

Draft BEREC report on the impact of generative AI on internet openness

1 October 2026

Contents

Executive Summary	2
1. Introduction	5
2. An overview of the competitive dynamics of generative and agentic AI in the internet ecosystem	7
2.1. From generative AI to agentic AI.....	7
2.2. Considerations about competition dynamics.....	8
2.3. The integration of generative and agentic AI in NI-ICS	10
3. Protocols in the context of generative and agentic AI.....	11
3.1. An overview of open and closed protocols.....	12
3.2. Model Context Protocol.....	13
3.3. WebMCP.....	16
3.4. Robots.txt – Robots Exclusion Protocol	19
4. The impact of generative AI on internet openness	21
4.1. The principles of internet openness	22
4.2. The impact of generative AI on users' access to online content and services	22
4.2.1. From search engines to response engines	23
4.2.2. Agentic AI: integration in and impact on digital services.....	26
4.3. The impact of generative AI on users' incentives to produce and distribute content and services online	27
4.3.1. Economic and informational consequences of generative AI: challenges posed by mediated access to content	27
4.3.2. Impact of synthetic content on internet diversity	32
5. Relevant regulatory frameworks to tackle the identified issues	33
5.1. The Open Internet Regulation (OIR)	33
5.2. The AI Act	34
5.3. The Digital Markets Act.....	36
5.4. The Digital Services Act	37
5.5. The Data Act	38
5.6. Coordination across instruments and enforcement authorities, and potential regulatory gaps	39
6. Conclusion	41
Annex: List of abbreviations.....	43

Executive Summary

Generative AI services have spread rapidly in the last five years and have deeply modified our everyday digital activities. Such technology can greatly benefit a variety of different actors, including end and business users, but also national regulatory authorities (NRAs) who increasingly use AI solutions to support their regulatory tasks (e.g. for market surveillance, competition analysis and spectrum management or for knowledge-sharing on AI regulation).

While generative AI can enrich user experience and complement traditional methods of accessing, generating and distributing online content and services, it can also affect the overall user experience on the internet as it has the potential to directly influence the way users access online content/services and possibly, therefore, their freedom of choice. Their widespread integration into a variety of digital services and products (e.g. search engines, messaging services, operating systems etc.) can exacerbate this possibility.

This draft BEREC report analyses the impact of generative AI on internet openness, understood as the ability of users to access and distribute information, content and services online without unlawful interference or discrimination, and the ability to innovate. The report examines the evolving structure of the AI ecosystem, the technical protocols that govern interactions between AI systems and online services, the implications for users and content providers, and the extent to which existing European regulatory frameworks can address the resulting challenges.

The analysis shows that generative AI is creating a new layer of intermediation within the internet ecosystem. Whereas search engines have traditionally organised access to information by ranking and presenting links, generative AI services increasingly provide direct responses, summaries and recommendations, while AI agents are beginning to perform actions on behalf of the users. As a result, generative AI is becoming a new gateway through which users access information, products and services. While these developments can generate significant efficiencies, reduce search costs and improve user experience, they also concentrate decision-making power within a limited number of digital intermediaries and raise questions regarding users' autonomy and freedom of choice, transparency, and competition.

The report finds that competitive dynamics in the AI ecosystem are characterised by strong structural dependencies between model developers, infrastructure/cloud providers and downstream applications. Although competition remains active in the development of foundation models, key infrastructure and distribution layers remain concentrated among a small number of Big Tech companies. The integration of AI functionalities into already dominant digital services reinforces existing market positions through defaults, preferential integration, access to proprietary data and ecosystem lock-in. These concerns become even more significant in the context of agentic AI, where users increasingly delegate both information retrieval and decision-making tasks to AI, and where only three actors account for over 84% of the AI agents sector.

Technical protocols are likely to play an increasingly important role in shaping the future AI ecosystem. They shape technical architectures and interfaces between services and can influence how markets develop. By adopting protocols, parties use common rules to exchange information, thereby decreasing costs to negotiate and coordinate. The report examines developments such as the Model Context Protocol (MCP) and WebMCP, which aim to enable interoperability between AI services, digital tools and online platforms. These protocols may reduce barriers to entry, facilitate switching and support innovation by allowing a wider range of services to connect to AI systems. At the same time, they may increase the influence of providers controlling user-facing AI interfaces, potentially creating new forms of dependency and concentration. The draft report also assesses the role of the Robots Exclusion Protocol (robots.txt), which remains the principal mechanism through which content providers express preferences regarding AI crawling and text-and-data mining. However, these existing mechanisms present some limitations, especially in terms of enforceability and transparency, which highlights the need for further technical and governance developments.

A key finding of this draft report concerns the transformation of access to online content and services. By generating synthesised answers rather than directing users to a range of sources, generative AI changes the relationship between users and online content. Evidence reviewed in the report suggests that AI-generated answers can reduce traffic to original websites and concentrate attention on a relatively small number of sources. This may limit source diversity, reduce transparency regarding how information is selected and prioritised, and increase users' reliance on AI-mediated representations of information. The rise of AI-powered browsers and agents reinforces these trends by allowing users to delegate increasing number of online activities to automated systems. Although these services can significantly improve convenience, they also raise concerns regarding transparency, accountability and user choice.

The report further highlights the growing tension between generative AI and the sustainability of the production of online content. Generative AI services rely heavily on content created by publishers, creators, researchers and public-interest organisations for training, retrieval and grounding purposes. However, AI-generated summaries and the expansion of zero-click experiences may reduce traffic, advertising revenues and subscription opportunities for those content providers. In parallel, increasing AI crawling activity creates additional costs and operational burdens for many websites and digital commons. The draft report therefore identifies a structural paradox: AI systems depend on the continued production of high-quality content while simultaneously weakening some of the economic incentives that support its creation and dissemination.

Although no single European regulatory instrument can fully address the broader implications of generative AI for internet openness, the analysis shows that several frameworks can tackle some of the identified challenges. The Digital Markets Act establishes a set of rules on some problematic gatekeeper behaviours, for example, imposes interoperability or bans self-preferencing, among others. The Digital Services Act provides mechanisms to identify and mitigate systemic risks affecting media pluralism and access to information. The AI Act

strengthens obligations relating to copyright compliance, transparency and respect for rights reservations, while the Data Act supports interoperability and switching in cloud-based environments. Effective coordination among all these frameworks, robust transparency obligations, and meaningful user choice will be key for internet openness and to ensure that the internet remains an engine of innovation for the benefit of users and society as a whole.

1. Introduction

Generative AI services represent a major innovation whose use has accelerated across many different sectors – such as professional services, ICT, telecommunications¹, finance, logistics, transportation – where firms use AI foundation models for summarisation, analytics, coding, workflow automation, human resource management² and customer engagement, among others.

BEREC has already produced several deliverables on AI and related topics, such as its BEREC High-level position on artificial intelligence and virtual worlds³, the BEREC Report on the impact of AI solutions in the telecommunications sector⁴, the BEREC Report on Cloud and Edge Computing Services⁵, and the BEREC external study on the trends and cloudification, virtualisation, and softwarisation in telecommunications⁶. BEREC has also organised an internal workshop to share and discuss about the AI solutions which NRA use to support regulatory tasks, the lessons that can be learnt (e.g. regarding barriers to implementation), as well as how and where the application of AI could be further explored.⁷ BEREC is also working on a report⁸ aimed at gathering factual information and providing a comprehensive assessment of the environmental impact of AI (that is increasingly integrated into many sectors, including in electronic communications networks and services), as AI systems are identified as particularly energy-intensive technologies. This is especially relevant, given that the information collected may allow for the inclusion of environmental considerations in the

¹ RTR (2024), Current Use of AI in the Austrian Telecommunications Sector, see: https://www.rtr.at/TKP/aktuelles/publikationen/publikationen/ai_in_austrian-telecomsector.en.html; and ARCEP (2025), Future Networks: AI and electronic communications networks, see: https://en.arcep.fr/uploads/tx_gspublication/future-networks-summary-IA-and-Electronic-Com-Networks_ENG_june2025.pdf

² RTR (2025), Artificial Intelligence in Human Resource Management. Findings from a representative business survey in the context of the European AI Act, see: https://www.rtr.at/ai_in_human_resource_management

³ BoR (24) 68, BEREC High-level position on artificial intelligence and virtual worlds, 15.03.2024, see: <https://www.berec.europa.eu/en/all-documents/berec/opinions/berec-high-level-position-on-artificial-intelligence-and-virtual-worlds>

⁴ BoR (23) 93, BEREC Report on the impact of Artificial Intelligence (AI) solutions in the telecommunications sector on regulation, 08.06.2023, see: <https://www.berec.europa.eu/en/document-categories/berec/reports/berec-report-on-the-impact-of-artificial-intelligence-ai-solutions-in-the-telecommunications-sector-on-regulation>

⁵ BoR (24) 136, BEREC Report on Cloud and Edge Computing Services, 03.10.2024, see: <https://www.berec.europa.eu/en/all-documents/berec/reports/berec-report-on-cloud-and-edge-computing-services>

⁶ BoR (23) 208, External study on the trends and cloudification, virtualization, and softwarization in telecommunications, 07.12.2023, see: <https://www.berec.europa.eu/en/document-categories/berec/reports/external-study-on-the-trends-and-cloudification-virtualization-and-softwarization-in-telecommunications>

⁷ One example of AI used by an NRA is the “AI Act Chatbot” by the Austrian AI Service Desk, established at RTR, developed to respond to inquiries concerning the EU AI Act. The chatbot is built on a retrieval-augmented generation (RAG) architecture, whereby a language model accesses a curated and trusted knowledge base to generate its responses. The responses to inquiries are available to users instantly, significantly enhancing RTR’s ability to inform the wider public about the AI Act. A key feature of the chatbot is its emphasis on transparency and explainability. The chatbot is also available in English, see: <https://chat.ki.rtr.at/en/>.

⁸ The draft report is scheduled to be published for public consultation in December 2026.

risk assessments as defined by the AI Act, with the aim of minimising the environmental footprint of AI systems in the EU digital markets.

Moreover, in the last years many regulatory and competition authorities in Europe have analysed, among others, the competitive dynamics of (generative) AI and foundation models⁹.

To complement this work, the current BEREC report focuses on the impact of generative AI on internet openness which can be defined as the ability of users to access and distribute information, content and services online, without unlawful interference or discrimination, and the ability to innovate. The current draft report also builds on a previous BEREC Report on the Internet Ecosystem¹⁰ which presented an analysis aimed at understanding how users' internet experience is affected by the different elements of the internet ecosystem in terms of competition and internet openness.

This report is organised as follows: Chapter 2 provides an overview of the competitive dynamics of generative AI, Chapter 3 presents a selection of protocols in the context of generative and agentic AI, Chapter 4 analyses the impact of this technology on internet openness, Chapter 5 focuses on the European regulatory frameworks which could tackle some of the identified issues and Chapter 6 concludes.

It should be noted that for the purpose of this report, the term “users” refers to both end consumers and business users. The term “end-user” in the quotes related to the Open Internet Regulation (OIR) is defined in Art. 2(14) of the European Electronic Communications Code (EECC), where “end-user” means a user not providing public communications networks (ECN) or publicly available communications services (ECS). Moreover, the present report uses the term “internet openness” which is to be understood more broadly than “open internet” in the context of OIR, but is used to examine how conditions and forms of intermediation across the wider internet ecosystem may affect user choice, access to online information and services, the ability of providers to reach users, and innovation.

⁹ Autoridade da Concorrência (2023), Competition and generative artificial intelligence, see: <https://www.concorrenca.pt/sites/default/files/documentos/Issues%20Paper%20-%20Competition%20and%20Generative%20Artificial%20Intelligence.pdf>; Autorité de la Concurrence (2024), Opinion of 28 June 2024 on the competitive functioning of the generative artificial intelligence sector, see: https://www.autoritedelaconcurrence.fr/sites/default/files/commitments/2024-09/24a05_eng.pdf; CMA (2024), AI Foundation Models, see: https://assets.publishing.service.gov.uk/media/661e5a4c7469198185bd3d62/AI_Foundation_Models_technical_update_report.pdf; Ofcom (2025), The Era of Answer Engines: Generative AI's impact on search experiences and online safety, see: <https://www.ofcom.org.uk/internet-based-services/technology/the-era-of-answer-engines-generative-ais-impact-on-search-experiences-and-online-safety>; Arcep (2026), Generative AI: challenges for the future of the open internet, see: https://www.arcep.fr/uploads/tx_gspublication/report-generative-AI-challenges-open-internet-january2026.pdf

¹⁰ BoR (22) 167, BEREC Report on the Internet Ecosystem, 12.12.2022, see: https://www.berec.europa.eu/en/document-categories/berec/reports/berec-report-on-the-internet-ecosystem?language_content_entity=en

2. An overview of the competitive dynamics of generative and agentic AI in the internet ecosystem

2.1. From generative AI to agentic AI

Since the creation of ELIZA around sixty years ago¹¹, the development of this technology has evolved into a generative AI ecosystem composed of foundational model developers (OpenAI, Anthropic, Google DeepMind, Meta, Mistral, Cohere, AI21, DeepSeek, and others), cloud and infrastructure providers (Microsoft Azure, Amazon Web Services, and Google Cloud and others), intermediaries between users and content providers such as search engines and browsers, and a variety of downstream products and services integrating AI.

Overall, the generative AI ecosystem is characterised not only by rapid innovation and diversified supply, but also by structural dependencies and tight linkages – cloud hosting, access to data, model distribution, training compute, and integration into end-user facing services –, which result in a variety of different partnerships. The relationship between AI model developers and cloud providers is intertwined: Microsoft¹² and Amazon¹³ have both a partnership with OpenAI; AWS provides preferred access to Anthropic models¹⁴ and distributes Meta’s Llama models¹⁵; and Google integrates its Gemini models while concurrently offering third-party models through its Google Cloud platform and has a collaboration agreement with Apple.¹⁶

Deeper change may come from the rise of the agentic AI systems¹⁷, which are autonomous AI systems capable of performing multiple sequential steps including access to both in-house and third-party tools and actions such as browsing, messaging, or sending instructions towards a high-level goal. In addition, these can continuously adapt strategies based on results. AI agents can become a new intermediation layer between users and digital services, potentially altering who acts as the gateway to end-users, how demand is steered across suppliers, and how end-users can understand what is done on their behalf. As analysed in the

¹¹ Eliza, a computer programme, created between 1964 and 1966 at MIT (USA), generated text by mimicking conversations using predefined rules and templates. Although limited, it had the ability to generate dialogues, pioneering the concept of automatic text generation. However, ELIZA cannot be classified as GenAI as we understand it today.

