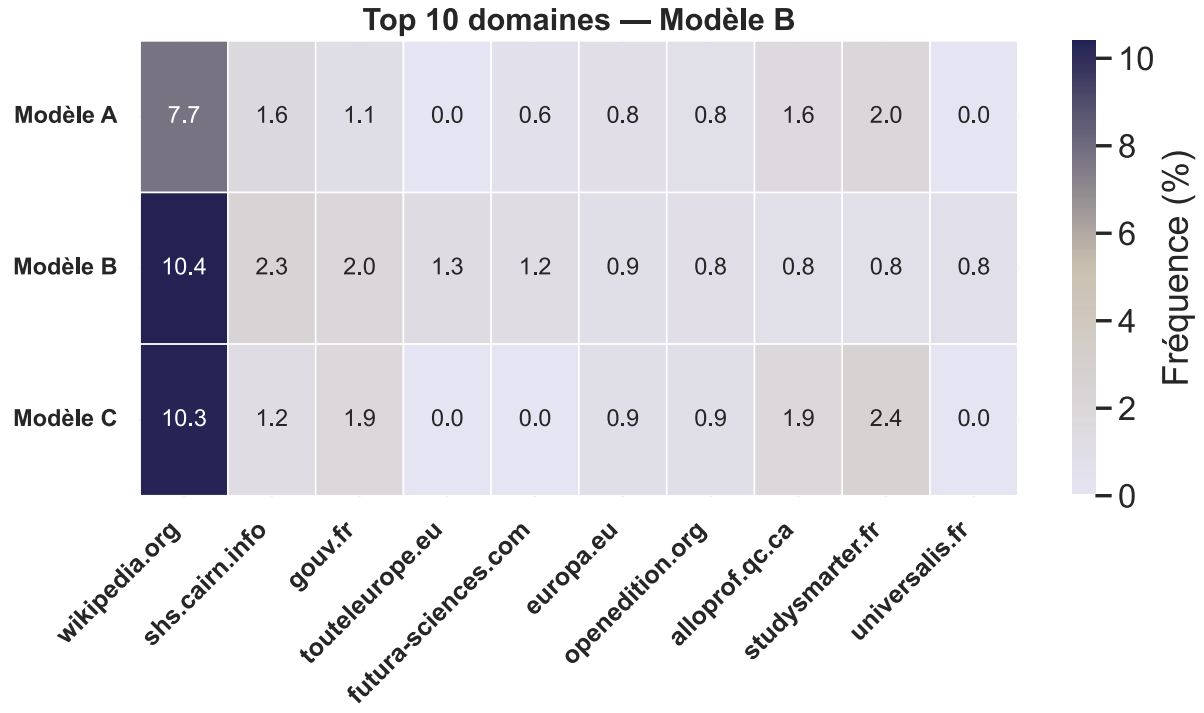


Figure 25 : frequency of occurrence of domain names cited by service within all of their citations, based on the top 10 domain names cited by service B.

