ISSN 1831-9424

European Commission

Joint Research Centre

Generative AI Outlook Report

Exploring the Intersection of Technology, Society, and Policy

2025

EUR 40337

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC142598

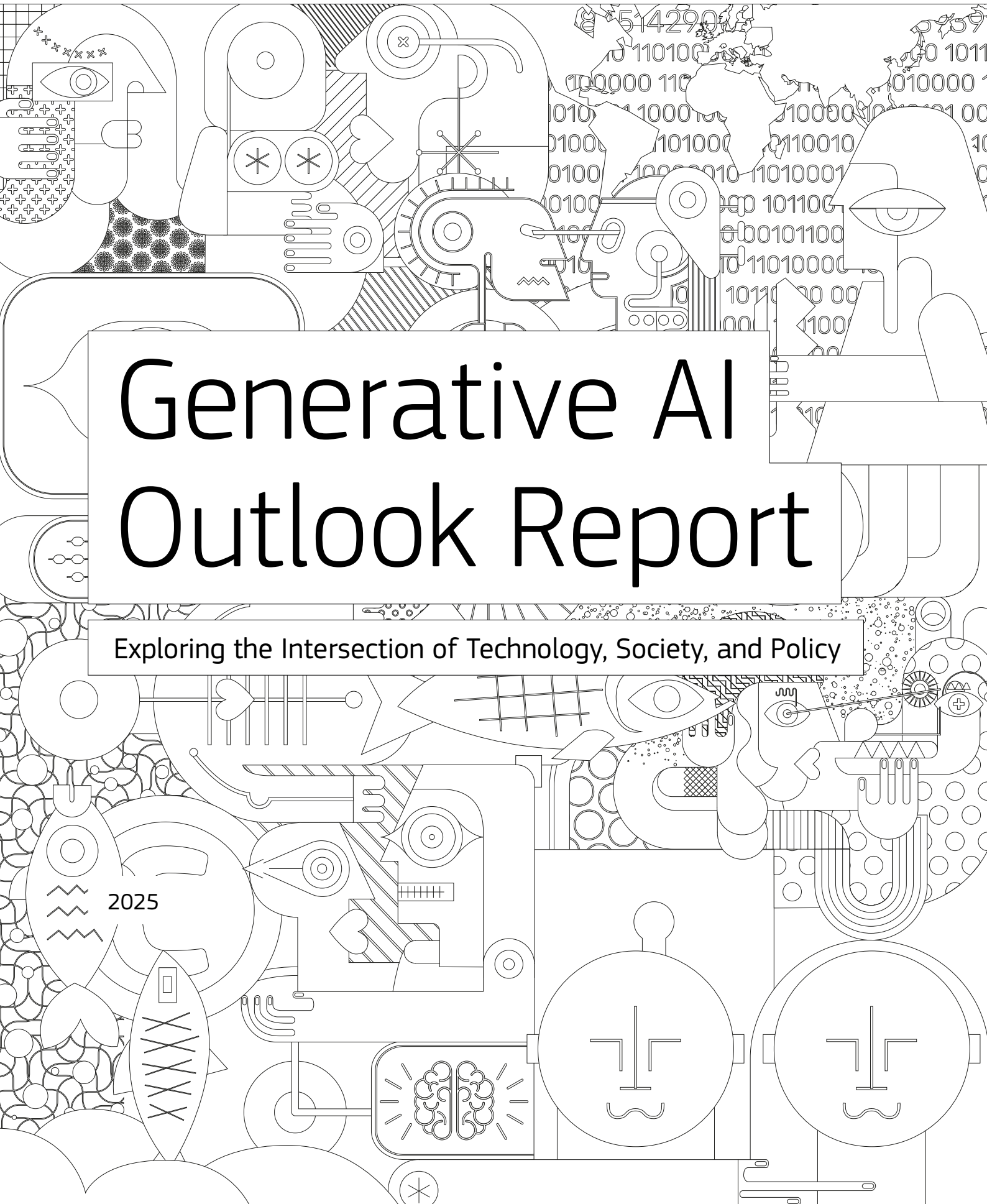
EUR 40337

Print	ISBN 978-92-68-28247-2	ISSN 1018-5593	doi:10.2760/0991238	KJ-01-25-309-EN-C
PDF	ISBN 978-92-68-28153-6	ISSN 1831-9424	doi:10.2760/1109679	KJ-01-25-309-EN-N

Luxembourg: Publications Office of the European Union, 2025

How to cite this report: Abendroth Dias, K., Arias Cabarcos, P., Bacco, F.M., Bassani, E., Bertoletti, A. et al., *Generative AI Outlook Report - Exploring the Intersection of Technology, Society and Policy*, Navajas Cawood, E., Vespe, M., Kotsev, A. and van Bavel, R. (editors), Publications Office of the European Union, Luxembourg, 2025, https://data.europa.eu/doi/10.2760/1109679, JRC142598.

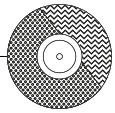
Layout: Carmen Capote de la Calle



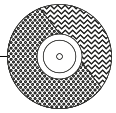
Generative AI Outlook Report

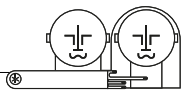
Exploring the Intersection of Technology, Society, and Policy

2025



CONTENTS





ABSTRACT

This Outlook report, prepared by the European Commission's Joint Research Centre (JRC), examines the transformative role of Generative AI (GenAI) with a specific emphasis on the European Union. It highlights the potential of GenAI for innovation, productivity, and societal change. GenAI is a disruptive technology due to its capability of producing human-like content at an unprecedented scale. As such, it holds multiple opportunities for advancements across various sectors, including healthcare, education, science, and creative industries. At the same time, GenAI also presents significant challenges, including the possibility to amplify misinformation, bias, labour disruption, and privacy concerns. All those issues are cross-cutting and therefore, the rapid development of GenAI requires a multidisciplinary approach to fully understand its implications.