¹² See: <https://blogs.microsoft.com/blog/2026/02/27/microsoft-and-openai-joint-statement-on-continuing-partnership/> and <https://blogs.microsoft.com/blog/2026/02/27/microsoft-and-openai-joint-statement-on-continuing-partnership/>

¹³ See: <https://www.aboutamazon.com/news/aws/amazon-open-ai-strategic-partnership-investment>

¹⁴ See <https://aws.amazon.com/es/blogs/machine-learning/introducing-claude-platform-on-aws-anthropics-native-platform-through-your-aws-account/>

¹⁵ See: <https://www.aboutamazon.com/news/aws/aws-meta-llama-4-models-available>

¹⁶ Under which the next generation of Apple Foundation Models will be based on Google’s Gemini models and cloud technology, see: <https://blog.google/company-news/inside-google/company-announcements/joint-statement-google-apple/>

¹⁷ OECD (2026), The agentic AI landscape and its conceptual foundations, see: https://www.oecd.org/content/dam/oecd/en/publications/reports/2026/02/the-agentic-ai-landscape-and-its-conceptual-foundations_a9d4b451/396cf758-en.pdf

recent report prepared by the French Competition authority¹⁸, from a regulatory perspective, AI agents raise important issues related to transparency, accountability, safety, cybersecurity, data protection, and human oversight, especially when they can make decisions or trigger actions with legal, financial, or societal consequences.

2.2. Considerations about competition dynamics

The analysis of competition dynamics in generative AI can focus on different aspects, from competition among AI foundation models to downstream integration of these models with other digital services as well as competition dynamics in the context of AI agents. The AI value chain can be divided into three main layers (infrastructure, modelling and deployment)¹⁹ and competitive level varies, with some layers being more concentrated than others. AI modelling is actively developing, with thousands of AI models available on the market today both open source and proprietary²⁰. However, the infrastructure and deployment, especially its integration in digital services and products, are dominated by few players²¹.

Firstly, generative AI services are typically integrated or linked with – often by default – other digital services, such as search tools (see Chapter 4), e-mail, social networks, number-independent interpersonal communication services (see Chapter 2.3), as well as editing and coding tools. This integration of AI allows large providers to add complementary functionalities to those services and tools, also facilitating the development of agentic AI, that are based on generative AI. However, it also raises several competition questions, since the exclusive or by-default integration of generative AI within other digital services/products has the potential to strengthen the market power of large vertically integrated players, to reinforce market concentration and thus increases the risk of anti-competitive practices, such as tied selling/supplying of services or self-preferencing^{22,23}. These players provide both services in

¹⁸ Autorité de la concurrence (2026), “AI agents: the Autorité de la concurrence issues its opinion on the competitive functioning of the AI agents sector”, see: <https://www.autoritedelaconcurrence.fr/en/press-release/ai-agents-autorite-de-la-concurrence-issues-its-opinion-competitive-functioning-ai>

¹⁹ Autorité de la Concurrence (2024), Opinion of 28 June 2024 on the competitive functioning of the generative artificial intelligence sector, see: https://www.autoritedelaconcurrence.fr/sites/default/files/commitments/2024-09/24a05_eng.pdf; CMA (2024), AI Foundation Models, see: https://assets.publishing.service.gov.uk/media/661e5a4c7469198185bd3d62/AI_Foundation_Models_technical_update_report.pdf

²⁰ CERRE (2025), A competition policy for cloud and AI, see: https://cerre.eu/wp-content/uploads/2025/06/A-Competition-Policy-for-Cloud-and-AI_FINAL.pdf.

²¹ BoR (25) 119, BEREC response to the European Commission's consultation on the first review of the Digital Markets Act, 24.09.2025, see: <https://www.berec.europa.eu/system/files/2025-09/BoR%20%2825%29%20119%20BEREC%20response%20to%20the%20European%20Commission%27s%20consultation%20on%20the%20first%20review%20of%20the%20Digital%20Markets%20Act.pdf>

²² BoR (25) 119, BEREC response to the European Commission's consultation on the first review of the Digital Markets Act, 24.09.2025, see: <https://www.berec.europa.eu/system/files/2025-09/BoR%20%2825%29%20119%20BEREC%20response%20to%20the%20European%20Commission%27s%20consultation%20on%20the%20first%20review%20of%20the%20Digital%20Markets%20Act.pdf>

²³ As a case related to these potential anticompetitive practices, the European Commission opened in December 2025 a formal investigation into possible anticompetitive conduct by Google in the use of online content for AI purposes, see: https://ec.europa.eu/commission/presscorner/detail/nl/ip_25_2964

highly concentrated markets (such as operating systems (OS), 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 can restrict users' freedom to choose the AI services that they prefer, as it also gives these vertically integrated players an advantage over third-party AI developers in their ability to reach users.

Big Tech companies can benefit from several advantages: their privileged or exclusive access to proprietary and large-scale data place them in a favourable position to develop AI models, and the integration of such models with other services they provide further reinforces their data-driven advantage, e.g. thanks to the collection of usage data which are highly valuable for improving the quality of the services.

Secondly, the way AI agents act can benefit competition and reduce users' search costs while also enabling entry for actors that are offering services accessible by the AI agents²⁴. However, this can simultaneously also compound the strength of large, vertically-integrated companies through the control over service usage flows, the verticalization of more services, defaults, control of essential inputs and lack of transparency. In fact, in a dynamic sector characterised by the presence of numerous AI agent developers, only a few actors account for the vast majority of users, with OpenAI, Google and Anthropic together holding more than 84% of the market globally.²⁵

Upstream and downstream competition dynamics mutually influence each other. Some firms that control current access points (OS, browsers, search, app distribution) are often best positioned to control the emerging agent layer as well, and large language model providers can influence downstream markets via their API terms, licensing, and technical limits, so competition depends on open, non-discriminatory access and interoperability. Hyperscale cloud providers also play a pivotal role, as advanced AI requires massive computational resources. Close partnerships between model developers and large cloud providers can create switching costs, enable bundling strategies, and provide preferential access, all of which tend to reinforce incumbent advantages. OS providers can "pick winners" via defaults, pre-installation, and access restrictions, and the role of technical protocols in the context of generative and agentic AI at both the device/OS layer and the application/AI layer is crucial, particularly with regard to interoperability (see Chapter 3).

²⁴ CERRE has also dedicated a report to this topic: CERRE (2024), AI Agents and Ecosystems Contestability, see: <https://cerre.eu/publications/ai-agents-and-ecosystems-contestability/>

²⁵ Autorité de la concurrence (2026), "AI agents: the Autorité de la concurrence issues its opinion on the competitive functioning of the AI agents sector", see: <https://www.autoritedelaconcurrence.fr/en/press-release/ai-agents-autorite-de-la-concurrence-issues-its-opinion-competitive-functioning-ai>

2.3. The integration of generative and agentic AI in NI-ICS

The rapid evolution of both generative and agentic AI has begun to transform number-independent interpersonal communication services (NI-ICS)²⁶. For example, the integrated AI functionality can suggest replies to NI-ICS users, summarize long message threads, instantly translate messages in different languages and produce images within chats. Agentic AI goes beyond simple content generation and can take actions: it may enable functionality within NI-ICS to schedule meetings across multiple calendars, monitor workflows and even enable intelligent spam and fraud detection. For businesses using NI-ICS for customer interactions, agentic AI can help by autonomously resolving common queries, handling multi-step tasks and escalating to humans only when needed. While integrated-AI functionality may result in increased productivity, intelligent spam and fraud detection, as well as faster data analysis and reduced costs, some challenges remain.

Tech firms providing AI services have also integrated these AI capabilities into NI-ICS services. In theory, AI services can be provided by the messaging service itself or by a third-party: (i) directly in-app by the NI-ICS provider (e.g., Meta AI in both WhatsApp and Facebook Messenger), (ii) via NI-ICS API by a third-party (e.g. ChatGPT as chat “partner” in WhatsApp²⁷), (iii) via OS (e.g. transcription of voice messages in WhatsApp), or (iv) by a combination of approaches (e.g. Google Gemini mobile app²⁸).²⁹

Moreover, in November 2025, OpenAI launched ChatGPT group chats, enabling multi-participant conversations where ChatGPT acts as an AI collaborator and could benefit from large network effects present already in messengers. Google has unveiled deep integration of its Gemini models into the Android ecosystem (e.g., on-device Gemini Nano, AI Mode in Search and Gemini integration with Google Messages).

From early 2025 onwards, Meta introduced Meta AI, a large language model (LLM)-powered assistant embedded directly into WhatsApp, operating through mentions (e.g., “@MetaAI”). It provides a dedicated assistant thread, and context-aware suggestions. Meta AI provides

²⁶ BoR (23) 92, BEREC report on interoperability of Number-Independent Interpersonal Communication Services (NI-ICS), 08.06.2023, see: <https://www.berec.europa.eu/en/document-categories/berec/reports/berec-report-on-interoperability-of-number-independent-interpersonal-communication-services-ni-ics>. Number-Independent Interpersonal Communication Services (NI-ICS) are interpersonal communication services which do not connect with publicly assigned numbering resources, or which do not enable communication with a number in national or international numbering plans.

²⁷ Since January 15 2026, ChatGPT is no longer available on WhatsApp, through an update of Meta’s WhatsApp Business Solution Terms, banning third-party general-purpose AI assistants from the application, although Meta has since lifted the ban (as of the end of May 2026); see the European Commission Press Release: [IP_26_310_EN.pdf](https://ec.europa.eu/press/2026/05/26/ip_26_310_en.pdf).

²⁸ On Android devices, Gemini mobile app can be used to make calls and send messages on WhatsApp. Other functions, like read and reply to notifications and summarising messages, are performed by the app with the support of the Utilities app, see: <https://support.google.com/gemini/answer/15574928?hl=en-GB>, for Google’s Press Release. Last consulted by BEREC on 26 May 2026.

²⁹ BoR (25) 119, BEREC response to the European Commission’s consultation on the first review of the Digital Markets Act, 24.09.2025, see: <https://www.berec.europa.eu/en/all-documents/berec/opinions/berec-response-to-the-european-commissions-consultation-on-the-first-review-of-the-digital-markets-act>

several capabilities inside NI-ICS, such as analysing attachments (images, PDFs), drafting responses and generating content and summarising of group discussions within WhatsApp. Furthermore, Meta AI for WhatsApp Business has agentic capabilities and is touted to be capable of handling customer operations autonomously.

Beyond the DMA, Meta's behaviour is being scrutinised under competition law. Meta is accused of banning third-party AI assistants (e.g., ChatGPT, Perplexity) from WhatsApp Business APIs, limiting access to its platform and strengthening Meta AI's exclusivity. This exclusion was deemed to be abusive, hence it triggered antitrust interventions at national and supranational level: i) in July 2025, the Italian Competition Authority (AGCM) opened a case (A576) on such a conduct and, later on, adopted interim measures ordering Meta to suspend the terms excluding competing AI Chatbots from WhatsApp³⁰; ii) in December 2025, also the European Commission opened an antitrust investigation (04/12/2025-AT 41034) on this conduct and imposed interim measures to prevent this policy from causing harm on the market.³¹

Certain functionalities of Meta AI – especially those needing tighter integration with WhatsApp due to requiring access to data of interpersonal conversations – may not be replicable for competing AI chatbots using the WhatsApp Business API. Therefore, even with non-discrimination of the WhatsApp Business API regarding competing AI chatbots, Meta AI has a competitive advantage.

3. Protocols in the context of generative and agentic AI

The BEREC Report on the Internet Ecosystem³² presented an analysis aimed at understanding how users' internet experience is affected by the different elements of the internet ecosystem in terms of competition and internet openness. One finding is that a dominant provider may develop or restrict certain technical architectures (e.g. native app from app stores vs. progressive web apps via the world wide web independently of app stores) in order to shape technical formats, editorial choices and business models within provider-

³⁰ AGCM case (2025), see: https://www.agcm.it/dotcmsdoc/allegati-news/A576_avvio.pdf and <https://en.agcm.it/en/media/press-releases/2025/11/A576> and for interim measures see: https://en.agcm.it/dotcmsdoc/pressrelease/A576_adoz%20misura%20cautelari_omi_signed_signed.pdf

³¹ Commission opens antitrust investigation into Meta's new policy regarding AI providers' access to WhatsApp, see: https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2896; for interim measures: Commission notifies Meta of possible interim measures to reverse exclusion of third-party AI assistants from WhatsApp, see: https://ec.europa.eu/commission/presscorner/detail/en/ip_26_310 and https://ec.europa.eu/commission/presscorner/detail/it/ip_25_2896

³² BoR (22) 167, BEREC Report on the Internet Ecosystem, 12.12.2022, see: https://www.berec.europa.eu/en/search-documents?saa_field_media_file=&field_document_number=BoR+%2822%29+167&saa_field_media_file=&search_api_document_fulltext=&field_tags=&field_category_name=All&created_min=&created_max=&items_per_page=10

specific ecosystems (e.g. of a mobile OS). Protocols are crucial components shaping these technical architectures and can therefore influence how markets develop.

This section provides an overview of selected protocols in the context of generative and agentic AI.

3.1. An overview of open and closed protocols

A protocol is a standardised set of rules that enables interactions between two or more parties, for example between machines, services, agents or people. By adopting protocols parties use common rules to exchange information, thereby decreasing costs to negotiate and coordinate on the design of such rules and their implementation (transaction costs). The adoption of a protocol by another additional user usually benefits both, this user and all the users which already adopted the protocol (network effects).