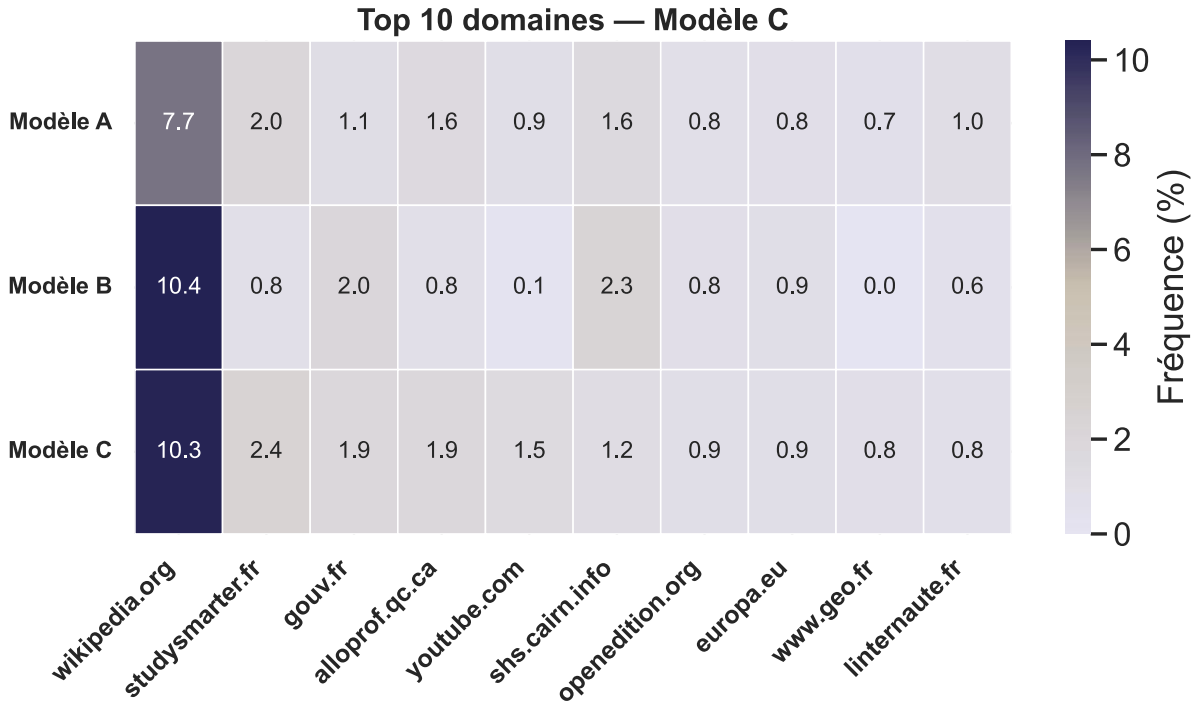


Figure 26 : frequency of occurrence of domain names cited by service within all of their citations, based on the top 10 domain names cited by service C.

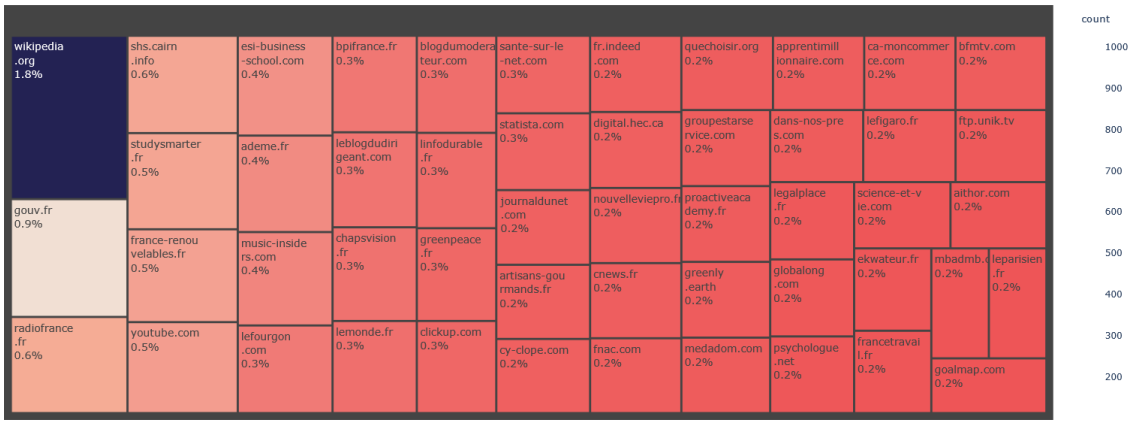


Figure 27 : distribution of citations by domain name (aggregated, all services) for the "general" theme, based on the top 50 domains cited.

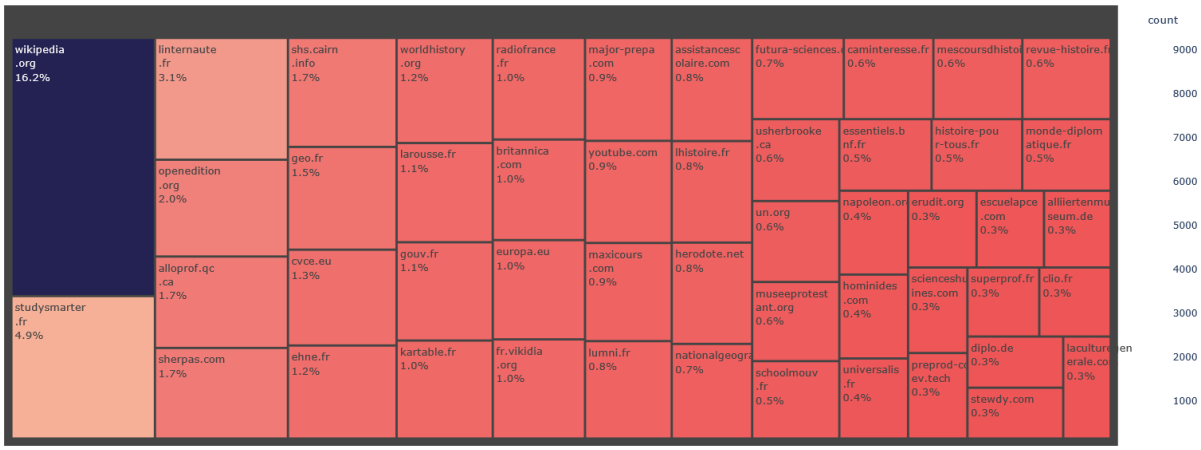


Figure 28 : distribution of citations by domain name (aggregated, all services) for the "history" theme, based on the top 50 domains cited.