Against this context, the Outlook report begins with an overview of the technological aspects of GenAI, detailing their current capabilities and outlining emerging trends. It then focuses on economic implications, examining how GenAI can transform industry dynamics and necessitate adaptation of skills and strategies. The societal impact of GenAI is also addressed, with focus on both the opportunities for inclusivity and the risks of bias and over-reliance. Considering these challenges, the regulatory framework section outlines the EU's current legislative framework, such as the AI Act and horizontal Data legislation to promote trustworthy and transparent AI practices. Finally, sector-specific 'deep dives' examine the opportunities and challenges that GenAI presents. This section underscores the need for careful management and strategic policy interventions to maximize its potential benefits while mitigating the risks. The report concludes that GenAI has the potential to bring significant social and economic impact in the EU, and that a comprehensive and nuanced policy approach is needed to navigate the challenges and opportunities while ensuring that technological developments are fully aligned with democratic values and EU legal framework.

FOREWORD

Ekaterina Zaharieva
European Commissioner For Startups, Research And Innovation

Can you tell if this foreword was written with generative AI? Today, this is an important question. The emergence of generative AI, with its power to create different types of content in a matter of seconds, is a revolution in the dynamic and disruptive landscape of digital technologies. Generative AI is evolving rapidly and is increasingly integrated into sectors well beyond traditional IT. It is redefining European industries and acting as a driver for innovation and economic growth, at an unprecedented pace.

There are **benefits** and new opportunities emerging before our eyes. In healthcare, generative AI can accelerate drug development, personalise patient care, or support early diagnosis. AI-driven cybersecurity solutions are helping to combat cyberattacks and detect misinformation narratives. Manufacturing sectors are revolutionising their business processes, with production tasks made fully autonomous.

In all these areas, Europe needs to ensure that we are not left behind by global competitors. We must equip our labour market to seize the advantages offered by AI by **upskilling and reskilling** our workforce. And we must create the conditions for EU-based generative AI startups to attract strategic investment to support their growth.

Of course, there are challenges. The likely productivity gains offered by generative AI have to be weighed against potentially negative consequences. Our creative industries now have to contend with new intellectual property issues linked to the use of AI. We must also **guarantee that AI supports rather than harms** young people, for example by avoiding over-reliance on AI-generated content in education, which could undermine critical thinking and lead to cognitive erosion.

I welcome this report's analysis of the interplay between technological innovation, societal needs, and policy responses. It gives much food for thought! It is a timely reminder that the development and deployment of generative AI are not only technical issues but require a **coordinated social and political approach**. Our common vision of making Europe a global AI leader, outlined in the AI Continent Action Plan, shows what our destination is. This report provides key scientific evidence to help us get there.

ACKNOWLEDGEMENTS

Ana Boskovic, Francesca Campolongo, Carmen Capote, Andrea Ceglia, Lorenzo Gabrielli, Miriam Giubilei, Eva Martínez, Pieter Kempeneers, Alberto Pena, Yves Punie, Francesca Siciliano, Tobias Wiesenthal.

The authors are very grateful for the comments and contributions by colleagues from Directorates General AGRI, CLIMA, CNECT, COMP, EAC, ECHO, EMPL, GROW, JUST, MOVE, RTD, and SG.

Authors

Kulani Abendroth-Dias, Patricia Arias Cabarcos, Manlio Bacco, Elias Bassani, Alice Bertoletti, Lorenzo Bertolini, Astrid Bertrand, Danai Bili, Philip Boucher, Romina Cachia, Mario Ceresa, Guillaume Chaslot, Stephane Chaudron, Valentin Comte, Cristian Consonni, Judith Cosgrove, Giuditta De Prato, François J. Dessart, Francesca Erica Di Girolamo, Stephanie Díaz, Néstor Duch-Brown, Anastasia Economou, Maria Eriksson, Josefina Fabiani, João Farinha, Eimear Farrell, Ana Fernández-Cruzado, David Fernández-Llorca, Roxana Fernández-Machado, Enrique Fernández Macías, Emilia Gómez, Claudius Benedict Griesinger, César Herrero, Juraj Hledik, Robert Jungnickel, Georgios Karopoulos, Sarah Klein, Alexander Kotsev, Bonka Kotseva, Kristina Kovacikova, Andraž Krašovec, Sarah Lemaire, Jens Linge, Montserrat López Cobo, Charles Macmillan, Anabela Marques Santos, Marco Minghini, Orsi Nagy, Igor Nai, Elena Navajas Cawood, Arman Noroozian, Daniel Nepelski, Daniele Paci, Andrea Pagano, Erasmo Purificato, Vittorio Reina, Theresa Reitis-Münstermann, Paula Rodriguez Müller, Arianna Sala, Ignacio Sánchez, Sven Schade, Mareike Sehrer, Alessandro Sellitto, João Soares da Silva, Josep Soler Garrido, Johan Stake, Gary Steri, Luca Tangi, Adeline Raluca Toader, Carlos Torrecilla Salinas, Juan Torrecillas Jodar, Jean-Paul Triaille, René Van Bavel, Michele Vespe, Daniel Villar Onrubia, João Vinagre.