Open protocols are often characterised by³³:

- Being developed by a non-profit organisation independent from specific providers which typically defines and updates the rules,
- The potential participation of interested parties (e.g. organisations, companies, individuals),
- Being used by different providers with demand for common rules and functionalities (“market acceptance”),
- Accessible and transparent documentation of rules,
- Free and open-source licensing.

The use of open protocols may facilitate innovation across markets thanks to a reduction in switching costs, (thereby boosting user choice), lowering of barriers to market entry and expansion, and to a maximisation of network effects. Usually, the architecture and program code of open protocols can be reviewed by anybody, which can benefit their security. Disadvantages of open protocols may be slower development of new functionalities, e.g. due to higher costs to coordinate or lower investment incentives when resulting benefits are shared by all adopters (free-riding). The Internet Protocol suite (TCP/IP)³⁴ – a foundational framework of protocols for the internet architecture – is often considered to be open.

Proprietary or closed protocols may facilitate faster changes of the rules of protocols due to lower need for coordination and thereby also facilitate the innovation, but only by the players

³³European Commission (2022), Common Assessment Method for Standards and Specifications, CAMSS Assessment MSP Scenario Methodology, see: <https://interoperable-europe.ec.europa.eu/collection/common-assessment-method-standards-and-specifications-camss/solution/camss-assessment-multi-stakeholder-platform-msp-scenario/distribution/camss-assessment-msp-scenario-methodology-v400> The assessment method is a framework designed to guide public administrations to select and implement ICT solutions that are interoperable, secure, and aligned with relevant frameworks and regulations (e.g. Regulation (EU) 2024/903, the Interoperable Europe Act).

³⁴ See also: <https://datatracker.ietf.org/doc/html/rfc1122>

who are developing or adopting them. They may also provide advantages regarding vendor accountability. Incentives to invest in the development of proprietary protocols may be higher, e.g. due to possibilities to recoup investments with fees or by bundling/tying of services or products. However, commercial practices or technical incompatibilities may also raise switching costs and lock-in³⁵. Some protocols relevant to consumer Internet of Things (IoT) – often developed by large providers of operating systems – are frequently considered proprietary.³⁶

3.2. Model Context Protocol

The Model Context Protocol (MCP) was developed as an open-source protocol by Anthropic in November 2024³⁷, and has gained widespread traction after its adoption by OpenAI, AWS, Azure, Perplexity AI, Microsoft Copilot Studio, Google, and major Software Development Kits (SDKs). In December 2025 MCP was donated by Anthropic as a contribution to the Agentic AI Foundation (AAIF), which was created by the Linux Foundation.

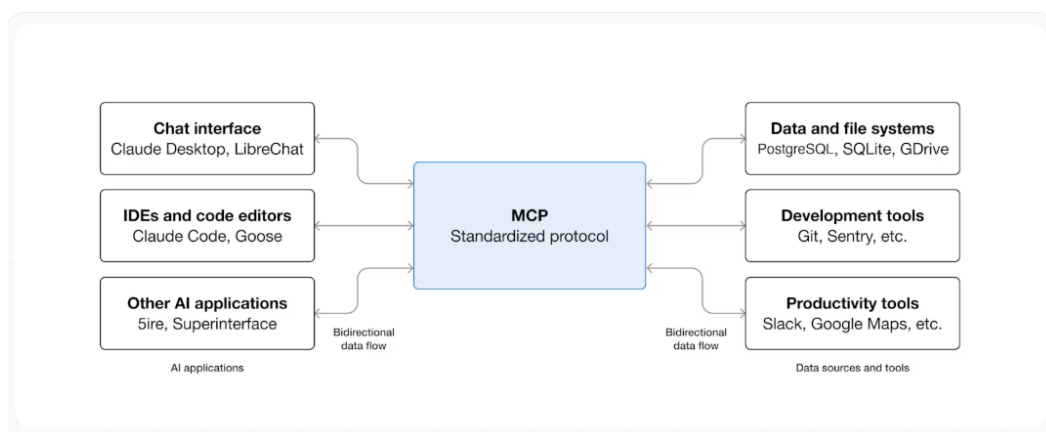


Figure 1: Diagram of Model Context Protocol³⁸

MCP facilitates and enhances the development and operation of AI agents³⁹ by providing a standardised interface between LLMs and the external systems that provide the LLMs with

³⁵ Competition and Markets Authority (2014), The Economics of open and closed systems, see: https://assets.publishing.service.gov.uk/media/5a75c284e5274a545822e01a/The_economics_of_open_and_closed_systems.pdf

³⁶ The Data Act aims to address some of the identified competition concerns in the IoT sector. See also: European Commission (2022), Final report - sector inquiry into consumer Internet of Things, see: https://competition-policy.ec.europa.eu/document/download/c74fd54d-7cbb-41c5-aad5-0175be7b8a99_en

³⁷ See: <https://www.anthropic.com/news/model-context-protocol>

³⁸ Source: <https://modelcontextprotocol.github.io/docs/getting-started/intro>

³⁹ There is no universal definition of AI agents. In this context AI agents are characterized as a system where an LLM self-directs its logic and tool selection by its own reasoning, rather than through predefined code.

context or that allow LLMs to trigger actions in other systems, i.e. real-time access to data and tools enabling the completion of a specific task (see Figure 1). The main MCP elements are:

- Host, i.e. the AI application the user usually interacts with. It orchestrates the whole session (e.g. ChatGPT interface, Claude Code),
- Client, i.e. a connector built inside the AI application. It handles the technical communication between the MCP host and one specific MCP server,
- Server, i.e. a lightweight program acting as translator between the client and specific external systems, which often will be a tool (an executable function AI can call to perform a task, e.g. create a new event in Google Calendar) or a data resource (passive, read-only information providing background context to the AI, e.g. retrieve content from specific files on a Google Drive). Also, the server enables clients to discover, which MCP server functionalities exist in a given system and are accessible through the MCP.

The strength of the protocol lies in allowing the seamless integration of MCP servers with AI applications, eliminating the need for bespoke connectors for each LLM-tool combination.⁴⁰ For the user of AI tools, MCP allows them to connect to a variety of services outside a specific set provided by their AI tool provider. This can be illustrated by the following scenario of planning a business trip. Before MCP, an AI chatbot might have had built-in plugins/connectors⁴¹ that worked only with certain travel companies, also the user had to manually check their calendar for date collisions. By contrast if travel companies have set up their service as MCP servers (and thereby are available for all AI applications which, as a prerequisite, support MCP), MCP enables an AI application to access the user's calendars of choice (e.g. work calendar, private calendars – if the calendar provider has implemented MCP support) and check for the most convenient travel dates and also can add the trip. In short, the use of MCP by AI agent services results in significant benefits for end-users in externalising the management of their tasks. News publisher and content producer (e.g. Reuters⁴²) already use MCP servers to enable access to content via AI applications.

The protocol aims to make it easy to set up MCP servers, and at the same time allow high complexity for advanced use cases, as they can interface with all existing MCP server providers, without needing further adjustments. As a consequence, new AI agents can be

⁴⁰ MCP addresses the so-called M x N integration problem by reducing it to a M + N solution. Previously, if a developer wanted to integrate e.g. M = 3 LLM models with N = 8 tools/data resources, they had to write 3 x 8 = 24 connectors, 1 for each LLM model – tool/data resource combination. MCP on the other hand provides a standardised interface (the MCP server) for connecting the tools/data resources to LLMs, meaning that a developer now has to write only 1 connector for each LLM model and each tool/data resource, resulting in M + N = 3 + 8 = 11 connectors and as a result substantially reducing the effort for designing new AI agents.

⁴¹ For example, ChatGPT's (deprecated) plugins for travel companies like Expedia, see: <https://openai.com/index/chatgpt-plugins/>. Also, these plugins had to be built specifically for each LLM, which led to the so-called M x N integration problem (see footnote above).

⁴² See: <https://www.reuters.com/media-center/reuters-launches-model-context-protocol-server-bring-trusted-news-directly-into-2026-07-08/>

deployed fast, for a wide variety of use cases and connecting to a variety of actors providing services. Providers already active in markets for (often business internal) productivity, cloud, data analytics and communication services are increasingly interested to provide AI solutions via MCP, often targeting the integration of business internal data resources.⁴³ Notably BirdyChat, a business messaging app and the first third-party NI-ICS on the market implementing interoperability with WhatsApp according to Article 7 DMA⁴⁴, also offers its own MCP server, giving AI assistants access to read and write messages⁴⁵. Still, at large, MCP adoption among dominant market players currently presents a mixed picture. Perplexity, for instance, uses MCP for the integration of custom tools and resources, i.e. connectors, in its answer engine for end users – but only in the paid version. At the same time, it is discontinuing the internal use of MCP in favour of accessing tools and resources via API and command line interface.

Regarding the prevalence of MCP server features⁴⁶ a clear shift can be observed, moving away from perception (gathering information from databases, internet search) and reasoning tasks (analysis, planning, resource management) towards actions directly affecting the environment, with the latter marking a transition from specific software extensions (e.g. for calendar, third-party services) to general-purpose ‘computer use’-type tools for browser automation, computer control, and remote mobile phone use. MCP shows the highest adoption rates in software development and IT tasks, followed by finance and business management tasks⁴⁷. There is also a growing trend of AI agents building their own functional capabilities⁴⁸.

However, security poses a major challenge in MCP: first, MCP lacks built-in security features like authorisation, which is optional, or encryption, which is delegated to the underlying transport layers, and host systems; unless strong external security measures are implemented data could be exposed during transport and storage. Also, in order for certain tasks to be fulfilled, AI agents might need persistent access not only to often security-sensitive tools (e.g. delete files) and data resources, but at the same time also sensitive areas to which systems might (inadvertently) be exposed. Furthermore, robust security implementation in MCP systems is a task whose complexity is often underestimated.

⁴³ For example: Microsoft MCP servers, see: <https://github.com/microsoft/mcp>

⁴⁴ See: https://www.europarl.europa.eu/doceo/document/E-10-2026-000134-ASW_EN.pdf, as of July 2026 Haiket, the second messaging service reaching interoperability with WhatsApp, is in its beta testing phase.

⁴⁵ See: <https://docs.birdy.chat/mcp.html>. As of summer 2026 this functionality is limited to internal BirdyChat chats and does not include interoperable WhatsApp third-party chats.

⁴⁶ Between November 2024 and January 2026, and based on download metrics, where 19,388 verified public MCP server repositories from GitHub and Smithery MCP registry (<https://smithery.ai/>) were evaluated. Merlin Stein (2026): How are AI agents used? Evidence from 177,000 MCP tools, see: <https://arxiv.org/abs/2603.23802>

⁴⁷ 67% of all MCP features evaluated and 90% of MCP server downloads accounting for software development and IT tasks, 18% of all MCP features evaluated and 5% of MCP server downloads accounting for finance and business management tasks. Merlin Stein (2026): How are AI agents used? Evidence from 177,000 MCP tools, see: <https://arxiv.org/abs/2603.23802>

⁴⁸ MCP servers created with the help of AI rising from 6% in January 2025 to 55% in January 2026. Merlin Stein (2026): How are AI agents used? Evidence from 177,000 MCP tools, see: <https://arxiv.org/abs/2603.23802>

In addition, there seem potential limitations with regard to the openness of MCP, as claimed by its creators. This concerns its hosting by the AAIF, whose maxim is providing a neutral and open foundation for the transparent and collaborative evolution of agentic AI. However, the highest level of governance and strategic influence over the foundation's direction, including sole voting rights, are granted to eight so-called Platinum members (AWS, Anthropic, Block, Bloomberg, Cloudflare, Google, Microsoft, and OpenAI), with the three Contributing Platinum Members Anthropic, Block and OpenAI holding most of the authority.⁴⁹ Moreover, modifications to the MCP codebase are carried out primarily by developers from the ranks of AAIF Platinum members⁵⁰.

Furthermore, MCP may pose additional burdens on service providers' business models. As the interaction from the third-party AI application with the MCP server takes place only on a machine-to-machine basis, the end-user never directly interacts with the service provider. In the above example of a travel agency, as the user never interacts directly with the travel agency website, showing ads, suggesting additional services etc is not possible. This may limit the interest of at least some service providers to make their service available through an MCP server. With the same reasoning that some providers choose to ban AI crawlers from their websites, charge additional fees for API use, they may also choose to not implement MCP in order to keep the direct interaction with end-user traffic.

On the one hand, MCP may potentially counteract fragmentation by fostering interoperability between AI applications, tools and data sources, and by removing compatibility issues across different open/proprietary APIs and formats, possibly facilitating switching between AI applications, tools and data resource providers, as well as levelling the playing field, allowing smaller companies to be part of this ecosystem. On the other hand, MCP may increase the influence of providers controlling user-facing AI interfaces and the design of the protocol, potentially creating new forms of dependency and concentration. Currently there is little information on the uptake of the protocol, especially for commercial use cases. Therefore, the effect of increasing integrations has yet to be seen.

3.3. WebMCP

In February 2026, Google announced the launch of WebMCP⁵¹ - a JavaScript API that lets websites expose structured tools to AI agents, enabling agents to reliably perform actions on web pages using declared inputs, outputs, and context-, which was jointly developed with

⁴⁹ According to The Agentic AI Foundation Charter (2025), there is a "Platinum Member Cap", i.e. "At no time may there be more than eight (8) Platinum Members", and any amendment to this charter or an increase of this cap requires approval by all three Contributing Platinum Members during the first three years, see: <https://github.com/aaif/foundation/blob/main/charter.md>. AAIF was founded with the project contributions of MCP by Anthropic, Goose AI agent framework by Block, and Agents.md coding agents specification by OpenAI, see: <https://aaif.io/press/linux-foundation-announces-the-formation-of-the-agentic-ai-foundation-aaif-anchored-by-new-project-contributions-including-model-context-protocol-mcp-goose-and-agents-md/>.