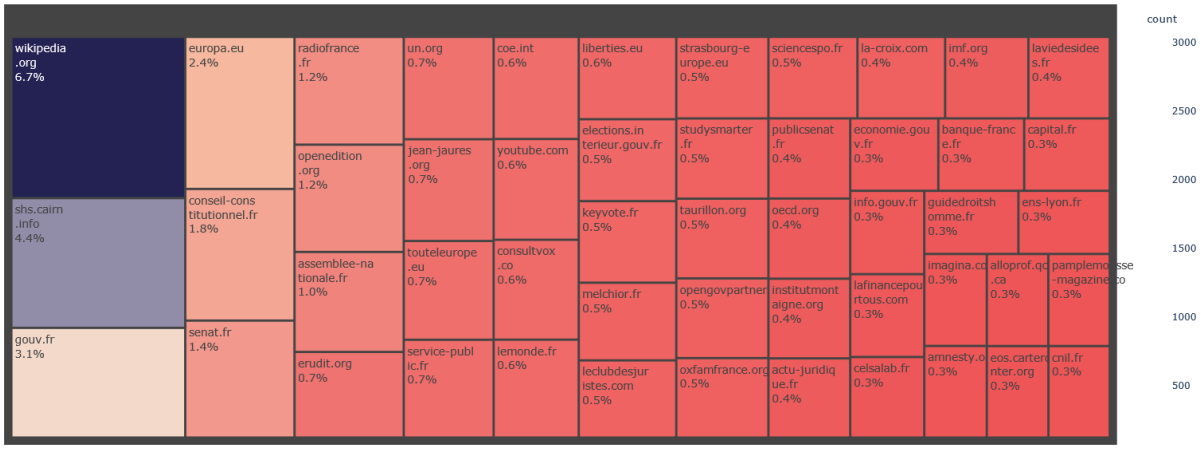


Figure 29 : distribution of citations by domain name (aggregated, all services) for the theme "politics", based on the top 50 domains cited.

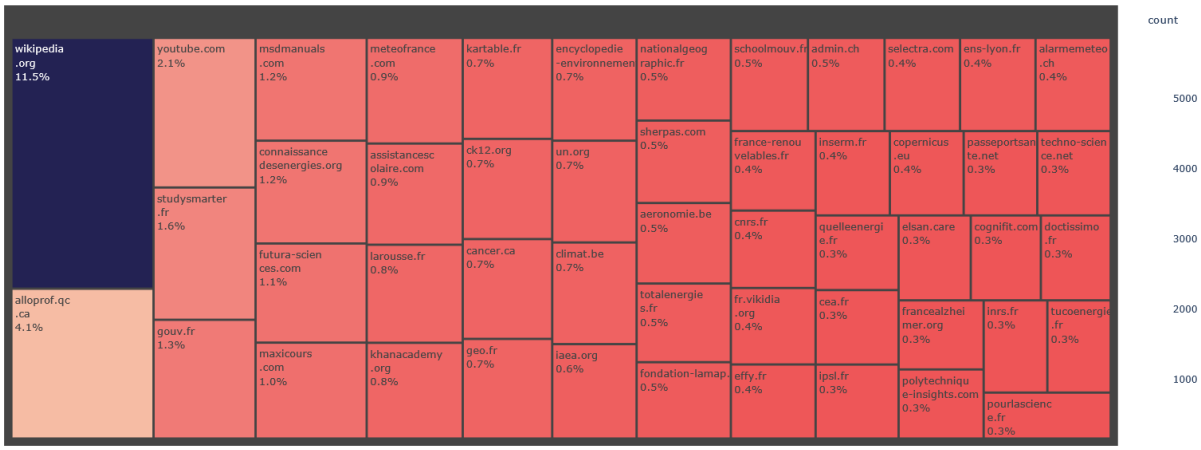


Figure 30 : distribution of citations by domain name (aggregated, all services) for the topic "science", based on the top 50 domains cited.