Contributors

Gwendolyn Bailey, Michela Bergamini, Lorenzo Bertolini, Emiliano Bruno, Elodie Carpentier, Chiara Chiarelli, Anders Friis Christensen, Marco Combetto, Diego D'Adda, Margherita Di Leo, Olivier Eulaerts, Marcelina Grabowska, Isabelle Hupont Torres, Uros Kostic, Sandy Manolios, Jaume Martín Bosch, Irena Mitton, Antonia Mochan, Andrea Musumeci, Ilyas Tiouassiouine.

Editors

Elena Navajas Cawood, Michele Vespe, Alexander Kotsev, Rene Van Bavel.

Disclaimer: During the preparation of this work, the editors used GPT@JRC in order to support the integration process of the contributions by the authors to the final report, as well as to harmonise the style. After using this tool, the editors and authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

GPT@JRC is a platform offering secure access to a wide variety of pre-trained Large Language Models to assist with written office tasks and support scientific work. GPT@JRC is hosted local at the JRC data centre and is part of a JRC-wide study on the potential applications of this new technology within the European Commission. User prompts and the generated responses are not shared with third parties.

EXECUTIVE SUMMARY

The potential of Generative Artificial Intelligence (GenAI) is currently reshaping our socio-techno-economic landscape. This "Outlook Report" is designed to guide policymakers through the multifaceted implications of GenAI. It is meant to offer a forward-looking analysis of current trends, future scenarios and policy discussions that come with this transformative technology. By drawing on the latest scientific knowledge and expert insights of the Joint Research Centre (JRC), the report serves as a resource for policymakers across various policy areas, including digital technologies, employment, competition, environment, health, education, industry, justice and fundamental rights, to name a few. While the report does not claim to be a definitive research analysis, it delivers anticipatory research insights into current trends that can assist policymakers in exploring broader thematic areas. This approach can ensure that even if the technology evolves at a fast pace, policymakers receive a comprehensive overview across multiple domains.

Introduction – **Section 1**

GenAI is not merely a technological advancement as it represents a fundamental shift in how digital technologies intersect and shape our society and the economy. The emergence of GenAI, from its roots in academic research to its current status as a transformative technology, is driven by key technological enablers. The development of AI algorithms capable of processing and learning from large datasets, and the availability of high-performance computing coupled with advancements in deep learning architectures, have been instrumental for the emerging paradigm shift embodied by GenAI. The adaptive nature of this new technology allows applications across diverse domains; general purpose models that can be used for downstream use cases without retraining, for a wide range of tasks. The European landscape is uniquely positioned to leverage its robust research environment, characterised by networks of academic institutions and private innovators, to drive progress and foster the efficient adoption of GenAI. However, the competitive pressures faced by European GenAI start-ups highlight the need for strategic investment to support their growth.

Technological Aspects – **Section 2**

The technological landscape of GenAI is in continuous evolution, already seeing emerging capability trends such as Agentic AI, Multi-modal AI, and Advanced AI Reasoning. These advancements have the potential to boost productivity and enhance significantly decision-making and versatility across sectors but also pose challenges related to accountability, governance, and bias. The development of standardised evaluation methodologies is essential to develop trust in GenAI models, as we continue exploring the capabilities of these new systems and increase our understanding of limitations. Policymakers must reflect on these advancements to ensure ethical oversight and enforce standards for transparency and explainability in AI systems to help address the ethical boundaries of AI development and facilitate a sustainable integration of GenAI technologies.

Economic Implications – **Section 3**

GenAI can impact economic structures by driving industry transformation and the emergence of new business models. It is expected to deliver substantial productivity gains and foster job creation across various sectors. Digital maturity is crucial for GenAI adoption, especially for SMEs, which need to develop digital skills, business processes, and infrastructure. Employment policies must consider the labour market dynamics induced by GenAI, including impacts on income inequality, occupational restructuring, and shifts in demand for skills. Encouraging workforce resilience, adaptability and training will help address these changing needs.

Societal Impact and Challenges – **Section 4**

GenAI offers both opportunities and challenges for societal advancement. On the positive side, GenAI can drive more inclusive and equitable access to resources and opportunities, boosting creative skills or making complex analysis and knowledge accessible to a broader audience. However, it also raises significant considerations, such as over-reliance on and bias in AI-generated content. Policymakers must pay attention to these challenges to ensure responsible deployment, with particular attention to the risks of disinformation, mental health issues, deep fakes, and the societal biases perpetuated through AI outputs. The rapid adoption of GenAI also highlights a potentially significant skills gap, necessitating coordinated efforts from businesses, educational institutions, and policymakers to train, upskill and reskill the workforce. By adopting comprehensive strategies focused on fostering AI literacy, societies can better prepare their workforces and citizens to harness the potential of GenAI effectively.

Regulatory Framework – **Section 5**

The regulatory landscape in the EU plays an essential role in shaping the development and use of GenAI. The AI Act and the General Data Protection Regulation (GDPR) are central to this effort, promoting innovation while ensuring transparency, trust, and protection of safety and fundamental rights. The AI Act mediates the development of GenAI systems with legal requirements that make AI systems more transparent and trustworthy. The Digital Services Act (DSA) requires that systemic risks posed by Very Large Online Platforms and Search Engines, including those stemming from the use of GenAI, are duly assessed and mitigated. These regulations are also designed to foster technological innovation in areas specifically relevant to trustworthy AI, such as watermarking and fingerprinting techniques. Policymakers must continue to work on the details of the application of these frameworks to address emerging challenges raised by GenAI applications, for example in the areas related to intellectual property and data protection.

Sectoral examples of benefits and challenges brought by GenAI – **Section 6**

In-depth analyses within the report reveal the transformative potential of GenAI in specific sectors, alongside the need for careful management of the associated risks and ethical considerations.