⁵⁰ See <https://github.com/modelcontextprotocol/modelcontextprotocol/graphs/contributors>.

⁵¹ See: <https://developer.chrome.com/blog/webmcp-epp>

Microsoft. The project was later moved to the W3C Web Machine Learning Community Group with the status as a community draft⁵². However, it remains to be seen to what extent the protocol will gain traction, as the incentives for website operators are still an open question at this time. The protocol is inspired by MCP and complementing it, providing a standardised and reliable way for web applications and web services to expose their functionalities to AI agents directly in the user's browser tab (i.e. in the frontend) and without the need for communication with such components hosted on a server (i.e. in the backend). MCP by contrast enables AI web access to backend web data and services. As to WebMCP, the website exposes the agent's actions directly inside the user's web session, and by this it allows the collaboration of human users and agents inside a shared user interface.

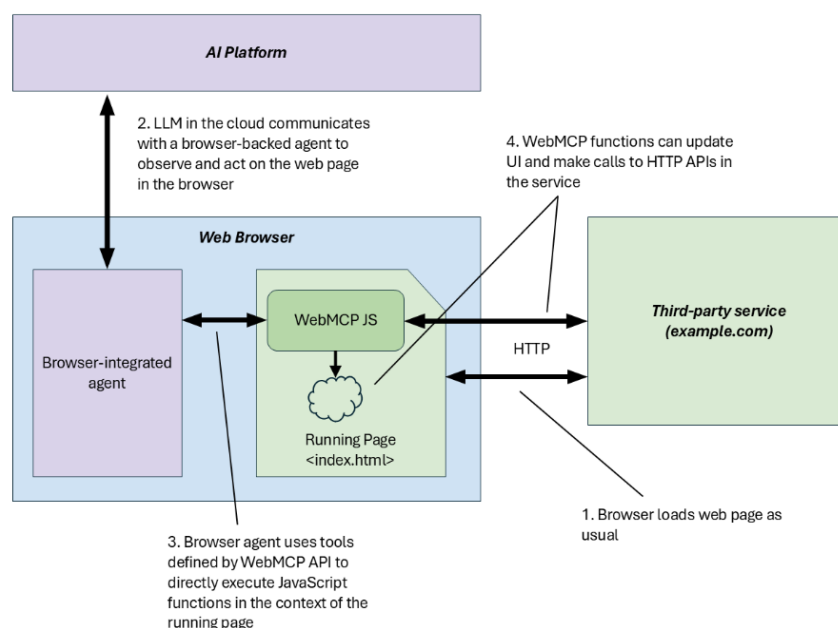


Figure 2: Diagram of WebMCP⁵³

Figure 2 presents the main WebMCP elements:

- Web page, loading as usual, but including tools in the form of JavaScript functions (e.g. filter a list, fill in a field, check a box), which can be called by browser agents, i.e. the web page can be thought of as an MCP server,

⁵² While the World Wide Web Consortium (W3C) develops open standards and guidelines aimed at ensuring accessibility and interoperability of the World Wide Web, it also hosts open community groups for collaboration of developers and stakeholders, although no official standards are created there. The W3C Web Machine Learning Community Group is aimed at developing machine learning inference in the browser, see: <https://webmachinelearning.github.io/webmcp/>

⁵³ Source: <https://github.com/webmachinelearning/webmcp>

- LLM-based AI platform⁵⁴ (e.g. Google's Gemini) orchestrating the whole process, providing reasoning and a user interface, and which is integrated with a certain web browser (e.g. Google's Chrome), which it observes and controls,
- Browser-integrated agent, receiving instructions from the AI platform, and translating them into WebMCP tool calls in the context of the running web page,
- Third-party services, which can be called by WebMCP tools (e.g. a request to a store's backend server to verify stock), and also can update the web pages' user interface (e.g. update shopping cart).

The functionality of WebMCP⁵⁵ can be illustrated using the following scenario of online shopping for a summer dress with the help of an AI assistant, which the user first instructs, giving the key facts (e.g. style, colour, price range). The AI assistant, typically knowing the user's favourite vendors from context (e.g. previous browser searches/shops visits), presents the user with an overview of potential shops compatible with the current task. After the user decides on a shop, the AI agent opens the user's browser (which it has to be integrated with), and navigates to the respective webpage with dresses. Already knowing the user's size from context, the agent first calls a tool included in the web page, enabling the choice of dresses a certain size and colour, and then another tool, enabling the showing of these dresses, by updating the user interface. WebMCP implies a shift for AI powered web browsers (e.g. Comet by Perplexity), which currently have to process the whole rendered code of websites or to interpret screenshots, towards websites becoming tool providers which can be directly called by AI agents inside these browsers, which thereby gain the ability to process tasks faster, more accurately, and more reliably. Thus, while WebMCP promises greater interoperability on some levels, it remains to be seen how it will affect third-party providers with regard to the integration of the AI service into the browser.

WebMCP has the potential to substitute to some degree traffic to websites manually requested by end-user with automated access to functionalities of websites. By controlling the user-facing interface the provider of the AI platform can play the role of a new intermediary – and performs e.g. tasks of structuring, ranking or filtering – which otherwise could be performed by the provider of other user-facing interfaces (websites, apps). On the one hand, this can shift more control towards the user-facing AI platform (including monetisation through advertising) at the expense of other user-facing interfaces. On the other hand, WebMCP – by providing a common technical interface between AI platform and browser agents integrated in websites – may also reduce consumer's barrier to switch to competing AI platforms or web browsers which may use WebMCP as well to replicate functionalities.

⁵⁴ Unlike the configuration shown in Figure 2, the AI platform could also be deployed locally on the user's premises.

⁵⁵ See: <https://github.com/webmachinelearning/webmcp>

3.4. Robots.txt – Robots Exclusion Protocol

The Robots Exclusion Protocol is an industry standard that existed before the advent of generative AI. Created as early as 1994, the protocol enables the machine-readable storage of rules on how website content may be processed by automated Web crawlers (“robots”) by creating a text file that can be called up under “/robots.txt”. The providers of the leading search engines adhere to these rules when indexing websites for search engines.⁵⁶ Following its development as an industry standard, the protocol was also formally standardised by Internet Engineering Task Force (IETF) in 2022.⁵⁷

Generative AI creates content, e.g. text, images, music, videos or programme code – guided by using human creativity as a reference. The foundations for the development of generative AI models are large amounts of high-quality data for training, which can be, and often are, protected by copyright. The European Copyright Directive defines “text and data mining” (TDM) as “*any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations*”.⁵⁸

To enable such use cases, the Copyright Directive authorises reproduction and extraction for TDM (including for commercial purposes) “*on condition that the use of works [...] has not been expressly reserved by their right-holders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.*”⁵⁹ AI developers often use web crawlers to collect data. These web crawlers automatically navigate to websites and read and store information available on them. The robots.txt protocol is an unofficial, de facto standard to provide website operators control over access, its modalities and thus usage restrictions. The European AI Act requires providers of general-purpose AI (GPAI) models to “*identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed*” as well as to publish summaries about the content used for training.⁶⁰ The GPAI Code of Practice refers to the use of the Robot Exclusion Protocol (robots.txt), as specified in the IETF Request for Comments No. 9309, (ii) subsequent versions of the protocol and (iii) other appropriate machine-readable protocols. Signatories of the GPAI Code of Practice, which also provide a designated online search engine (and therefore have a potential dual role in crawling related to indexing search engines as well as AI) should also ensure that rights reservations

⁵⁶ See: <https://developers.google.com/search/docs/crawling-indexing/robots/intro> and <https://www.bing.com/webmasters/help/how-to-create-a-robots-txt-file-cb7c31ec>; and <https://duckduckgo.com/duckduckgo-help-pages/results/duckduckbot>

⁵⁷ See: <https://datatracker.ietf.org/doc/rfc9309/>

⁵⁸ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, Art. 2(2), see: <https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng>

⁵⁹ Directive (EU) 2019/790, Art. 4(3), see: <https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng>

⁶⁰ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art 53 1. (c) and (d), see: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

do “not directly lead to adverse effects on the indexing of the content, domain(s) and/or URL(s), for which a rights reservation has been expressed, in their search engine”.⁶¹

The advantage of the Robots Exclusion Protocol is that a centralised overview of all rules that should apply to crawlers in each website is in a file that always has the same name ('robots.txt'). The protocol nevertheless also has some limitations:

1. Reference to crawler names: unless all crawlers should be addressed with the same rule, website owners have to specify the names of the crawlers to be addressed. These names may change frequently due to market entries or exits or companies changing the names of their crawlers. Google alone identifies several handfuls of crawlers.⁶²
2. General inclusion/exclusion using “Allow” or “Disallow”: it is not possible to set TDM preferences granularly. For example, by differentiating between “Training (commercial)”, “Training (scientific)”, “Data mining”, etc. Especially in the AI context, setting different preferences for search engine indexing, AI training and AI assistants is only possible if different crawlers are used.
3. No assurance of consideration of preferences: from a technical point of view, the Robots Exclusion Protocol can be used to set preferences, but actual implementation from the crawler side cannot be technically enforced in the website side. Companies' crawlers can adhere to the rules set out in robots.txt but can – technically – also simply ignore them. In practice it seems hard for rightsholders to verify whether their preferences communicated via robots.txt are respected – by AI developers in general or by a web crawler of an individual AI developer. This raises questions about the practical enforcement of the copyright directive via robots.txt.

Despite these limitations, the Robots Exclusion Protocol is currently the most common method of setting preferences for TDM crawlers due to its existing standardisation and historical distribution.

The Austrian AI service desk set up at RTR conducted an empirical analysis of the development of TDM preferences of Austrian websites between 2018 and 2025 based on robots.txt of the dataset available at Common Crawl.⁶³ At least two thirds of the websites use

⁶¹ Code of Practice for General-Purpose AI Models. Copyright Chapter, see: <https://ec.europa.eu/newsroom/dae/redirection/document/118115>

⁶² See: <https://developers.google.com/search/docs/crawling-indexing/overview-google-crawlers>
 Google was accused of using the “Googlebot” crawler to provide search results both in the form of a list of website addresses and in the form of summaries created using generative AI (“AI search”). Hence website operators do not have sufficient control over the use of website content by Google. Search results summarised with generative AI would lead to a loss of revenue compared to the list of website results due to the low number of user accesses to the website itself. See also: <https://pressgazette.co.uk/media-law/google-ai-scraping-crawlers-financial-times-news-publishers/> and <https://www.reuters.com/legal/litigation/googles-ai-overviews-hit-by-eu-antitrust-complaint-independent-publishers-2025-07-04/>. At the same time, some consumers also prefer results summarized with AI. See also: <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/new-front-door-to-the-internet-winning-in-the-age-of-ai-search> and <https://www.reuters.com/legal/litigation/googles-ai-overviews-hit-by-eu-antitrust-complaint-independent-publishers-2025-07-04/>.

⁶³ See: <https://commoncrawl.org/>

the Robots Exclusion Protocol. Around 75% of the websites using the protocol do not differentiate rights for different crawlers, but signal the same rules for all of them. The crawler Googlebot, used at least for indexing for the Google search engine, is restricted only by very few Austrian websites. Restricting the crawler can lead to lower visibility of the website in the search engine. AI crawlers play an increasing role in restrictions since 2023, when LLMs began to be more widely available. Crawlers for the collection of training data for AI models (“AI training”) are more often restricted than crawlers building a search index for AI summarized results for search engines (“AI search”) or AI crawlers accessing websites in real-time after user input (“AI assistants”), Website operators allow access by AI crawlers, notably since 2025.⁶⁴

Alternative protocols to signal TDM preferences are, amongst others, the W3C Text and Data Mining Reservation Protocol (TDMRep)⁶⁵ and – both in development – the IETF AI Preferences⁶⁶ as well as the Really Simple Licensing Protocol⁶⁷. They address some of the shortcomings of the Robots Exclusion Protocol highlighted above, for example by allowing granular rules and conditions for access by crawlers (e.g. on payment or restricting data usage to scientific purposes).

The European Commission launched a consultation on protocols for reserving rights from text and data mining under the AI Act and the GPAI Code of Practice⁶⁸ as well as published a study on the feasibility of a central registry of Text and Data Mining opt-out expressed by rightsholders.⁶⁹ Both confirm the need to monitor capabilities and practical enforcement of existing obligations and tools to enable rightsholders sufficient control of their content.

4. The impact of generative AI on internet openness

The rapid and significant uptake of generative AI services in the last five years has greatly benefitted a variety of different actors, including end consumers and business users. This technology can enrich user experience and complement traditional methods of accessing, generating and distributing information, content and services online in novel and promising ways.

⁶⁴RTR (2026), Text und Data Mining: Welche Botschaften senden österreichische Webseiten an KI-Anbieter? (Available only in German), see: <https://www.rtr.at/TKP/aktuelles/publikationen/publikationen/text-und-data-mining.de.html>

⁶⁵ See: <https://www.w3.org/2022/tdmrep/>

⁶⁶ See: <https://datatracker.ietf.org/doc/draft-ietf-aipref-attach/>

⁶⁷ See: <https://rslstandard.org>

⁶⁸ See: <https://digital-strategy.ec.europa.eu/en/consultations/commission-launches-consultation-protocols-reserving-rights-text-and-data-mining-under-ai-act-and>

⁶⁹ See: <https://digital-strategy.ec.europa.eu/en/library/new-feasibility-study-introducing-eu-level-registry-text-and-data-mining-opt-out> <https://digital-strategy.ec.europa.eu/en/library/new-feasibility-study-introducing-eu-level-registry-text-and-data-mining-opt-out>

However, generative AI services can also affect the overall user experience on the internet as they have the potential to directly influence the way users access online content and/or services and possibly their freedom of choice.

After presenting the principles of internet openness (Chapter 4.1), some preliminary elements of analysis are proposed to assess the impact of generative AI on the ability of users to access (Chapter 4.2), produce and distribute (Chapter 4.3) information, content and services online.