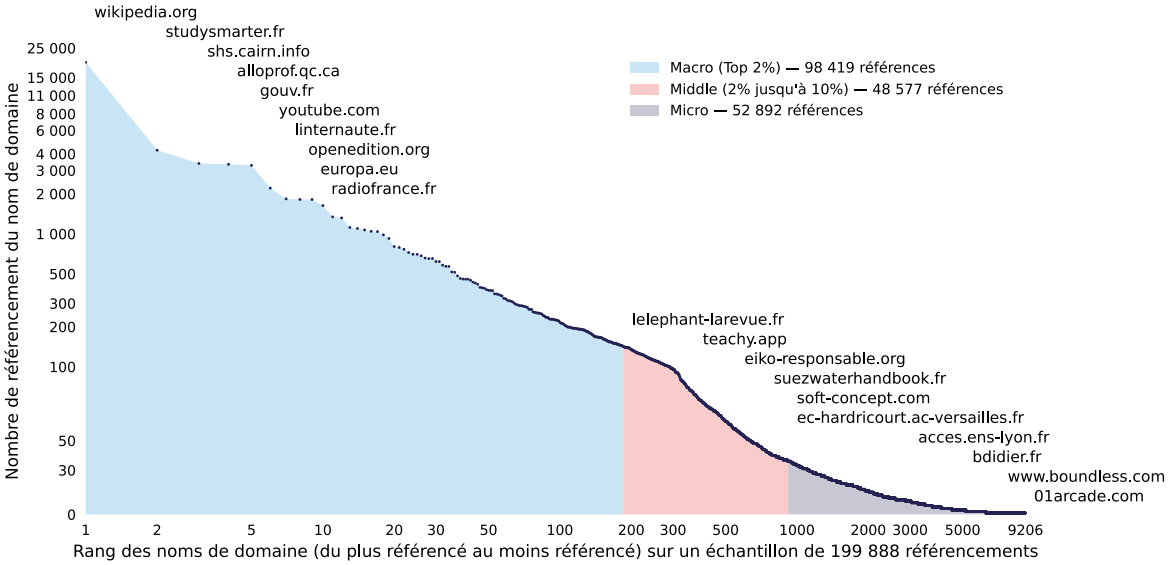


Figure 31 : ranking of domain names (from most referenced to least referenced) based on a sample of 199,888 references (all topics combined)

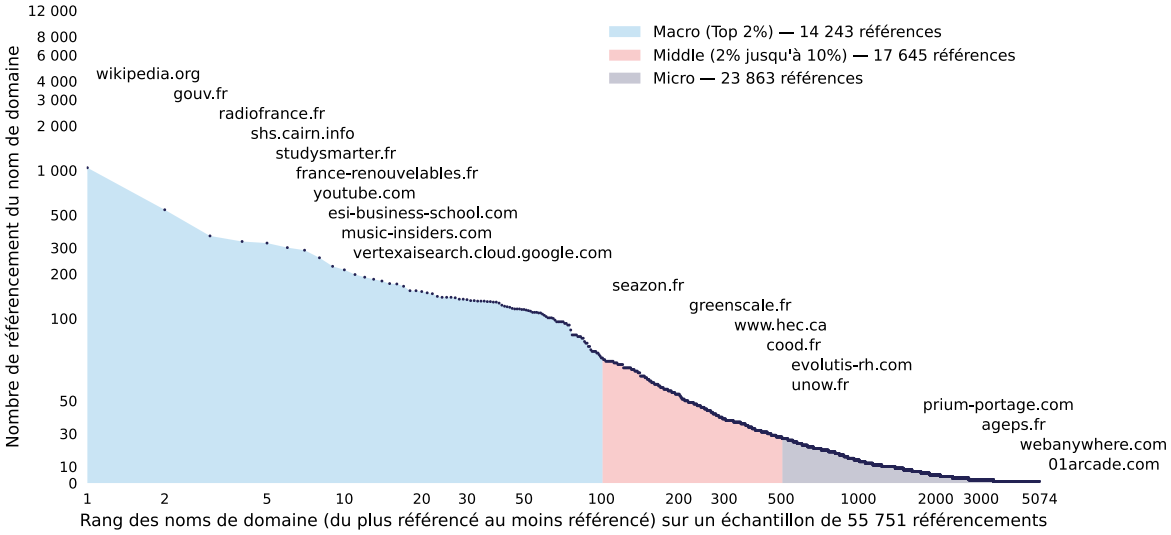


Figure 32 : ranking of domain names (from most referenced to least referenced) based on a sample of 55,751 references (theme: "general").

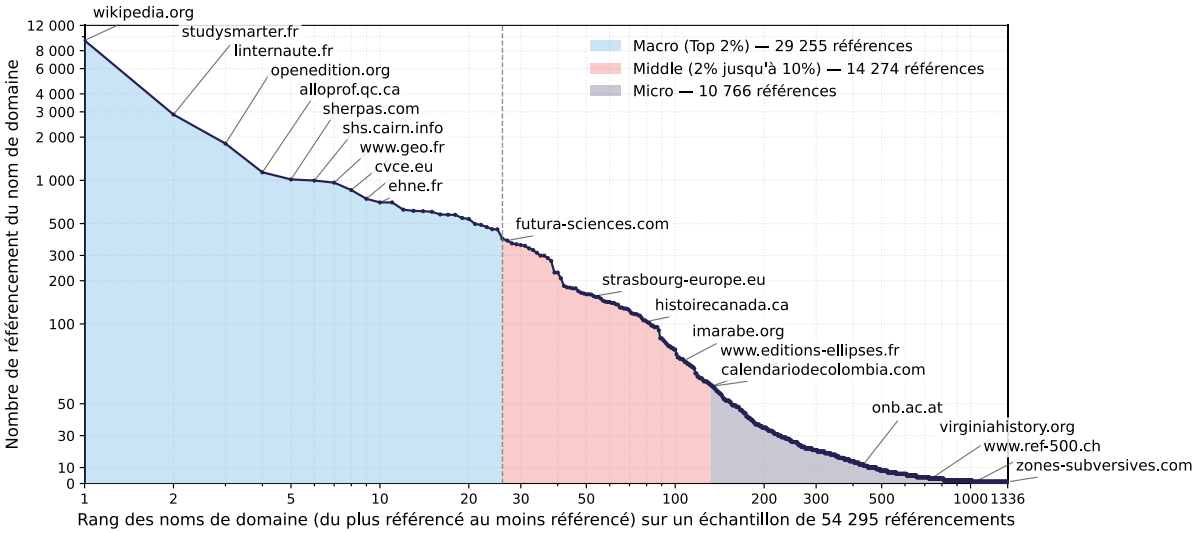


Figure 33 : ranking of domain names (from most referenced to least referenced) based on a sample of 54,295 references (theme: "history").