Conclusions – **Section 7**

The report concludes highlighting the potential of GenAI to bring significant social and economic impact in the EU, and that a comprehensive and nuanced policy approach is needed to navigate the challenges and opportunities while ensuring that technological developments are fully aligned with democratic values and EU legal framework.

SECTORIAL EXAMPLES OF OPPORTUNITIES AND CHALLENGES BROUGHT BY GENAI

EDUCATION

OPPORTUNITIES

GenAI has the potential to redefine teaching and learning. This technology can help deliver more personalised learning experiences, adjusting the difficulty and nature of tasks based on a student's performance and interests. Likewise, it could democratise access to personal tutoring as well as enable, problem-solving, critical thinking and new of way of creativity.

CHALLENGES

There is a risk of over-reliance on AI for task completion and productivity gains rather than deeper conceptual exploration and learning. This could undermine critical thinking, problem solving, and the role of educators. Moreover, more research is needed to better understand the extent to which the use of GenAI can effectively enhance teaching and learning. Ensuring that AI tools are used to complement traditional teaching methods, rather than replace them, is crucial. Additionally, there is a need to safeguard against deceptive manipulation, bias and ensure the ethical use of AI in educational settings.

CYBERSECURITY

OPPORTUNITIES

GenAI has the potential to enhance threat detection and response capabilities. The usage of GenAIis can lead to more robust and proactive cybersecurity measures, benefiting both experts and ordinary users.

CHALLENGES

GenAI introduces complex cybersecurity challenges, including traditional threats and AI-specific vulnerabilities like data and model poisoning, adversarial attacks, and prompt injections. As AI systems become more embedded in cybersecurity, ensuring that they are used responsibly and do not introduce new vulnerabilities is essential.

CREATIVE INDUSTRIES

OPPORTUNITIES

GenAI is revolutionising content creation, enabling artists and designers to generate innovative works by analysing audience preferences and trends. This technology allows for the creation of AI-generated music, video, and art, fostering new business models focused on digital experiences.

CHALLENGES

One major challenge is the potential for homogenisation of styles, as AI models often rely on existing trends rather than creating entirely new ones. Significant concerns emerge about intellectual property rights, as AI-generated works may infringe on the creations of original artists.

SCIENCE

OPPORTUNITIES

GenAI is reshaping the scientific process by offering unprecedented efficiency and creativity, and allowing the development of novel approaches to support scientific work. It facilitates advancements by democratising access to scientific tools and fostering collaboration across disciplines, thereby accelerating research and innovation.

CHALLENGES

The integration of AI in science poses risks such as potential biases and the reinforcement of dominant narratives. Ensuring that AI tools are used to complement human expertise, rather than overshadow it, maintaining scientific integrity will be crucial.

HEALTH

OPPORTUNITIES

GenAI improves diagnostic accuracy and personalises patient care by analysing large datasets to detect patterns and predict disease progression. It supports early diagnosis and treatment planning, enhancing healthcare efficiency and empowering patients.

CHALLENGES

Data privacy and ethical use require careful attention. There are also concerns about data bias, the propagation of health inequities, and the potential deskilling of clinicians. Addressing these challenges requires responsible use within healthcare workflows and significant investments in IT infrastructure.

PUBLIC SECTOR

OPPORTUNITIES

GenAI has the potential to transform public sector management and service delivery by improving efficiency, transparency and responsiveness. AI-driven solutions can enhance decision-making processes, improve citizen engagement and optimise resource allocation, leading to better public services.

CHALLENGES

The adoption of GenAI requires effective governance and regulatory approaches to ensure safe, ethical, and lawful use. Ensuring transparency, accountability and oversight in AI systems is crucial to maintaining public trust and addressing potential biases.

1
INTRODUCTION

This chapter provides a fundamental understanding of Generative AI (GenAI), exploring its evolution from research to widespread adoption. It begins by examining the emergence of GenAI, highlighting key technological enablers and its impact across various sectors. The chapter then transitions to the current state of technology, identifying major players and technological advancements that define its landscape today. Attention is given to the infrastructures, data and models that underpin GenAI, crucial for its scalability and efficacy, while keeping safety and responsible use as priorities. Finally, the chapter underscores the strategic importance of GenAI for EU policymakers, discussing its potential to enhance the EU's digital sovereignty and competitiveness. Key issues addressed include the challenges of regulation, ethical considerations, and the socio-economic impact of GenAI.

1.1 The Emergence of Generative AI: From Research to Widespread Adoption

KEY MESSAGES

- The emergence of GenAI represents a paradigm shift in the field of artificial intelligence, characterised by the use of generative models to create text, images, or other types of content.
- From its roots in academic research to its current state as a transformative technology, GenAI continues to rapidly evolve, driven by technological advancements and a robust research ecosystem.
- As the EU and other global players navigate the opportunities and challenges presented by GenAI, a strategic and ethical approach will be essential to harness its full potential for societal and economic benefit.

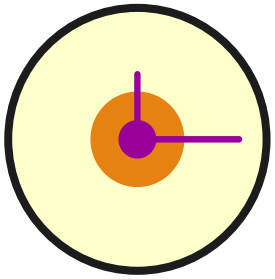
DEFINITION AND SCOPE OF GENERATIVE AI

GenAI refers to a class of artificial intelligence that focuses on the creation of new content, whether it be text, images, video, music or code. Unlike traditional descriptive or predictive AI models, GenAI models learn from vast datasets to generate original outputs that mimic human creativity. This capability has positioned GenAI as a transformative technology with applications ranging from healthcare to scientific research and beyond (see List of abbreviations and definitions).