4.1. The principles of internet openness

Internet openness – as defined for the purpose of this report in Chapter 1 – is related to the principle of net neutrality. It means that internet endpoint connectivity via the internet access service must be open so that end users are not hindered when accessing and distributing content or using applications via the internet. The OIR reflects this by providing that “[e]nd-users⁷⁰ shall have 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, via their internet access service” (Article 3(1) OIR). The regulation “aims to protect end-users and simultaneously to guarantee the continued functioning of the internet ecosystem as an engine of innovation” and safeguard the internet “as an open platform for innovation with low access barriers for end-users, providers of content, applications and services and providers of internet access services” (Recitals (1) and (3) OIR).

The provisions of the OIR apply to internet service providers (ISPs). However, nowadays ISPs are no longer the only gateways capable of materially influencing how users access online information and services. At higher layers of the internet ecosystem, OS, browsers and app stores are also involved⁷¹. A variety of actors and applications, which are not providers of internet access services and thus not covered by the obligations of the OIR, are used to access the digital ecosystem and can structure, filter or prioritise users’ access to content and services available on the internet, thus exerting a power of intermediation. In that sense, these players can challenge the openness of the internet, understood more broadly than in the context of the OIR.

4.2. The impact of generative AI on users’ access to online content and services

As explained above, generative and agentic AI may significantly increase intermediation power whether offered as a stand-alone service or integrated into other digital services (e.g.

⁷⁰ For the definition of “end-user” see Chapter 1.

⁷¹ BoR (22) 167, BEREC Report on the Internet Ecosystem, 12.12.2022, see: https://www.berec.europa.eu/en/document-categories/berec/reports/berec-report-on-the-internet-ecosystem?language_content_entity=en

a search engine)⁷². A chatbot or the AI agent can retrieve information, generate answers in natural language and perform complex tasks, such as buying products on the user's behalf. Moreover, on some services, users will commonly be provided with an AI-generated content/summary as a result of their request, hinting at specific sources or with chatbot's access interface integrated in the search bar. This leads to a shift from search engines (structuring the available information by providing access to links, which are ordered depending on a set of parameters⁷³) to answering engines⁷⁴ which rather produce an *ad hoc*⁷⁵ answer in natural language that summarises the requested information. By creating *synthetic* content upon request and not only organising information, these services increasingly act as important channels to content and services available on the internet (gateways).⁷⁶

Generative and agentic AI can reduce search and transaction costs, and provide an alternative access to information and services. However, this intermediation can raise concerns in terms of internet openness, when the conditions under which content, sources or services are selected, ranked or accessed are insufficiently transparent, where users have limited ability to influence or override those choices, or where technical or commercial arrangements systematically favour certain providers or restrict access to alternatives.

This has important consequences for users' access to content and services.

4.2.1. From search engines to response engines

Search engines have already impacted users' access to online content by introducing an initial form of algorithmic ranking: by presenting results in the form of ordered lists, their ranking algorithm influence the visibility of content and focus attention on a limited number of domains. While the challenges and problems raised by search engines are targeted by existing regulation (see Chapter 5), the specific impact of generative AI on content distribution deserves to be monitored.

One specificity of generative AI relates to the role of users and the interactions with online services and content. Generative AI services have changed the structure of access to content

⁷² On the competitive functioning of the agentic AI sector, see Autorité de la concurrence (2026), Avis relatif au fonctionnement concurrentiel du secteur des agents d'intelligence artificielle (in French), see: https://www.autoritedelaconcurrence.fr/sites/default/files/integral_texts/2026-07/Avis_26a05_ADLC_compressed.pdf

⁷³ Some of them are known, such as the presence of the keyword in the text, or the number of links pointing to the resource (in the case of PageRank), some are not disclosed to the end-user.

⁷⁴ Ofcom (2025), The Era of Answer Engines: Generative AI's impact on search experiences and online safety, see: <https://www.ofcom.org.uk/internet-based-services/technology/the-era-of-answer-engines-generative-ai-impact-on-search-experiences-and-online-safety>

⁷⁵ The fact that generative AI tools produce ad hoc output, tailored for the users' request, should not undermine the fact that LLMs use existing data and content to produce it.

⁷⁶ The recent Google I/O presentation confirms this major change, for instance: Sarah Perez (2026), Google Search as you know it is over, TechCrunch, see: <https://techcrunch.com/2026/05/19/google-search-as-you-know-it-is-over/> By means of consequence, googling the word 'disregard' (to check the spelling for example) became impossible for some users: Russel Brandom (2026), You can no longer Google the word 'disregard', TechCrunch, see: <https://techcrunch.com/2026/05/22/you-can-no-longer-google-the-word-disregard/>

by replacing the logic of searching documents with keywords with that of synthesised answer with a limited number of sources which are structured, filtered and summarised by the model. Some authors observe that users are confronted with a trade-off: accepting lower transparency in exchange for access to personalized content that may reflect information, in particular when the user can describe the information they look for rather than name it (“tip-of-the-tongue” problem)⁷⁷. The internet user’s ability to understand and interpret the context of the information received may therefore be limited, due to the lack of transparency. These effects are reinforced by the default integration of generative AI features into digital services interfaces, which can limit users’ ability to control their use⁷⁸.

To illustrate this trend, recent studies also show that users on Google are less likely to click on the link to the source when the answer to their query is already provided in the AI Overviews summary⁷⁹. When users interact directly with a conversational agent, traffic to third-party sites tends to decrease (see Chapter 4.3.1), reducing the visibility of lesser-known domain names⁸⁰ (niche sources of content, for example, but also content produced by smaller organisations that do not have the same economic power than bigger ones). These effects on users’ behaviour can be amplified by the phenomenon of cognitive disengagement highlighted by academic literature⁸¹.

Another particular issue concerns the degree of user choice over the sources underlying generative AI responses. While traditional search engines typically present users with a list of results from which they can select, generative AI services provide a synthesised response based on a set of sources selected by the service. This reduces the immediate visibility of alternative sources compared with a traditional search interface. At the same time, conversational AI services may allow users to refine their queries, request additional or alternative sources and further specify the information they are looking for. The degree of user choice and control therefore depends significantly on the design and functionalities of the service, including how sources are presented and to what extent users can influence or refine their selection.

⁷⁷ Chirag Shah, Emily M. Bender (2024), [Envisioning Information Access Systems: What Makes for Good Tools and a Healthy Web? ACM Transactions on the Web](https://dl.acm.org/doi/10.1145/3649468), Volume 18, Issue 3, see: <https://dl.acm.org/doi/10.1145/3649468>

⁷⁸ Anaëlle Beignon, Thomas Thibault, Nolwenn Maudet (2025), Imposing AI: Deceptive design patterns against sustainability, see: <https://arxiv.org/pdf/2508.08672>

⁷⁹ Athena Chapekis and Anna Lieb (2025), Google users are less likely to click on links when an AI summary appears in the results, Pew Research Centre, see: <https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/>

⁸⁰ Athena Chapekis and Anna Lieb (2025), Google users are less likely to click on links when an AI summary appears in the results, Pew Research Centre, see: <https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/>

⁸¹ Among other studies, see notably: Matthias Stadler, Maria Bennet, Michael Sailer (2024), Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry, *Computers in Human Behavior*, see: <https://www.sciencedirect.com/science/article/pii/S0747563224002541>; also see: Nataliya Kosmyna, et al. (2025), Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task, see: <https://arxiv.org/abs/2506.08872>

The risk of lock-in and the concentration of sources in generative AI answers also raises crucial challenges, similar to those raised by search engines ranking algorithms. For instance, Arcep conducted, with the support of PEReN (the French Centre of Expertise of Digital Regulation), a series of tests on three major AI models and analysed the responses provided. The study shows that the service tends to rely on a limited number of sources, hindering the access to a diversity of information: 2% of the websites referenced account for 49% of the citations. 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 citations⁸². It should be noted that this study compared the sources distribution of generative AI tools with Google Search top results: the convergence is very limited.

When using generative AI services, and in particular AI agents, users are delegating⁸³ their choices: they let the agent select sources, summarise them or act on their behalf. With new “AI-powered” browsers such as Atlas or Comet, users are no longer navigating from link to link, but getting an entirely mediated access to the web and its content. This unconscious delegation is funded on trust: the user has to be confident in what the machine will do on their behalf. But as highlighted by Shah and Bender, among others⁸⁴, 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. For example, a new study by the New York Times suggests that AI Overviews is only answering correctly 90% of the time, which means one out of ten answers is wrong⁸⁵.

This outcome is due to the probabilistic nature of large language models: generative AI tools basically work by predicting the next most probable word in a sentence⁸⁶, which is not necessarily the right one.

In the field of AI, the term “hallucination” is used, among others, to refer to situations where an AI model’s output contains incorrect information yet presented as fact⁸⁷. While LLM have a tendency to “hallucinate” because of their probabilistic nature, factual “hallucinations” can be limited by retrieval augmented generation (RAG), therefore adding an external, authoritative knowledge source (e.g. documents). Besides, benchmarks for LLMs reward

⁸² Arcep (2026), Generative AI: challenges for the future of the open internet, see: https://www.arcep.fr/uploads/tx_gspublication/report-generative-AI-challenges-open-internet-january2026.pdf

⁸³ Bernhard Rieder (2006), *Metatechnologies and delegation: towards a society-oriented design in the Web 2.0 era*, Doctoral thesis, University of Paris VIII Vincennes-Saint Denis, see: <https://theses.hal.science/tel-00179980v1>

⁸⁴ Chirag Shah, Emily M. Bender (2024), Envisioning Information Access Systems: What Makes for Good Tools and a Healthy Web? ACM Transactions on the Web, Volume 18, Issue 3, see: <https://dl.acm.org/doi/10.1145/3649468>

⁸⁵ Tripp Mickle, Cade Metz, Dylan Freedman, Teresa Mondria Terol and Keith Collins (2026), How Accurate Are Google’s A.I. Overviews? New York Times, see: <https://www.nytimes.com/2026/04/07/technology/google-ai-overviews-accuracy.html>

⁸⁶ See: <https://time.com/6280533/ai-chatbots-improv-machines/>

⁸⁷ OECD.AI Policy Observatory, see: <https://oecd.ai/en/genai/issues/risks-and-unknowns>

guessing answers and disincentivise abstention (e.g. answering with “I don’t know”). New benchmarks might help to address “hallucinations”.⁸⁸ Other strategies include improved, complete training data and the development of mechanisms to validate and verify generated information through external sources.⁸⁹

4.2.2. Agentic AI: integration in and impact on digital services

The emergence of a single interface, as it is the case with generative AI services, is not only changing online visibility and attention dynamics regarding content, but it can also affect users’ access to services, especially in the context of agentic AI. AI agents can compare different services or offers and act on the user’s behalf, and thus reduce search and transaction costs. At the same time, this means that part of the selection process is delegated to and internalised within the AI service. Where the criteria governing such selection are not sufficiently transparent or controllable by the user, this can affect effective user choice and, depending on how the agent selects or integrates underlying services, the competitive opportunities of alternative service providers.

Indeed, agentic AI reinforces users’ delegation for a variety of activities, such as booking a journey, making a purchase, generating a document, submitting an administrative request. Rather than browsing multiple sites or choosing from a range of applications, users delegate some of these decisions to the AI agent when using this service. While such delegation can increase convenience and efficiency under some circumstances, it can also concentrate the user’s interaction within a single interface. This highlights the importance of the criteria and arrangements determining which underlying services are selected.

Depending on their design and on the technical and commercial integrations on which they rely, AI agents may influence how demand is distributed among different services. 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 interconnect 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 for internet openness.

While protocols such as MCP (see Chapter 3.2) aim to facilitate vertical interoperability, easing the connection with agents with the use of a standardised API, some challenges remain.

The design and operation of LLM-based systems may create a risk of unequal brand visibility. With the progressive shift of AI companies towards economic models supported by

⁸⁸ See: <https://iain.so/why-ai-models-hallucinate> and <https://openai.com/index/why-language-models-hallucinate/> and <https://artificialanalysis.ai/evaluations/omniscience>

⁸⁹ See: <https://www.rtr.at/rtr/service/ki-servicestelle/faq/FAQ.en.html>

advertising⁹⁰, AI training the proliferation of bilateral agreements and partnerships between AI providers and service and application providers and token representation⁹¹, ensuring transparency will become even more crucial to assess the potential impact of partnerships on e.g. favouring specific sources. 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 interface⁹². The payment network Visa also announced a collaboration with OpenAI⁹³. 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.

4.3. The impact of generative AI on users' incentives to produce and distribute content and services online

The impact of generative AI on visibility and traffic flows mentioned under Chapter 4.2.1 also affects content providers in the broad sense (e.g. websites, media outlets, independent creators), and their incentives to produce and share content and services online.

4.3.1. Economic and informational consequences of generative AI: challenges posed by mediated access to content

Challenges are heterogeneous and affect different areas, from the discoverability of sources for users to the sustainability of the production of online (quality) content which is often used by AI services without (fair) remuneration. A prerequisite for content producers to maintain and identify sustainable business models is sufficient control on how their content is used.

The question of discoverability, which can be defined as the “*online availability [of the content], and its ability to be found among a vast array of other content, particularly by someone who*”

⁹⁰ The introduction of advertising in AI companies' economic models has been analysed by researchers as a shift from the attention economy to the intention economy. See: Yaqub Chaudhary and Jonnie Penn (2024), Beware the Intention Economy: Collection and Commodification of Intent via Large Language Models, Harvard Data Science Review, Special Issue 5, see: <https://hdsr.mitpress.mit.edu/pub/ujvharkk/release/1>

⁹¹ See: Sonia Davidovic, Hervé Toupe (2026), How Agentic AI Will Reshape Payment, International Monetary Fund, <https://www.imf.org/en/-/media/files/publications/imf-notes/2026/english/insea2026004.pdf> and Amariles, D. R., Charlotin, D., & He-Guelton, L. (2026). AI Agents in Payments: Applications, Risks and Regulations. *European Journal of Risk Regulation*, 1–24. doi:10.1017/err.2026.10103 <https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/ai-agents-in-payments-applications-risks-and-regulations/C2EF22C4CD9513A9A6459D10680A5D12>

⁹² Reuters (2025), OpenAI partners with Etsy, Shopify on ChatGPT payment checkout, see: <https://www.reuters.com/world/americas/openai-partners-with-etsy-shopify-chatgpt-checkout-2025-09-29/>

⁹³ AP News (2026), see: <https://apnews.com/article/visa-chatgpt-openai-shopping-mastercard-d769dec86344cb4977c98789e8ec492f>.