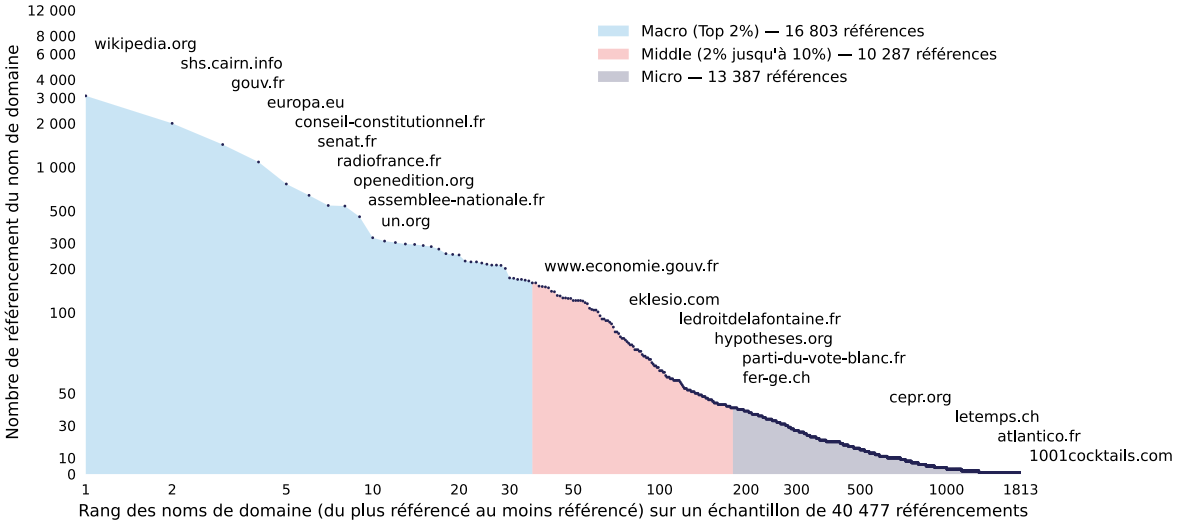


Figure 34 : ranking of domain names (from most referenced to least referenced) based on a sample of 40,477 references (theme: "politics").

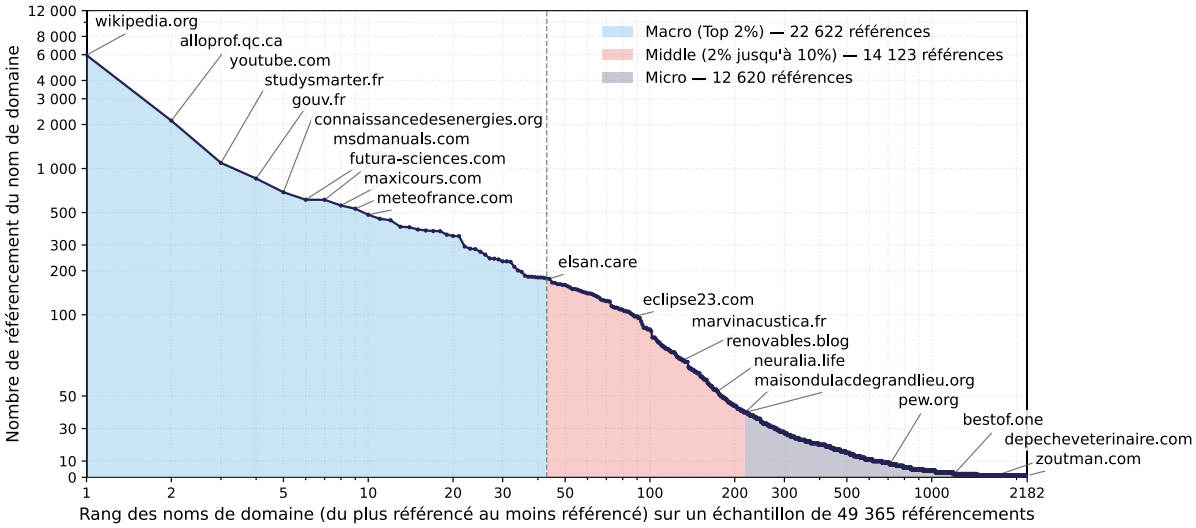


Figure 35 : ranking of domain names (from most referenced to least referenced) based on a sample of 49,365 references (topic: "science")