HISTORICAL PERSPECTIVE

The development of GenAI can be traced back to foundational research in machine learning and neural networks. Initially, AI focused on rule-based systems and narrow applications. However, advances in computational power and algorithmic design have enabled the transition of large and deep neural network models to practical applications. This evolution reflects a broader shift of AI as a technology with widespread commercial and societal implications.

The ground-breaking nature of the technological revolution driving GenAI is evinced only by the astonishing pace of its commercial developments (see Section 4.3 for the extended version of the timeline as extracted from media outlets).



Timeline of main topics about media reporting peaks from mainstream news sources clustered around main GenAI themes using EMM1 (further analysis reported in Section 4.3 where the volume of outlets is also discussed and analysed over time).

JANUARY 2023

- ChatGPT and AI, with OpenAI's ChatGPT gain popularity for diverse applications
- Microsoft Invests in OpenAI and ChatGPT to challenge Google
- GenAI discussions at Davos

FEBRUARY 2023

- ChatGPT goes viral since its release, tech investors pour billions into it
- Tech Giants' race start with Google, Microsoft, and others competing with GenAI
- Meta creates new top-level product group focused on GenAI and releases LLaMA for AI research

MARCH 2023

- OpenAI's GPT-4 is released, concerns about biases, security and job displacement. Goldman Sachs predicts GenAI could automate 300 million jobs worldwide
- OpenAI, Google, and Microsoft invest heavily in AI development and application
- Microsoft integrates GenAI, into its Office apps; Google introduces GenAI features in Workspace apps

APRIL 2023

- Alibaba launches a ChatGPT-like AI model. Amazon launches Bedrock, a cloud service for GenAI, to compete with Microsoft and Google
- G7 nations discuss AI regulations, focusing on "ChatGPT" and GenAI, to ensure trustworthy technology.
- Elon Musk launches AIVenture"TruthGPT", despite calling for a pause in AI development, and purchased 10,000 GPUs to support the project

MAY 2023

- Japan discusses AI regulations, focusing on ChatGPT, to address concerns about misinformation, copyright, and data privacy, while exploring AI's potential benefits in education, business, and administration
- Rapid evolution of GenAI facilitates improved products while raising concerns about security, privacy, and potential risks
- Nvidia's stock rises due to increasing demand for its chips driven by the booming GenAI industry.
- Samsung and Apple ban employees from using GenAI tools like ChatGPT due to security concerns and risk of data breaches.
- Google integrates GenAI into search, ads, and products, enhancing user experience and advertising capabilities

JUNE 2023

- GenAI application in companies and for education
- Various companies partner with Google, Microsoft, and Nvidia to develop GenAI solutions
- The EP has agreed on stricter rules for AI, including a ban on biometric surveillance and requiring GenAI systems like ChatGPT to disclose AI-generated content

NOVEMBER 2023

- Sam Altman, CEO of OpenAI, was fired and then reinstated, causing shockwaves in the tech industry, with OpenAI's board citing lack of candor and investors pushing for his return
- Elon Musk's xAI Unveils Grok, a GenAI Chatbot to Rival ChatGPT
- First anniversary of ChatGPT

JANUARY 2024

- GenAI is transforming industries, with companies like Sony, Honda, and Microsoft investing in AI-powered services, and experts predicting significant growth in the AI market, with applications in fields like healthcare, finance, and education.
- Samsung partners with Google to integrate GenAI into Galaxy S24 series
- WHO warns of risks associated with GenAI in healthcare, despite its potential benefits in drug development and disease diagnosis
- Microsoft surpasses Apple as world's most valuable company due to its lead in GenAI

MAY 2024

- OpenAI launches GPT-4o
- Google integrates GenAI into Search Engine, aiming to improve search results and compete with Microsoft and OpenAI
- TikTok to Label AI-Generated Content
- AI transforms industries, workplaces, and lives, with GenAI revolutionising operations, client experiences, and health systems
- The EU adopts world-leading AI Regulation, aiming to harmonise AI rules, promote secure systems, and prevent disinformation, with most rules applying from 2026
- Google introduces Gemini

JUNE 2024

- Apple and Meta discuss integrating Meta's GenAI into Apple Intelligence, while Apple postpones its AI system launch in the EU due to regulatory concerns.
- Nvidia becomes world's most valuable company amid GenAI boom
- GenAI and cloud computing innovations
- Researchers develop new ways to detect errors in large language models, as tech companies integrate AI and LLMs into various applications

OCTOBER 2024

- Qualcomm, Google, and other tech companies are advancing GenAI - Impact, Applications, and Future
- OpenAI's ChatGPT has revolutionised AI, with new features and expansions, marking a significant shift in the tech industry.
- Researchers question large language models' ability to reason, despite advancements, highlighting limitations in mathematical reasoning and potential for errors
- The new digital economy is driven by Generative AI and virtual reality, sparking a social, ethical, and cultural revolution
- Apple introduces Apple Intelligence

JANUARY 2025

- AI and GenAI Trends and Applications 2025: improving efficiency, and enabling new applications, but also raising concerns about data security, job displacement, and ethical dilemmas
- The rise of China's AI rival to ChatGPT: DeepSeek developed a large language model rivaling US AI giants at a fraction of the cost, disrupting markets and sparking global interest.
- A US soldier used ChatGPT to plan an attack where he exploded a Tesla Cybertruck outside a Trump hotel in Las Vegas
- DeepSeek's low-cost AI model globally sparking market turmoil and concerns over America's dominance in AI

FEBRUARY 2025

- China's AI market is rapidly growing, driven by innovations like DeepSeek
- Amazon's new Alexa with GenAI Capabilities
- AI Summit in Paris
- Advances and Impact of DeepSeek and GenAI
- Multiple countries ban Chinese AI start-up DeepSeek due to security concerns and data privacy issues
- Governments and experts are working to regulate AI technology
- Studies show that over-reliance on GenAI can negatively impact critical thinking skills and cognitive abilities in knowledge workers and students

KEY TECHNOLOGICAL ENABLERS

Several technological advancements have played a crucial role in the rise of GenAI. Central to this is the development of AI algorithms that can effectively process and learn from large datasets. The introduction of large and computationally intensive deep learning architectures (e.g. the Transformer)[1], which enable models to understand context and generate coherent content, has been instrumental in enhancing GenAI's capabilities.