*was not specifically searching for it*⁹⁴ is key⁹⁵. Classic search engines were relying on a mix of technologies, that, while being not fully transparent, still permitted content providers to adapt with Search Engine Optimisation (SEO) strategies. Nowadays, AI functionalities integrated into search ecosystems, such as Google AI Overviews and AI Mode, build upon traditional search infrastructure – including crawling, indexing, and ranking mechanisms – while adding retrieval and generative capabilities that synthesise information from online sources. The interaction between these components affects the discoverability of content in ways that differ from stand-alone generative AI services, which may rely on different retrieval, source-selection, and citation mechanisms. These differences in discoverability imply changes for content providers.

Some providers, such as newspapers, were already relying on search engines so their content would be discoverable to users: most of them let the web crawlers visit their websites in that regard. However, as crawlers increasingly started collecting data for different purposes (i.e. not only indexing content for search engines, but also fuelling training datasets for LLMs and retrieving up to date relevant information for AI generated responses, see Chapter 3.4), as well as relevant data for AI agents, new aspects of the way generative AI services transform discoverability online emerged.

In turn, LLM models have been pretrained on increasingly large and diverse corpora – often composed of hundreds of billions of tokens drawn from webpages, books, Wikipedia, news articles, source code and social media conversations which may be referred to also as digital commons. One example of these corpora, at least in the initial phase of chatbots' diffusion, has been considered the Colossal Clean Crawled Corpus (C4) launched initially by Google⁹⁶.

Over time, open catalogues of digitised collections offered by research centres, universities, museums and, more generally, associations defending public interests⁹⁷ have been experiencing growing disruption from the scraping action of robots, technically crawlers, which access collections in order to add them to a corpus of training data or to retrieve updated

⁹⁴ French Ministry of Culture (2020), Online discoverability of French-language cultural content, see: <https://www.culture.gouv.fr/thematiques/europe-et-international/Publications/Decouvrabilite-en-ligne-des-contenus-culturels-francophones>

⁹⁵ Including for example for cultural, academic or specialised content. See : Hubert Guillaud (2024), IA, réducteur culturel: vers un monde de similitudes, Dans les algorithmes (in French), <https://danslesalgorithmes.net/2024/12/03/ia-reducteur-culturel-vers-un-monde-de-similitudes/>

⁹⁶ Madhavi Singh and Fiona M. Scott Morton, A Roadmap for a Monopolization Case Against Google, Monopsony Power and AI Overviews, University of Pennsylvania Law Review, Vol. 175, p. 13, see: <https://www.antitrustinstitute.org/wp-content/uploads/2025/10/Singh-Morton-A-Roadmap-for-a-Monopolization-Case-Against-Google-Monopsony-Power-and-AI-Overviews.pdf>

⁹⁷ To the original perimeter of data commons institutions and bodies which included Galleries, Libraries, Archives and Museums (GLAMS), new commons such as urban commons (content related to public infrastructure, housing, social services) and digital commons (open software and other digital resources often protected under the Creative Commons license). See: Vasilis Avdikos, Mina Dragouni, Martha Michailidou and Dimitris Pettas (2024), Rethinking GLAMs as commons: a conceptual framework, Open Research Europe 2024, last updated 30 Oct 2025, see: <https://open-research-europe.ec.europa.eu/articles/3-157/v2>

relevant information⁹⁸: AI crawlers are scraping the web at an increasing rate, which sometimes is difficult to bear for these commons' infrastructure, thus disrupting access to end-users. As explained in Chapter 3.4, existing consent mechanisms based on technological solutions, such as the *robots.txt* protocol, are not always respected by all the stakeholders, generating inconsistencies and legal uncertainty⁹⁹. Websites hosting "data commons" are also exposed to new costs, risks, and externalities when exposed to continuous request of access to content (RAG queries): mitigating the growing afflux of bots can be done either by investing in infrastructure or by paying third-parties that offer DDOS (Distributed Denial of Service) protections.¹⁰⁰ These costs can discourage content creators or providers from sharing it online.

Moreover, as mentioned earlier, the integration of AI chatbots into search engines – such as Google AI Overviews and AI Mode – and, more broadly, the bundling of chatbot functionalities with search services – ChatGPT, Claude, Copilot, and Perplexity *et alia* offer search functions – is contributing to the so-called 'zero-click' phenomenon mentioned above. Generative AI systems and AI-enhanced search interfaces increasingly provide users with direct answers, summaries, or synthetic outputs, without requiring them to visit the original source. As a result, the share of news searches resulting in no click reportedly rose from 56% to nearly 69% between May 2024 and May 2025 in the USA, while search referral traffic to news publishers declined from more than 2.3 billion visits to less than 1.7 billion visits over the same period.¹⁰¹

⁹⁸ Are AI Bots Knocking Cultural Heritage Offline?, see: <https://www.glamelab.org/products/are-ai-bots-knocking-cultural-heritage-offline/> ; also Arcep (2026), Generative AI: challenges for the future of the open internet, p. 43, see: https://www.arcep.fr/uploads/tx_gspublication/report-generative-AI-challenges-open-internet-january2026.pdf

⁹⁹ Inconsistencies alongside the enforcement of opt out mechanisms based on protocols have been highlighted such as those allowed by the Robot Exclusion Protocol which is considered to "*place an immense burden on website owners to correctly anticipate all agents who may crawl their domain for undesired downstream uses cases*". Some authors suggest that an alternative scheme might give website owners control over *how* their webpages are used rather than *who* can use them, see: Longpre et al. (2024), Consent in crisis: The rapid decline of the AI data commons, p. 11, see: <https://arxiv.org/abs/2407.14933>

Currently the CMA decision mentioned above is a decision which tries to introduce more granularity for website owners in controlling the legitimate access to their content. Please see publisher conduct requirement here: <https://www.gov.uk/find-digital-markets-measures/google-search-publisher-conduct-requirement>

¹⁰⁰ Traditionally, a DDoS attack occurs when many computers (often a botnet) flood a website with requests, overwhelming its servers and making the site slow or unavailable, see: <https://www.ituonline.com/blogs/understanding-ddos-attacks/>

¹⁰¹ Similarweb, The Impact of Generative AI: Publishers, see slide number 3 titled "A new path to the headlines: GEN AI's growing role in news consumption" (last download 07 August 2026 h. 16.25) from the following link: <https://www.similarweb.com/corp/reports/the-impact-of-generative-ai-publishers-us/>. According to a European analysis published by the British Institute for Public Policy Research (IPPR), "*popular AI tools used by millions to access news are drawing on a narrow and inconsistent range of sources, often sidelining the UK's most trusted journalism and reshaping which voices are heard*", see IPPR "*Revealed: ChatGPT draws more on GB News, Al Jazeera, and Marie Claire than the BBC, IPPR analysis shows*" at the following link: <https://www.ippr.org/media-office/revealed-chatgpt-draws-more-on-gb-news-al-jazeera-and-marie-claire-than-the-bbc-ippr-analysis-shows>. For a comparison USA/EU related to AIO AI MODE see SearchEngineLand, D. Goodwin Article (2024) at the following link (last download and update 07 August 2026): <https://searchengineland.com/google-search-zero-click-study-2024-443869>, where stated: "A majority of Google searches – 58.5% in the U.S. and 59.7% in the EU – result in zero clicks. A zero-click search happens when users end their session or enter a new query without clicking on any results",

At the same time, the growing use of conversational AI tools for news access is accelerating such a pattern of consumption: news-related prompts on ChatGPT increased by 212% between January 2024 and May 2025, whereas equivalent Google searches declined by 5%.¹⁰² Hence, visibility no longer necessarily translates into traffic, audience engagement, advertising revenues, or subscriptions, and this shift undermines the current revenue models that historically financed the production of journalistic and creative content, particularly where monetisation depends on referral traffic, advertising impressions, or user conversion. More broadly, it may reduce incentives to invest in original human-generated content, even though such content remains essential for the training, updating, and overall quality of generative AI systems themselves.

The negative impact for publishers seems exacerbated by the fact that they do not seem to be allowed to opt out of the use of their content for AI-generated search functionalities such as Google AI Overviews and AI Mode, without losing visibility in traditional Google Search results.¹⁰³ On this point further analysis and deeper transparency is needed because it emerges that visibility is twofold: it includes visibility of website content to end-users (i.e. content in the list of websites on the right side of AI Overview or AI Mode texts) and visibility to retrieval systems which is essential for inference and grounding.¹⁰⁴

The possibility for very large private actors to exploit high-quality third-party content, without fair compensation nor effective opt-out mechanisms also affects media pluralism and the diversity of the internet. This diversity could be defined as a variety of opinions, sources and points of view being available and visible with no single intermediary or limited group of actors unduly determining which information becomes visible or influential. A recent survey which examines the performance of ChatGPT, Gemini, Claude, Google AI and Perplexity in five news areas (domestic policy, foreign policy, the economy, climate and migration) shows that AI news systems do not broaden the information environment in a systematic way. They repeatedly draw on a relatively narrow group of established sources which are usually quite

¹⁰² Similarweb, The Impact of Generative AI: Publishers, cit., slide number 1 under the same heading (last download and update 07 August 2026) at this link: <https://www.similarweb.com/corp/reports/the-impact-of-generative-ai-publishers-us/>

¹⁰³ Indeed, in the UK, the Competition and Market Authority (CMA), after having designated Google with Strategic Market Status (SMS) in general search services, explicitly concluded that publishers should be given a meaningful choice regarding the use of their content in AI search features and required Google to create opt-out mechanisms for AI Overviews and AI Mode. The CMA described this as a "world first" measure intended to strengthen publishers' bargaining position. See "CMA secures fairer deal for publishers and improves Google search services in UK" 2026, at: <https://www.gov.uk/government/news/cma-secures-fairer-deal-for-publishers-and-improves-google-search-services-in-uk>. More in detail: see the ad hoc Publisher conduct requirement, at: https://assets.publishing.service.gov.uk/media/6a1f0098b95db968c8f3bdb9/Publisher_CR_final_decision.pdf

¹⁰⁴ Technically, crawling is done by robots scanning the webpages, be it for indexing or for retrieving training content; these robots can declare their identity to the webserver there are retrieving the page from, so the webserver can block or limit some crawlers. Some crawlers can do both indexing and retrieving, and in some cases it may be difficult for content providers to identify the purpose behind their request. This raises some challenges as they may unwittingly block the indexing crawlers, while wanting to block the retrieving activities.. See also Arcep (2026), Generative AI: challenges for the future of the open internet: https://www.arcep.fr/uploads/tx_gspublication/report-generative-AI-challenges-open-internet-january2026.pdf

popular, distribute visibility very unevenly, and present different political and institutional source mixes, depending on the AI system and the topic.¹⁰⁵

Always in the remit of AI mediated access to news areas, on the side of enforcement actions, the Italian regulator AGCOM's decision No. 104/26/CONS (FIEG/Google case)¹⁰⁶, a referral to Article 65(2) DSA to the EU Commission,¹⁰⁷ highlights the risk that reductions in the visibility and discoverability of editorial content – particularly affecting smaller and independent publishers – may occur, as minority or less viral views are not “weighted”. Moreover, the criteria for selecting and ranking original sources remain unclear, resulting in the exclusion of many links that nonetheless contributed to generating the output.¹⁰⁸ These elements may have significant implications for freedom of expression, access to information, and media pluralism. Furthermore, AGCOM emphasises that the reliability and verifiability of information are weakened: AI-generated outputs may contain hallucinations, inaccuracies, or misleading summaries.¹⁰⁹ Against this backdrop, users – who generally do not understand how LLM operate (algorithmic design) and are not always aware of the limitations of these systems – often lack meaningful means to verify the veracity of AI-generated outputs or to determine the extent to which they faithfully reflect the underlying sources. By decision 127/06/CONS,¹¹⁰ AGCOM has established a roundtable bringing together Google, other platforms and publishers to discuss copyright, AI and media pluralism in order to shed light and find shared solutions in such a remit.

Furthermore, the shift from search engine optimisation to AI-oriented optimisation may also impact the way internet content is created since it would aim at being included in AI summaries and therefore cited as sources. Websites which are readable for humans could change to provide information intended solely for robots, for example in a specific format¹¹¹, that could not be easy to read for a human, ultimately making the content difficult to access for users.

¹⁰⁵ The study about the changing landscape of opinion formation. Vivien Benert and Stina Timmermann (2026), *Meinungsbildung im Wandel. Wie KI-systeme Quellen bei der Nachrichtensuche selektieren*, Agora Digitale Transformation GMBH Berlin, see: https://agoradigital.de/wp-content/uploads/2026/07/ADT_Studie_Meinungsbildung-im-Wandel.pdf

¹⁰⁶ AGCOM (2026), see: <https://www.agcom.it/comunicazione/comunicati-stampa/editoria-agcom-invia-segnalazione-alla-commissione-europea-sui>.

¹⁰⁷ Reuters (2026), see: <https://www.reuters.com/sustainability/society-equity/italys-media-regulator-asks-eu-investigate-google-ai-search-tools-over-publisher-2026-04-30/>

¹⁰⁸ Arcep (2026), *Generative AI: challenges for the future of the open internet*, see: https://www.arcep.fr/uploads/tx_gspublication/report-generative-AI-challenges-open-internet-january2026.pdf

¹⁰⁹ This phenomenon is well described and contextualised in several studies and papers, one of the most update is M.Huang and Others, “Answer bubbles: information exposure in AI mediated search” where Google AIO resulted to produce “*compressed content-heavy information optimised for quick consumption within a seach interface*” par. 5, p. 4, see: <https://arxiv.org/abs/2603.16138?utm>

¹¹⁰ AGCOM (2026), see: <https://www.agcom.it/provvedimenti/delibera-127-26-cons>

¹¹¹ Cloudflare proposed a “markdown for agents” format for example, see: <https://thenewstack.io/cloudflares-markdown-for-agents-automatically-make-websites-more-aifriendly/>.