2.2. Do the links cited by the services studied point to real and accessible web pages?

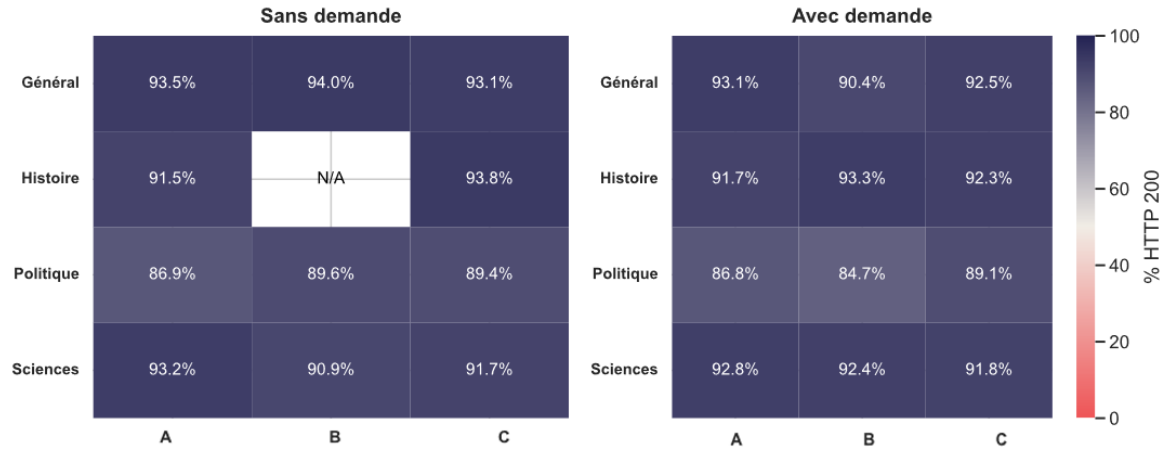


Figure 36 : Proportion of valid links (code 200) for each combination [service x type of question], with (right) or without explicit source request (left).

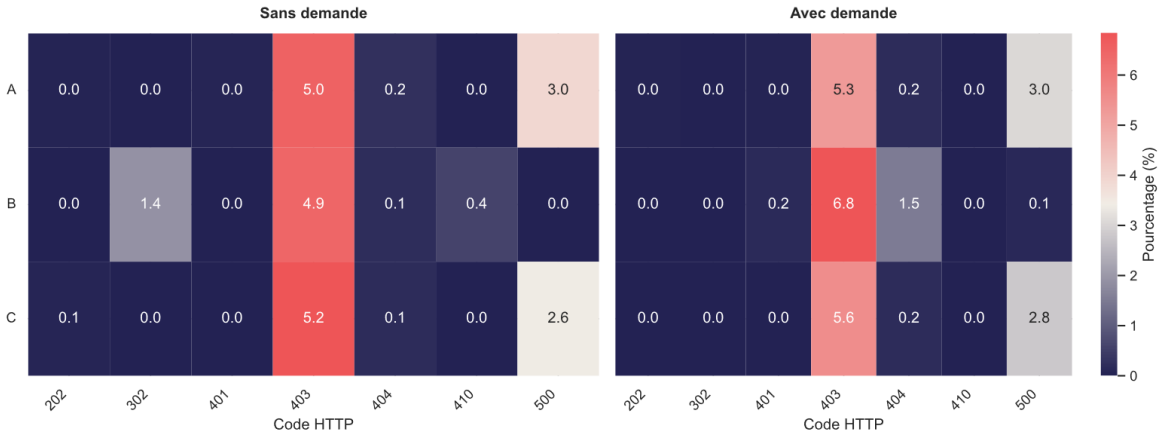


Figure 37 : Percentage of errors for each combination [service x HTTP code type (401/403, etc.)], with (right) or without explicit source request (left).

2.3. Do partnerships between the AI providers studied and the media have an impact on the dynamics of citations on current affairs issues?

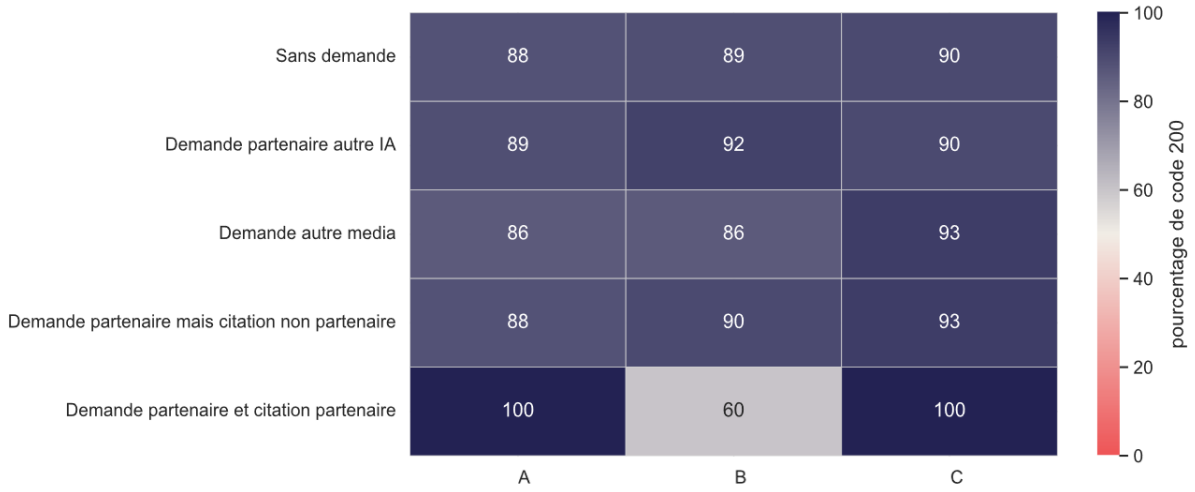


Figure 38 : distribution of functional links by service and type of request.