Infrastructural elements such as GPUs (Graphics Processing Units) and TPUs (Tensor Processing Units) have also been critical. These hardware components facilitate the intensive computations required for training and running large-scale GenAI models, making them indispensable to the technology's success.

Moreover, the availability of massive datasets has provided the raw material for training GenAI models. In conjunction with advancements in 5G connectivity, high-performance computing and the development of Large Language Models (LLMs) that enable realistic natural language communication, these datasets allow for the development and training of GenAI models at a larger scale, ensuring that the technology can meet the demands of diverse applications.

RESEARCH AND INNOVATION LANDSCAPE

The GenAI research landscape is characterised by networks of academic institutions and private sector innovators. China leads in the total number of academic publications related to the technology, with the EU ranking second, while facing funding gaps compared to other actors which affects the innovation potential (see Section 3.1). This research environment is supported by a network of universities, research institutions, and collaborative projects that drive innovation and knowledge sharing.

1. Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

The EU's emphasis on ethical AI and data privacy aims at a safe transition from research to widespread adoption. This focus ensures that GenAI applications align with European values, promoting trust and acceptance among stakeholders while also supporting competitiveness. This makes the EU approach distinctive, compared to that of other global players.

ADOPTION AND IMPACT

The widespread adoption of GenAI is evident across various sectors, including public administration, education, healthcare, and industry. In education, GenAI tools have the potential to transform teaching and learning processes when properly combined with adequate instructional methods. In healthcare, for example, GenAI aids in medical imaging and drug discovery, offering new possibilities for diagnosis and treatment, supporting patient empowerment and personalised medicine. These are analysed in Section 6.

The impact of GenAI extends beyond specific applications, influencing broader societal and economic dynamics. As a driver of innovation, GenAI presents opportunities for economic growth and job creation, while also posing challenges related to skills gaps and workforce displacement as discussed in Section 3.

CHALLENGES AND CONSIDERATIONS

Despite its potential, the emergence of GenAI is not without challenges. Ethical considerations, such as bias in AI-generated content and the need for transparency in AI decision-making, remain critical issues that must be addressed to ensure responsible deployment. Furthermore, the regulatory landscape plays a central role in shaping the development and use of GenAI. Policymakers must navigate complex considerations related to market dynamics, environmental impact, data protection, intellectual property, misinformation and disinformation, as well as the ethical implications of AI applications as expanded in Sections 3-5.

1.2 Current State of the Technology Key Players

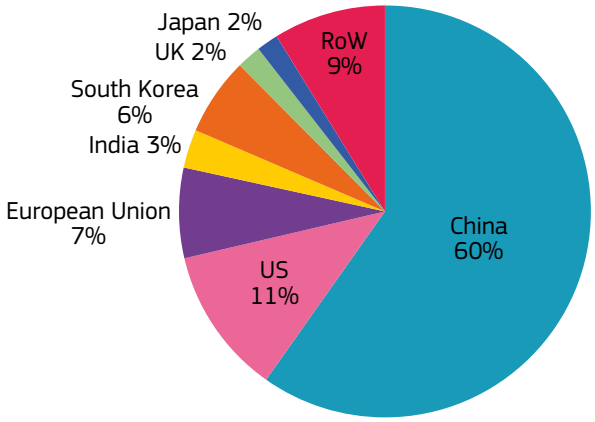
KEY MESSAGES

- The EU's strong research environment, ranking second globally in GenAI publications, provides a competitive advantage combined with the focus on ethical AI. However, funding and investment challenges might affect the potential for growth, with EU GenAI start-ups facing a significant venture capital funding gap compared to the US.
- By leveraging on its strengths and addressing these challenges, the EU can continue to play a leading role in the development and deployment of GenAI technologies.

GLOBAL DISTRIBUTION OF GENAI PLAYERS

GenAI activities increasingly account for a significant portion of the digital ecosystem, with over 72,000 players[2] engaged in more than 149,000 activities. Activities in this context refer to research publications, innovation (patenting), and business and investment activities. Figure 1 illustrates the competitive nature of the GenAI landscape and the regions that are at the forefront of its development. Regarding the number of players and activities, China leads with the highest share of players and activities, followed by the United States. EU GenAI players are roughly split between business (37%), innovation (33%) and research (31%) activities, with a higher proportion of research activity compared to the global share. As can be seen in Figure 1, the EU makes up 7% of global players, coming third to China (60%) and the US (12%). South Korea follows the EU closely with 6% of global players. The UK and Japan account for 2% of global players each according to the JRC DGTES Dataset. It should be noted however that while China leads in the number of GenAI players, the US remains the centre of global commercial innovation and deployment with companies such as OpenAI, Anthropic, Google DeepMind and Microsoft deploying GPT-4, Claude 3.5, and Gemini Ultra.

Figure 1. Global distribution of GenAI players 2009-2024.



Source: JRC DGTES Dataset.