4.3.2. Impact of synthetic content on internet diversity

While generative AI deployment can stimulate innovation and promote content, the significant challenges in terms of value distribution and the overload of synthetic creation can have a negative impact on content distribution and innovation online, resulting in eroding internet diversity.

Furthermore, while hindering incentives for content production and diffusion, generative AI services, are increasingly used for content creation, at a point where a significant part of the web now is occupied by synthetic content¹¹², including news and audiovisual platforms¹¹³. As more content consumed online is synthetic, the risk is that the more minor sources struggle more to be visible by end-users and to get ad revenues, which could have an impact on the diversity of the content produced on the internet. In addition, by increasing the amount of content available, it can make it more difficult to verify sources and assess the quality of online information.

According to prospective studies, this AI-slop dynamic could eventually contribute to an “artificialisation of the web”, as the stock of publicly available human-generated data suitable for AI training becomes increasingly constrained¹¹⁴. The theory of model collapse¹¹⁵ 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 term¹¹⁶. 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 user.

To conclude, generative AI, and especially retrieval-based and agentic AI systems, create a paradoxical dynamic: these systems depend on continuous access to high-quality online content and data while simultaneously weakening the economic incentives to produce and

¹¹² Next, Inquiry about AI generated newspapers_(in French), see: <https://next.innk/173214/recap-nous-avons-decouvert-des-milliers-de-sites-dinfo-generes-par-ia-tous-nos-articles/>, Next partnered with the French audience institute Mediametrie to make an observatory, see: <https://geste.fr/les-sites-dinformation-generes-par-lia-poursuivent-leur-progression-et-sont-desormais-consultes-par-pres-de-40-des-francais/>. On a more international scale, the same kind of phenomenon was observed on Medium, see: <https://www.wired.com/story/ai-generated-medium-posts-content-moderation/> and Google News, see: <https://www.404media.co/google-news-is-boosting-garbage-ai-generated-articles/>.

¹¹³ An Empirical Analysis of AI Slop in Music Streaming, Stanley Wu, Josephine Passananti, Viresh Mittal, Wenxin Ding, Haitao Zheng, Ben Y. Zhao, see: <https://oecd.ai/en/genai/issues/risks-and-unknowns>

¹¹⁴ Pablo Villalobos, Anson Ho, Jaime Sevilla, Tamay Besiroglu, Lennart Heim and Marius Hobbhahn (2026), Will We Run Out of Data? Limits of LLM Scaling Based on Human-Generated Data, see: <https://epoch.ai/publications/will-we-run-out-of-data-limits-of-llm-scaling-based-on-human-generated-data>

¹¹⁵ Ilya Shumailov, Zakhar Shumaylov, Yiren Zhao, et al. (2024), AI models collapse when trained on recursively generated data, Nature 631, p. 755–759, see: <https://www.nature.com/articles/s41586-024-07566-y>

¹¹⁶ Inria (2025), Towards a risk of collapse of generative AI?_(in French), see: <https://www.inria.fr/fr/risque-effondrement-collapse-ia-generatives>.

share such content. Future developments may therefore depend not only on compensation mechanisms, but also on the evolution and standardisation of technical governance tools enabling content providers to control different AI-related uses of their content, including training, retrieval, grounding, and AI-generated summarisation. This, in turn, may require greater transparency regarding how content is accessed, processed, and integrated into platform retrieval and generative systems.

5. Relevant regulatory frameworks to tackle the identified issues

The European Union has progressively developed a body of legislation that addresses many dimensions of the phenomenon under analysis: the concentration of intermediation power, the weakening of incentives to produce and share content, the transparency deficit of AI-generated outputs, and the erosion of user control over access to information. This chapter surveys the provisions of the existing EU regulatory frameworks that are most directly relevant to these concerns and that could help mitigate the risks to internet openness identified in this report.

The general normative backdrop for this analysis is the European Declaration on Digital Rights and Principles for the Digital Decade (2023/C 23/01)¹¹⁷, which in its Chapter III affirms the right of everyone to access a reliable, trusted, and diverse online environment. Although not binding as a directly enforceable instrument, the Declaration provides interpretive guidance for existing legislation and represents a political commitment that should inform future regulatory action. It underlines that the design of digital services must respect human dignity, autonomy, and access to a plurality of information sources, and that individuals should not be subject to algorithmic manipulation without transparency and effective recourse.

5.1. The Open Internet Regulation (OIR)

As mentioned in Chapter 4.1, although the Open Internet Regulation (OIR) establishes the principles of open access to the internet in general terms, its obligations fall only on providers of Internet access services (network connectivity). They do not reach content and application provision, where the intermediation power of generative AI services is exercised. The Commission's January 2026 proposal for the Digital Networks Act¹¹⁸ largely preserves this architecture, keeping the network-layer focus of the OIR rules on the connectivity between end-points intact. Consequently, the OIR does not constitute an operative tool for addressing the concerns analysed in this report.

¹¹⁷ European Commission (2022), see: <https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles>.

¹¹⁸ European Commission (2026), see: <https://digital-strategy.ec.europa.eu/en/policies/digital-networks-act>

Nevertheless, some of the broader policy objectives reflected in the OIR, in particular effective end-user choice, the ability to access and provide content, applications and services, and the preservation of the internet as an environment for innovation, provide a useful analytical reference when assessing the effects of new forms of intermediation at higher layers of the internet ecosystem. As these forms of intermediation evolve, the relevance and adequacy of the available instruments, including, where appropriate, the DMA, the DSA or the AI Act, may need to be assessed in light of the specific issues identified.

5.2. The AI Act

The purpose of the AI Act¹¹⁹ is to promote the uptake of human-centric and trustworthy AI, while ensuring a high level of protection of health, safety, fundamental rights, democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation. The legislation follows a ‘risk-based’ approach, which means the higher the risk to cause harm to society, the stricter the rules.

The AI Act establishes a specific regulatory regime for general-purpose AI (GPAI) models, which encompasses the LLMs underlying generative AI services. Article 53 of the Act obliges providers of GPAI models to maintain technical documentation, data requirements, transparency towards downstream actors and to comply with copyright law including through respect for rights reservations through state-of-the-art technologies (e.g. reservations expressed via machine-readable means such as the robots.txt protocol), and to publish summaries of the content used for training purposes. These provisions are directly relevant to the sustainability concerns of content creation examined in Chapter 4.3, as they aim to preserve the ability of publishers and other rights-holders to control the use of their works in AI training. However, their effectiveness depends on the robustness of the technical standards used to express such reservations and on the enforcement capacity.

In parallel, the Copyright in the Digital Single Market Directive¹²⁰ provides an exception or limitation to copyright and related rights for text and data mining (TDM), including for commercial purposes, notably through Article 4, while preserving the right of rights-holders to opt out in an appropriate manner, such as via machine-readable means. In practice, this opt-out is primarily exercised through the robots.txt protocol, which constitutes the main technical interface between content providers and AI developers. As identified in Chapter 3.4, this mechanism remains limited in its granularity, verifiability, and enforceability, resulting in a gap between the legal entitlement to opt out and the technical means available to exercise that right. The AI Act reinforces compliance with such reservations by requiring GPAI providers to identify and comply with them, including through state-of-the-art technologies. The GPAI Code

¹¹⁹ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), see: <http://data.europa.eu/eli/reg/2024/1689/oj>.

¹²⁰ Directive (EU) 2019/790, see: <https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng>

of Practice¹²¹, published in July 2025 as a voluntary tool under Article 56, further operationalises these obligations, including through commitments concerning robots.txt and other appropriate machine-readable protocols used to express TDM rights reservations.

In Article 4, the AI Act also provides measures for providers and deployers of any AI system, regardless of the risk classification, to support the development of AI literacy. The AI Omnibus¹²² amended¹²³ the AI Act and thereby, also requires the Commission and the Member States to support and facilitate the efforts of providers and deployers of AI systems in fulfilling their obligation. AI literacy is defined in Article 3 no 56 AI Act as *“AI skills, knowledge and understanding that allow providers, deployers and affected persons, taking into account their respective rights and obligations in the context of this Regulation, to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause”*

Additionally, Article 51 of the AI Act establishes that a GPAI model shall be classified as a GPAI model with systemic risk if either (i) it has high impact capabilities evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks, or (ii) based on a decision of the Commission, *ex officio* or following a qualified alert from the scientific panel, it has capabilities or an impact equivalent to those in point (i) having regard to the criteria set out in Annex XIII. Annex XIII refers to a list of criteria for the designation of general-purpose AI models with systemic risk referred to in Article 51, including whether it has a high impact on the internal market due to its reach, which shall be presumed when it has been made available to at least 10,000 registered business users established in the Union. All providers of GPAI models that present a systemic risk, be they open source or not, must comply with further obligations detailed in Article 55, such as performing model evaluations, adversarial testing, mitigating systemic risks, track and report serious incidents and ensure cybersecurity protections. These provisions could enhance the transparency and reliability of these models.

Standardisation supporting the implementation of the AI Act is also still developing. The first European standard developed in support of the AI Act, was approved by CEN and CENELEC in July 2026.¹²⁴ Subject to the publication of its reference in the Official Journal of the European Union, it is intended to provide presumption of conformity with Article 17 of the Act, which covers e.g. systems, techniques, procedures and actions to be used for the development, quality control, quality assurance, and data management of high-risk AI systems.

¹²¹ European Commission (2025), see: <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai>

¹²² Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI), see: <http://data.europa.eu/eli/reg/2026/1744/oj>.

¹²³ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), see the consolidated version of the AI Act, see: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A02024R1689-20260727>.

¹²⁴ EN 18286:2026: Artificial Intelligence - Quality Management System for AI Act Regulatory Purposes

Of particular relevance to the issues concerning user awareness are the transparency obligations laid down in Article 50 of the AI Act. These include, under the conditions set out in that provision, obligations to inform natural persons when they are interacting with an AI system and requirements concerning the marking and disclosure of certain AI-generated or manipulated content. Together with the broader AI literacy framework and the promotion of public awareness and understanding foreseen by the AI Act, these provisions may contribute to greater knowledge regarding the use, capabilities and limitations of AI systems and to more informed interaction with AI-generated or AI-mediated content.

The practical implementation of several of these provisions is, however, still at an early stage. The transparency obligations under Article 50, for instance, only became applicable on 2 August 2026. Moreover, important parts of the AI Act are not yet applicable, notably key requirements concerning high-risk AI systems, which will apply progressively from December 2027 and August 2028. At the same time, relevant standards, guidance and national implementation and enforcement structures are still being developed or operationalised. Further experience will therefore be necessary to assess the practical effectiveness of the existing framework and its contribution to addressing some of the issues identified in this report.

5.3. The Digital Markets Act

The DMA¹²⁵ is the instrument most directly relevant to the competitive dynamics that are impacted by the role of gatekeepers.

Several of the challenges described in Chapter 4 materialise precisely because generative AI is being embedded within services that are already designated as core platform services (CPS) under the DMA. Search engines, browsers, operating systems, and NI-ICS designated under the DMA are now integrating AI functionalities (e.g., AI Overviews, AI Mode, AI-powered browsing, in-app AI) in ways that reinforce the gatekeeping role of the companies concerned. Gatekeepers must comply with the obligations set out in the DMA also when they deploy AI in the context of their designated CPS¹²⁶. The DMA's prohibition on self-preferencing (Article 6(5)) is potentially relevant where a gatekeeper uses its existing CPS to privilege its own AI outputs over competing services or content providers. The obligation not to combine personal data from different sources without consent (Article 5(2)) constrains the data accumulation practices that underpin the training advantages of vertically integrated AI providers. The requirement to provide third-parties with fair and non-discriminatory access to ancillary services (Article 6(12)) is relevant to the platform dynamics discussed in Chapter 2.3, including

¹²⁵ 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 (Digital Markets Act), see: <https://eur-lex.europa.eu/eli/reg/2022/1925/oj/eng>

¹²⁶ European Commission (2024), High-Level Group for the Digital Markets Act Public Statement on Artificial Intelligence, see: https://digital-markets-act.ec.europa.eu/high-level-group-digital-markets-act-public-statement-artificial-intelligence-2024-05-22_en

the case of a designated gatekeeper for NI-ICS excluding competing AI chatbots from messaging infrastructure.

Lastly, designating cloud service providers as gatekeepers could also have an impact on the market for generative AI services¹²⁷: designated hyperscalers could be prevented by the provisions of the DMA from using their customers' AI services' data to fuel their own AI services in order to compete directly with their customers, as well as switching (including proportionate termination conditions) and interoperability obligations that could reduce lock-in effects in the AI infrastructure layer described in Chapter 2.2.

Whilst the implementation of the DMA to its full potential can positively affect the competitive dynamics in the market for generative AI services, this regulation cannot address the impact on internet openness at the core, even if generative AI services were to be designated themselves and hence made subject to obligations towards their upstream markets. Even in that scenario, DMA's provisions would not be sufficient to address the full range of concerns raised in this report: measures on fair remuneration of publishers and content-producers would be needed to ensure a viable and diverse content ecosystem, going beyond what competition law and the DMA can deliver. The DMA focuses on the competitive and economic relations among the concerned actors and does not intervene on the content or quality of AI outputs, nor does it directly address the disincentives for content production described in Chapter 4.3 or the transparency deficit experienced by users.

5.4. The Digital Services Act

The DSA¹²⁸ addresses the responsibilities of online intermediaries in relation to the content they host and the tools they are required to provide in order to ensure a safe, predictable and trusted online environment that facilitates innovation and where fundamental rights are respected. Two dimensions of the DSA are of particular relevance to the challenges identified in this report: the systemic risk obligations applicable to very large online platforms (VLOPs) and very large online search engines (VLOSEs), and the emerging question of whether generative AI services might themselves fall within the scope of DSA obligations¹²⁹. In addition, the DSA includes broader transparency and accountability obligations applicable to intermediary services, including those laid down in Articles 24 to 27, irrespective of whether they qualify as VLOPs or VLOSEs.