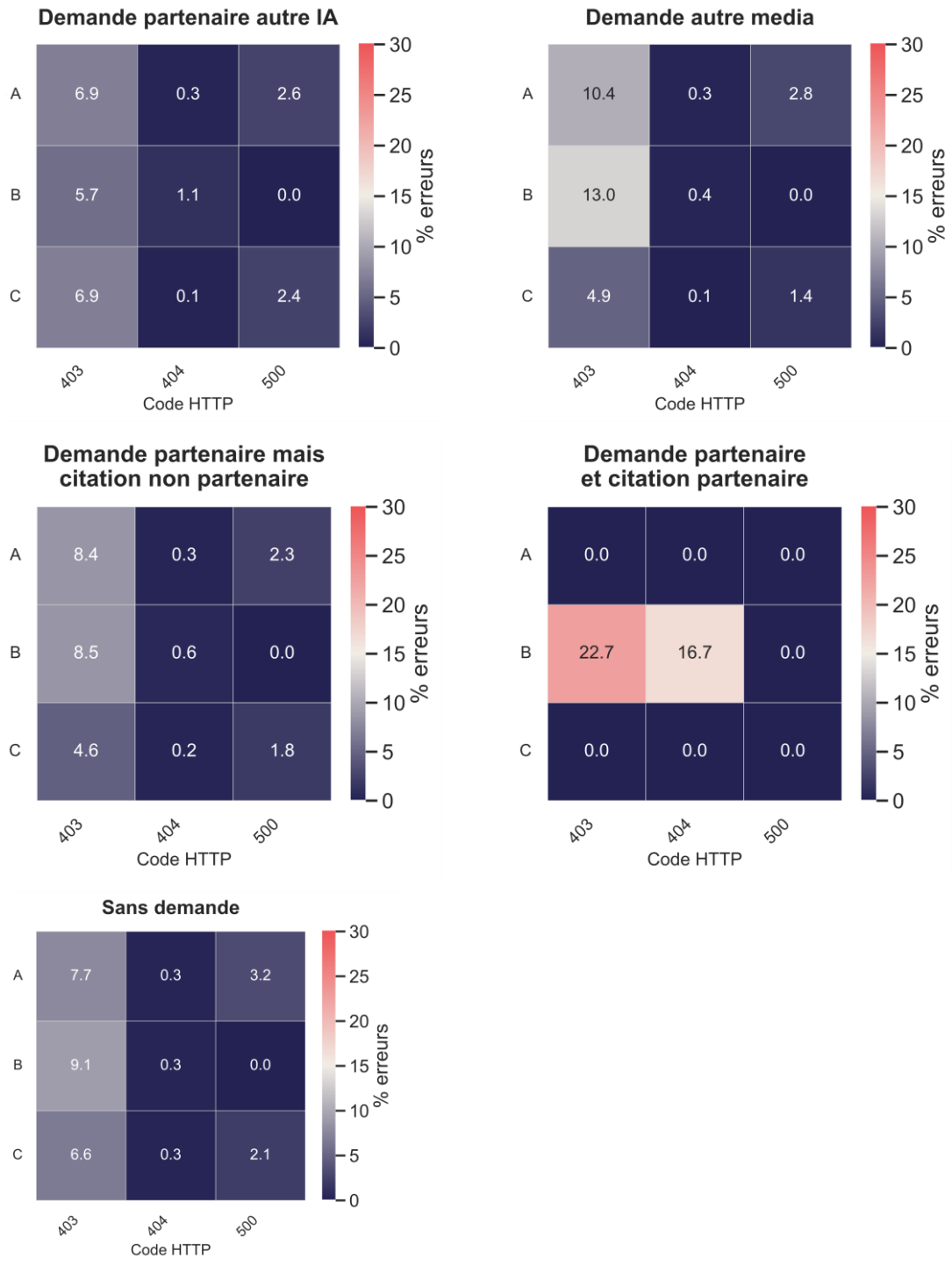


Figure 39 : Distribution of HTTP error codes by request type and service

2.4. Do the sources cited by the services studied actually support the content of their responses?

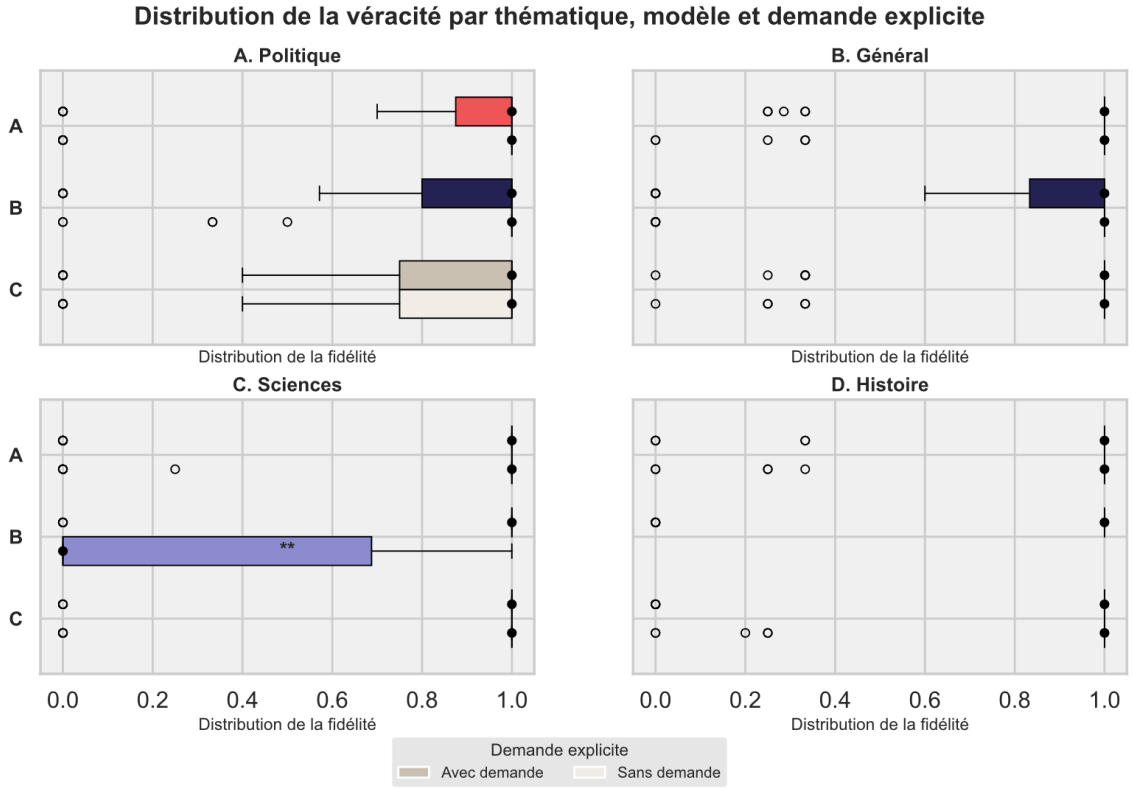


Figure 40 : Accuracy of responses generated by each service, broken down by topic

Reading note: the Figure 40 is a diagram called a "box plot". The left side of the box represents the 1st quartile and the right side represents the 3rd quartile. The solid dots represent the median, while the empty dots represent outliers. For example, for the general theme, service A obtains a median corresponding to a loyalty score of 1 with explicit request, with three specific values. Thus, even though almost all responses are loyal for this theme, in very rare cases the service may use sources that do not support the response.

This document was produced by Arcep

Olivier Corolleur, Director General
Rodolphe Le Ruyet, Advisor to the President

INTERNET, DATA, PRESS, POSTAL SERVICES AND USERS DEPARTMENT

Olivier Delclos, Director

Open Internet Unit

Sandrine Elmi Hersi, Head of Unit

Oriane Piquer-Louis, Project Manager

Vivien Guéant, Expert

Data Services and Cloud Unit

Léo Quentin, Head of Unit

Philéas Samir, Project Manager

Postal Markets and Press Distribution Unit

Bérénice Nguyen, Head of Unit