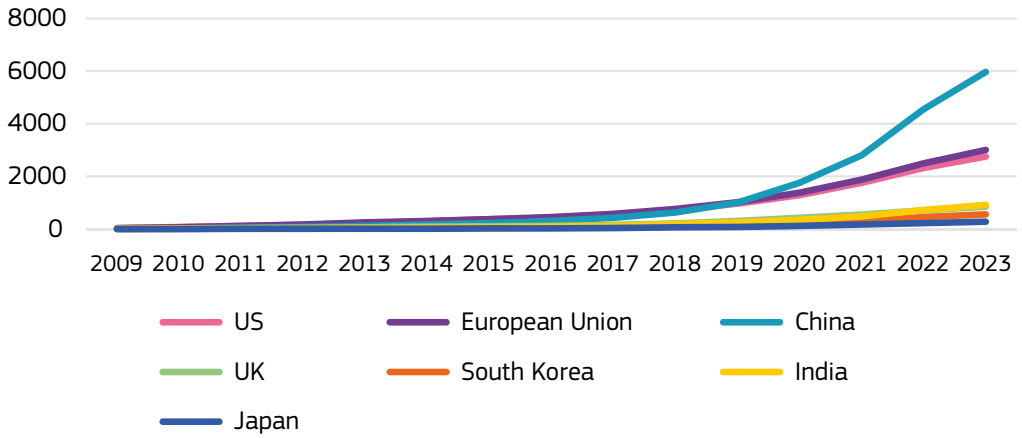
EU'S POSITION IN THE GENAI LANDSCAPE: RESEARCH AND INNOVATION

The EU maintains a strong position in GenAI, particularly in research and innovation. It ranks second globally in terms of academic publications on GenAI, highlighting its robust research environment (see Figure 2). These findings complement those by Renda et al. (2025)[3] showing that Europe precedes the US in terms of scientific publications, second only to China.

2. A player is an organisation that conducts research, innovates or has a business related to GenAI. Methodological information: Calza, E., et al., A policy oriented analytical approach to map the digital ecosystem (DGTES), Publications Office of the European Union, Luxembourg, 2022.

3. European Commission: Directorate-General for Research and Innovation, Renda, A., Balland, P.-A., Soete, L. and Christophilopoulos, E., *A European model for artificial intelligence*, Publications Office of the European Union, 2025, https://data.europa.eu/doi/10.2777/8034640

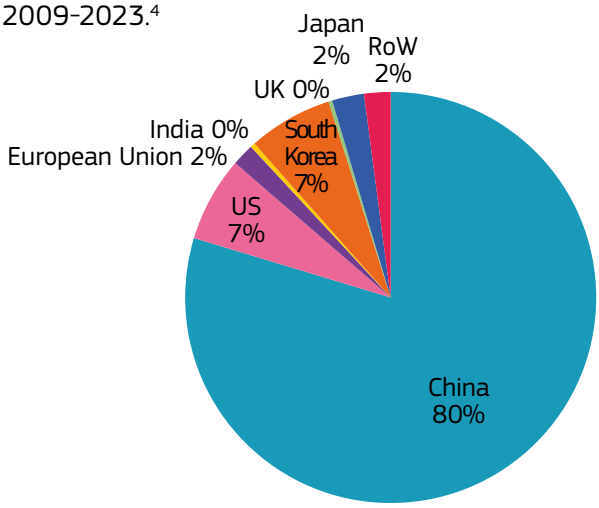
Figure 2. Research publications on GenAI in selected geographies 2009-2023.



Source: JRC DGTES Dataset.

EU research and innovation activities on Gen AI accelerated over the latest years, growing at an average rate of 32% annually between 2019 and 2021. Patents on GenAI grew exponentially over the last decade, accumulating over 120,000 filed patents by 2024. However, EU patent filings still comprise only 2% of global patent filings as shown in Figure 3, indicating a need for sustained investment in developing the GenAI innovation patenting ecosystem. The EU's position in innovation lags behind South Korea and the US, which have filed 7% and 6% of global patents respectively. Zooming in on the EU, 33% of players engaged in patent activities are located in Germany, followed by France (12%), the Netherlands and Spain (9%). Note that for the purposes of this analysis, we focus on priority patent filings, which refer to the first patent application filed for an innovation. A priority patent filing establishes a priority date, i.e. an official date from which the novelty and originality of the invention can be claimed. The use of priority patent applications is considered one of the most effective ways to account for innovation activity, because they represent the first step in seeking protection for an invention, avoid the risk of over-counting when the invention's protection is sought for different markets, and track it with less delay that the granted patent.

Figure 3. EU GenAI priority patent applications as a share of global AI priority patent applications 2009-2023.[4]



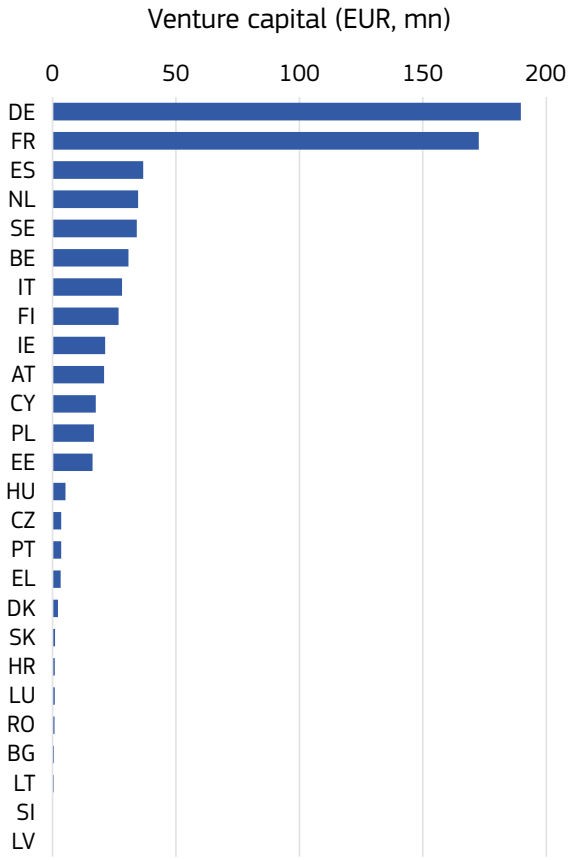
Source: JRC DGTES Dataset.