¹²⁷ In June 2026, the European Commission reached the preliminary position that Amazon's and Microsoft's cloud services AWS and Azure should be designated under the DMA, see: https://ec.europa.eu/commission/presscorner/detail/en/ip_26_1444

¹²⁸ 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 (Digital Services Act), see: <https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng>, <https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng>

¹²⁹ Some generative AI services can fall within the scope of the DSA, but not simply because they use AI. The key question is whether the service qualifies as an intermediary service, hosting service, online platform, or online search engine under the DSA.

VLOSEs and VLOPs designated under the DSA are required to assess and mitigate systemic risks arising from their services, including risks to the exercise of fundamental rights, media freedom and pluralism, and the protection of public health and safety (Article 34 DSA).

Some providers subject to DSA 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. AGCOM concluded (decision No. 104/26/CONS above) that potential risks to media pluralism and discoverability that are within the realms of possibility and touched upon in Chapter 4.3 (possibly including the zero-click phenomenon, the reduced visibility of smaller publishers, and the concentration of citation patterns documented by national regulators) can fall within the scope of Article 34 and Article 35).

The second dimension concerns the possible qualification of stand-alone generative AI services as intermediary services (particularly, online search engines) directly falling within the scope of DSA obligations¹³⁰. After OpenAI notified the Commission that ChatGPT had reached the active user threshold, which could potentially trigger a VLOP or VLOSE designation¹³¹, the European Commission designated ChatGPT as a VLOSE on 31 August 2026¹³². Consequently, ChatGPT must comply with the additional DSA obligations for VLOSEs, including assessing and mitigating systemic risks related to the dissemination of illegal content, negative effects on minors and users' physical and mental well-being, and threats to fundamental rights, electoral processes and public security.

5.5. The Data Act

The Data Act¹³³ (DA) is primarily concerned with access to and portability of data generated by connected products and related services, and with switching and interoperability obligations for cloud services, where authorities with experience in electronic communications services are considered to be well-placed to ensure the application and enforcement.¹³⁴ Its relevance to the concerns of this report operates mainly through its effects on competitive dynamics in the cloud markets, on which generative AI relies, but not directly on internet openness as such.

Chapter VI of the DA establishes obligations facilitating the switching of cloud service providers and the portability of data and applications across cloud environments. Since the AI infrastructure layer is dominated by a small number of hyperscale cloud providers with which

¹³⁰ In fact, a standalone generative AI model that merely generates responses to prompts and does not act as an intermediary between users and third-party content may fall outside the DSA.

¹³¹ Open AI's EU monthly active recipients for ChatGPT search, see: <https://help.openai.com/en/articles/8959649-eu-digital-services-act-dsa>

¹³² Commission designates ChatGPT, Reddit, Roblox under Digital Services Act, see: https://ec.europa.eu/commission/presscorner/detail/en/ip_26_1772.

¹³³ Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act), see: <https://eur-lex.europa.eu/eli/reg/2023/2854/oj/eng>

¹³⁴ Defined as "data processing services" under the Data Act.

AI model developers maintain exclusive or preferential partnerships, switching costs are structurally high. The DA's obligations reduce the dependence of AI developers on specific infrastructure providers, which can lower barriers to entry for new AI services, creating a more contestable infrastructure layer and reducing the risk of ecosystem lock-in described in Chapter 2.2. Also, some AI services might qualify as data processing services, and fall under SaaS category (for example, AI-as-a-service products) and therefore be directly subject to obligations under the DA easing the termination of contracts and the switching of providers for their users (data portability, interoperability via APIs, prohibition of unfair terms). These obligations would be particularly useful, for example, in the context of agentic AI, as it necessitates interoperable protocols between different AI services and third-party services.

Therefore, while the DA may affect competition dynamics, it would not specifically address the openness of generative AI services.

5.6. Coordination across instruments and enforcement authorities, and potential regulatory gaps

The risks to internet openness arising from generative AI do not belong neatly to a single regulatory framework but can be addressed by a variety of provisions under the different regulations. At the same time, several relevant provisions have only recently become applicable or are still being implemented. Priority should therefore be given to the effective implementation and enforcement of the existing framework and to monitoring its practical effects, also in order to establish a sound evidence base for assessing whether material regulatory gaps remain.

The AI Act, with its risk-based approach specifically covering AI, is the first comprehensive legal framework on AI. A systemic risk assessment under the DSA may identify media pluralism concerns resulting in specific mitigation measures. Additionally, the enforcement of copyright TDM opt-out mechanisms touches simultaneously on intellectual property law, AI Act compliance obligations, and the governance of technical protocols. The concentration of information access enabled by AI answering engines has implications for the DMA, competition law, media pluralism under the DSA, and fundamental rights¹³⁵ at the same time.

A coordinated enforcement landscape would prevent producing incoherence, duplication, and gaps that sophisticated market players are well positioned to exploit through regulatory

¹³⁵ Depending on its design and use, an AI-powered search tool may have implications for several rights protected by the Charter of Fundamental Rights of the European Union, including: i) Freedom of expression and information (Article 11), insofar as search tools shape access to information, prioritise certain sources, summarise content, or influence what users see and rely upon; ii) Protection of personal data (Article 8) and respect for private life (Article 7), where personal information is collected, processed, profiled, or incorporated into search results and generated responses; iii) Non-discrimination and equality (Articles 20 and 21), if algorithmic outputs reproduce or amplify biases; iv) Consumer protection (Article 38), particularly where AI-generated answers may be inaccurate, misleading, or insufficiently transparent; v) Human dignity and autonomy (Article 1), to the extent that users are influenced by opaque AI-generated outputs without understanding their provenance or limitations.

arbitrage¹³⁶. In this context, particular attention should also be given to the timely designation and operationalisation of the relevant national competent authorities (e.g. for the Data Act and the AI Act), as delays in this regard have already occurred in several Member States. Ensuring that these authorities have adequate resources to carry out their tasks effectively is equally important, as shortcomings in either respect may hinder the regulations' enforcement, as well as the cooperation between authorities, and lead to additional regulatory uncertainty to users and providers.

A further dimension concerns user agency. In the context of generative and agentic AI, while delegating information retrieval, selection or certain actions to an AI system can provide some benefits to users, it can also create risks related to overreliance, bias, reduced transparency, privacy, accountability. A key element of meaningful user agency is therefore that such delegation is sufficiently transparent and informed, so that users understand when they are relying on AI-mediated information or decisions and are aware of the relevant capabilities and limitations of the system. The transparency obligations under Article 50 of the AI Act, together with efforts to promote AI literacy, can make an important contribution in this respect, provided that they are effectively implemented and enforced. AI literacy may also help users actively mitigate some of the concerns outlined in this report, for example by requesting alternative sources or perspectives, critically checking AI-generated information, or making use of available settings and controls when configuring AI agents. Such user-side measures cannot, however, substitute for appropriate transparency and safeguards at service level.

Further experience will be important in assessing whether these existing provisions provide users with sufficient information in practice, including in relation to the way information sources are selected and presented. Where relevant information asymmetries or limitations to user choice remain, further consideration could be given to how users exercise meaningful control over relevant aspects of AI-mediated information retrieval and the choice of AI service provider.

In this context, similar consideration could be given to assessing the transparency of the criteria used by generative AI for the selection of information sources, as it would allow users to understand how this selection works and therefore assess the information that is presented in a more informed and potentially critical way, helping address bias, detect potential hallucinations, and information asymmetry. Specific transparency measures aiming at addressing how the credibility of the selected sources is assessed (e.g. recency vs authority trade-offs) could also be useful in this regard, along with establishing confidence indicators.

Such information would need to be presented in a user-friendly way to avoid information overload and confusion, as raw transparency does not always translate to meaningful understanding. Depending on the characteristics of the service and when appropriate, users may also benefit from meaningful, actionable choices, including the ability to access non-AI-mediated results, to consult original sources, and to use alternative providers without

¹³⁶ This is particularly relevant when different frameworks apply, as there may be additional difficulties that result from the definitions which in some cases do not total coincide across the regulations.

disproportionate switching costs. Publisher and content creator control over the use of their content requires, on top of the legal entitlements provided by copyright law and the AI Act, the technical infrastructure (standardised protocols, verifiable opt-out signals, and auditable compliance mechanisms) necessary to exercise those entitlements in practice. The governance of these technical standards is essentially a regulatory matter (as it is closely connected to the effective implementation of legal rights and public policy objectives), and public authorities have a legitimate and constructive role to play in multi-stakeholder forums at European and international level, helping to ensure that standards are interoperable, transparent, balanced, and capable of translating legal entitlements into practical and enforceable mechanisms.

Additionally, encouraging generative AI providers to disclose information on the reliability and limitations of their output (including through the collection of data on the reliability of the models and bias indicators), based on commonly established criteria and specific measuring rules, could also provide significant benefits. Approaches aimed at promoting diversity in source selection and avoiding excessive homogenisation could also be further explored.

One legislative development also merits attention. As the future Digital Fairness Act is expected to tackle several problems consumers face online, it could be a relevant tool to consider for the introduction of legal safeguards for users/consumers on the use of AI systems. However, with its scope being limited to consumers, it will likely only partially address the issues at stake.

The functioning of crawlers and exclusion protocols should also continue to be monitored, in particular to assess the practical effectiveness and enforceability of existing mechanisms enabling publishers and content creators to express their preferences regarding the use of their content. Further experience will also be important in assessing the interaction between such preferences and the visibility of content in search results. Where relevant shortcomings remain, consideration could be given to whether further measures are needed, taking into account the relevant obligations already introduced by the AI Act and their operationalisation through the GPAI Code of Practice.

In any case, a new monetising model for publishers/content creators – other than the current one, mainly based on traffic and advertising revenues – might be a road worth exploring. Central rights management registries, usage-based remuneration tied to model outputs and public-interest funding for high-quality information ecosystems could be alternatives worth discussing with stakeholders.

6. Conclusion

Generative and agentic AI are reshaping our digital activities and creating a new layer of intermediation within the internet ecosystem. While these technologies offer substantial benefits in terms of innovation, productivity and accessibility, they also risk concentrating

control over access to content and services, reducing the visibility and viability of content creators, and reinforcing existing market power.

Several key findings emerge from the analysis. First, the report finds that competitive dynamics in the AI ecosystem are characterised by strong structural dependencies and that distribution layers remain concentrated among a small number of Big Tech companies. Second, while technical protocols – such as MCP, WebMCP and robots.txt – may reduce barriers to entry, facilitate switching and support innovation by allowing a wider range of services to connect to AI systems, they also present some challenges and limitations, especially in terms of openness, enforceability and transparency, which highlights the need for further technical and governance developments.

Finally, two further key findings of this draft report concern the transformation of access to online content and services, and the growing tension between generative AI and the sustainability of the production of online (quality) content. Evidence reviewed in the report suggests that AI-generated answers can reduce traffic to original websites and concentrate attention on a relatively small number of sources, and that the rise of AI-powered browsers and agents reinforces these trends by allowing users to delegate increasing number of online activities to automated systems. Moreover, the draft report identifies a structural paradox: AI systems depend on the continued production of high-quality content while simultaneously weakening some of the economic incentives that support its creation and dissemination.

The European Union has progressively developed a body of legislation that addresses and tries to mitigate many challenges identified in this report: the concentration of intermediation power, the transparency deficit of AI-generated outputs, and the erosion of user control over access to information. However, the risks to internet openness arising from generative AI do not belong neatly to a single regulatory framework. Some aspects can be partly addressed by a variety of provisions under different regulations, such as the DMA, the DSA, the AI Act and the DA, that could be further implemented to reach its full potential.

Therefore, preserving internet openness in this evolving environment will require effective enforcement of existing legislation, closer coordination among regulatory authorities, robust transparency obligations, meaningful user choice, well-functioning interoperability frameworks and continued development of technical governance mechanisms that ensure a sustainable balance between AI innovation, competition, openness and the long-term health of the internet ecosystem. Equally important are effective mechanisms to protect publisher rights and support a diverse trustworthy digital environment.

Finally, to mitigate some of the identified risks, it is key to provide users with the information tools to enhance their control, understanding, and rights. In this respect, the transparency obligations under the AI Act, together with AI literacy, can make an important contribution, provided that they are effectively implemented and enforced. Regulatory initiatives (e.g. the above-mentioned regulations, as well as the Digital Fairness Act and others) and awareness campaigns aiming at promoting more informed and empowered online users can further play a relevant role in this context.

Annex: List of abbreviations

AAIF	Agentic AI Foundation
AI	Artificial Intelligence
AI Act	Artificial Intelligence Act
AGCOM	Autorità per le Garanzie nelle Comunicazioni
Arcep	Autorité de régulation des communications électroniques, des postes et de la distribution de la presse
AWS	Amazon Web Services
BEREC	Body of European Regulators for Electronic Communications
CPS	Core Platform Services
DA	Data Act
DDOS	Distributed Denial of Service
DMA	Digital Markets Act
DSA	Digital Services Act
EC	European Commission
ECN	Electronic communications networks
ECS	Electronic communications services
EECC	European Electronic Communications Code
GPAI	General-purpose AI
IETF	Internet Engineering Task Force
IoT	Internet of Things
IP	Internet Protocol
ISP	Internet service providers
LLM	Large language model
MCP	Model Context Protocol

NI-ICS	Number-independent interpersonal communication services
NRA	National Regulatory Authority
OIR	Open Internet Regulation
OS	Operating Systems
PEReN	French Centre of Expertise of Digital Regulation
RTR	Rundfunk und Telekom Regulierungs-GMBH
SDK	Software Development Kit
SEO	Search Engine Optimisation
TDM	Text and Data Mining
TDMRep	Text and Data Mining Reservation Protocol
VLOPs	Very Large Online Platforms
VLOSEs	Very Large Online Search Engines