Nesrine Aouaifia, Project Manager

ECONOMY, MARKETS AND DIGITAL DEPARTMENT

Anne Yvrande-Billon, Director

Blaise Soury-Lavergne, Deputy Director

Economic Analysis and Digital Intelligence Unit

Marion Panfili, Head of Unit

Chiara Caccinelli, Deputy Head of Unit

Nathan Salmon, Project Manager

EUROPE AND INTERNATIONAL DEPARTMENT

Anne Lenfant, Director

Europe Unit

Emmanuel Leroux, Head of Unit

Boris Gartner, Project Manager

COMMUNICATION AND PARTNERSHIPS DEPARTMENT

Clémentine Beaumont, Director

Lydia Lacroix, Project Manager

LEGAL AFFAIRS DEPARTMENT

Élisabeth Suel, Director

Open Infrastructure and Networks Unit

Henda Boucetta, Head of Unit

Paul Bougard, Project Manager

Mobile Market and Scarce Resources Unit

Léa Ployaert, Head of Unit

Suzelle Ayité, Project Manager

Many thanks to the PEReN team for their contribution!

The Center of expertise for digital platform regulation (PEReN) supports government departments and independent administrations involved in the regulation of digital platforms and artificial intelligence. As a centre of technical expertise, it pools state-of-the-art skills in data science, algorithmic processing and auditing, AI evaluation and software development to produce tools, studies and digital consulting services. It also conducts 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: com@arcep.fr

14 January 2026

ISSN No. 2258-3106