

General-purpose artificial intelligence

General-purpose artificial intelligence (AI) technologies, such as ChatGPT, are quickly transforming the way AI systems are built and deployed. While these technologies are expected to bring huge benefits in the coming years, spurring innovation in many sectors, their disruptive nature raises policy questions around privacy and intellectual property rights, liability and accountability, and concerns about their potential to spread disinformation and misinformation. EU lawmakers need to strike a delicate balance between fostering the deployment of these technologies while making sure adequate safeguards are in place.

Notion of general-purpose AI (foundation models)

While there is no globally agreed definition of artificial intelligence, scientists largely [share](#) the view that technically speaking there are two broad categories of AI technologies: 'artificial narrow intelligence' (ANI) and 'artificial general intelligence' (AGI). ANI technologies, such as image and speech recognition systems, also called weak AI, are trained on well-labelled datasets to perform specific tasks and operate within a predefined environment. By contrast, AGI technologies, also referred to as strong AI, are machines designed to perform a wide range of intelligent tasks, think abstractly and adapt to new situations. While only a few years ago AGI development [seemed](#) moderate, quick-paced technological breakthroughs, including the use of [large language model](#) (LLM) techniques have since radically changed the potential of these technologies. A new wave of AGI technologies with [generative capabilities](#) – referred to as '[general purpose AI](#)' or '[foundation models](#)' – are being trained on a broad set of unlabelled data that can be used for different tasks with minimal fine-tuning. These [underlying models](#) are made accessible to downstream developers through application programming interface (API) and open-source access, and are used today as infrastructure by many companies to provide end users with downstream services.

Applications: Chat GPT and other general-purpose AI tools

In 2020, research laboratory [OpenAI](#) – which has since entered into a commercial [partnership](#) with Microsoft – released [GPT-3](#), a language model trained on large internet datasets that is able to perform a wide range of natural language processing tasks (including language translation, summarisation and question answering). In 2021, OpenAI released [DALL-E](#), a [deep-learning model](#) that can generate digital images from natural language descriptions. In December 2022, it launched its chatbot [ChatGPT](#), based on GPT-3 and trained on [machine learning](#) models using internet data to generate any type of text. Launched in March 2023, [GPT-4](#), the newest general-purpose AI tool, is expected to have even more [applications](#) in areas such as creative writing, art generation and computer coding.

General-purpose AI tools are now reaching the general public. In March 2023, Microsoft launched a new [AI-powered Bing search engine and Edge browser](#) incorporating a chat function that brings more context to search results. It also released a [GPT-4 platform](#) allowing businesses to build their own applications (for instance for summarising long-form content and helping write software). Google and its subsidiary [DeepMind](#) are also developing general-purpose AI tools; examples include the conversational AI service, [Bard](#). Google [unveiled](#) a range of [generative AI tools](#) in March 2023, giving businesses and governments the ability to generate text, images, code, videos, audio, and to build their own applications. Developers are using these '[foundation models](#)' to roll out and offer a flurry of [new AI services](#) to end users.

General-purpose AI tools have the potential to [transform](#) many areas, for example by creating new search engine architectures or personalised therapy bots, or assisting developers in their programming tasks. According to a [Gartner study](#), investments in generative AI solutions are now worth over US\$1.7 billion. The study predicts that in the coming years generative AI will have a strong impact on the health, manufacturing, automotive, aerospace and defence sectors, among others. Generative AI can be used in [medical education](#) and potentially in clinical decision-making or in the design of new drugs and materials. It could even become a key [source of information](#) in developing countries to address shortages of expertise.



Concerns and calls for regulation

The key characteristics [identified](#) in general-purpose AI models – their large size, opacity and potential to develop unexpected capabilities beyond those intended by their producers – raise a host of questions. Studies have [documented](#) that large language models (LLMs), such as ChatGPT, present ethical and social risks. They can [discriminate](#) unfairly and perpetuate stereotypes and social biases, use toxic language (for instance inciting hate or violence), present a risk for personal and sensitive information, provide false or misleading information, increase the efficacy of disinformation campaigns, and cause a range of human-computer interaction harms (such as leading users to overestimate the capabilities of AI and use it in unsafe ways). Despite engineers' attempts to mitigate those risks, LLMs, such as GPT-4, still pose challenges to users' [safety](#) and fundamental rights (for instance by producing convincing text that is subtly false, or showing increased adeptness at providing illicit advice), and can generate harmful and [criminal content](#).

Since general-purpose AI models are trained by scraping, analysing and processing publicly available data from the internet, privacy experts [stress](#) that privacy issues arise around [plagiarism](#), transparency, consent and lawful grounds for data processing. These models represent a challenge for [education](#) systems and for [common-pool resources](#) such as public repositories. Furthermore, the emergence of LLMs [raises](#) many questions, including as regards intellectual property rights infringement and distribution of copyrighted materials without permission. Some experts [warn](#) that AI-generated creativity could significantly disrupt the creative industries (in areas such as graphic design or music composition for instance). They are calling for incentives to bolster innovation and the commercialisation of AI-generated creativity on the one hand, and for measures to protect the value of human creativity on the other. The question of what liability regime should be used when general-purpose AI systems cause damage has also been [raised](#). These models are also [expected](#) to have a significant impact on the labour market, including in terms of [work tasks](#).

Against this backdrop, experts [argue](#) that there is a strong need to govern the diffusion of general-purpose AI tools, given their impact on society and the economy. They are also [calling](#) for oversight and monitoring of LLMs through evaluation and testing mechanisms, [stressing](#) the danger of allowing these tools to stay in the hands of just a few companies and governments, and [highlighting](#) the need to assess the complex dependencies between companies developing and companies deploying general-purpose AI tools. AI experts are also [calling](#) for a 6-month pause, at least, in the training of AI systems more powerful than GPT-4.

General-purpose AI (foundation models) in the proposed EU AI act

EU lawmakers are currently engaged in protracted [negotiations](#) to define an EU regulatory framework for AI that would subject 'high-risk' AI systems to a set of requirements and obligations in the EU. The exact scope of a proposed [artificial intelligence act](#) (AI act) is a bone of [contention](#). While the European Commission's original proposal did not contain any specific provisions on general-purpose AI technologies, the Council has [proposed](#) that they should be [considered](#). Scientists have meanwhile [warned](#) that any approach classifying AI systems as high-risk or not depending on their intended purpose would create a [loophole](#) for general purpose systems, since the future AI act would regulate the specific uses of an AI application but not its underlying foundation models.

In this context, a number of stakeholders, such as the [Future of Life Institute](#), have called for general-purpose AI to be included in the scope of the AI act. Some academics favouring this approach have suggested modifying the proposal accordingly. Helberger and Diakopoulos [propose](#) to consider creating a separate risk category for general-purpose AI systems. These would be subject to legal obligations and requirements that fit their characteristics, and to a systemic risk monitoring system similar to the one under the Digital Services Act (DSA). Hacker, Engel and Mauer [argue](#) that the AI act should focus on specific high-risk applications of general-purpose AI and include obligations regarding transparency, risk management and non-discrimination; the DSA's content moderation rules (for instance notice and action mechanisms, and trusted flaggers) should be expanded to cover such general-purpose AI. Küspert, Moës and Dunlop [call](#) for the general-purpose AI regulation to be made future-proof, inter alia, by addressing the complexity of the value chain, taking into account open-source strategies and adapting compliance and policy enforcement to different business models. For Engler and Renda, the act should [discourage](#) API access for general-purpose AI use in high-risk AI systems, introduce soft commitments for general-purpose AI system providers (such as a voluntary code of conduct) and clarify players' responsibilities along value chains.