CHALLENGES IN FUNDING AND INVESTMENT

European GenAI start-ups face challenges in securing funding, with venture capital (VC) investment in US companies being significantly higher. This disparity highlights the competitive pressures faced by EU-based start-ups and the need for strategic investment to support their growth. While Germany and France have attracted sizeable amounts of VC over time (see Figure 4), more robust investment is needed to foster and further develop a vibrant EU GenAI ecosystem.

4. China's significant volume of generative AI patent filings underscores its commitment to technological leadership, although it also highlights the importance of emphasising high-quality, priority patent applications that reflect substantive innovation and international relevance.

Figure 4. Total amount (in million EUR) of VC related to GenAI received by EU country 2009-2024.



Source: JRC DGTES Dataset.

Balancing these considerations with the need for innovation and competitiveness, together with the gap of venture capital investments with respect to other areas at global scale is a key challenge for European GenAI players.

OPPORTUNITIES FOR GROWTH AND INNOVATION

The increasing demand for AI-driven solutions across sectors presents significant opportunities for GenAI. As industries seek to harness the power of AI to improve efficiency, productivity, and creativity, GenAI players are well-positioned to capitalise on these trends and deliver innovative solutions.

Cross-border collaborations and partnerships offer additional avenues for growth. By leveraging the diverse expertise and resources available across the EU, GenAI players can enhance their capabilities and expand their reach in the global market.

1.3 The Foundations for GenAI: Infrastructures, Data and Models

KEY MESSAGES

- It is essential to consider the challenges and opportunities associated with data, infrastructures, and models in each domain collectively to promote a balanced AI development. AI Factories alongside Common European Data Spaces serve as a flagship initiative that can integrate these critical components in a coordinated and trustworthy manner while interconnecting fragmented data infrastructures and establishing governance approaches grounded in EU values and existing legal frameworks.
- As GenAI continues to evolve, it will be crucial to address challenges related to data governance, interoperability, privacy, and computational capabilities to unlock its full potential. Moreover, open source models should be prioritised to enhance innovation, transparency, and explainability.
- It is critically important to enhance the understanding and progress towards AI-ready data. Additionally, investing in data visiting approaches, which involve moving algorithms instead of data, can help alleviate network pressure caused by exchanging large datasets.
- Increasing energy demands and cybersecurity are among the most critical challenges facing the development of GenAI models in the near future. Investigating AI energy efficiency, particularly through the development of smaller models and the use of specialised hardware, is essential to addressing these issues.

THE ROLE OF DATA IN GENAI

Data are the lifeblood of GenAI, serving as the primary input for training, refining and validating AI models. The ability of GenAI to generate new content hinges on its capacity to learn from extensive datasets, which provide diverse and rich information, such as text, audio and images, thus calling for the development of multimodal foundation models. That is why, access to high quality, diverse datasets is a crucial determinant of the effectiveness and competitiveness of GenAI applications. The sheer volume of data raises important questions about data availability, accessibility and management. While it may be challenging to fully eliminate bias and define representativeness, being aware of bias and developing methods to measure it are important steps in improving data quality and relevance.

DATA VOLUME, INTEROPERABILITY AND ACCESSIBILITY

The rapid advancement of the first generation of GenAI, particularly Large Language Models (LLMs), has been largely driven by training on vast amounts of user-generated content available on the internet. However, much of the existing web-based public domain data has already been utilised in current models, and AI-generated content is rapidly spreading into various areas. This is problematic, as an increasing body of evidence (e.g., Shumailov et al., 2024)[5] shows that models can collapse or their performance can deteriorate rapidly if trained repeatedly on AI-generated data. While training on such data is not inherently problematic, the real issue is the progressive distribution shift that occurs as a result, as it reduces the model's ability to accurately predict low-probability events. Those, in turn, are crucial for addressing questions related to minority or marginalised groups but also for complex system predictions required in industrial applications of LLMs. As AI-generated content increasingly saturates the internet, a pressing challenge is how to identify data produced by GenAI and prevent them from contaminating training loops or at least manage their integration in a manner that avoids poisoning. Finally, addressing issues related to data provenance would positively impact multiple AI applications. This is of significant societal importance, for example, ensuring data provenance can help mitigate the propagation of disinformation.

All of this, combined, underscores the need to explore new "unseen" datasets that can be integrated into upcoming GenAI applications. Such access would enable the development of novel applications and ultimately provide a competitive advantage to enterprises and nations that capitalise on this opportunity. At the same time, the data landscape in Europe is very fragmented, which imposes further challenges related to the interoperability and accessibility of data that need to be addressed through common interoperability standards, technical building blocks and governance approaches. Aligning data sharing practices to adhere to the FAIR principles,[6] i.e. ensuring that data are Findable, Accessible, Interoperable, and Reusable, becomes even more prominent in a fragmented digital ecosystem. In response to those challenges, the Common European Data Spaces being developed within the broader context of the forthcoming European Data Union Strategy aim to address the technical and organisational aspects related to the sharing of heterogeneous data at scale.[7]

DATA PRIVACY AND SECURITY CONCERNS

As GenAI applications become more widespread, the processing of personal and sensitive data requires robust safeguards to protect individual privacy and prevent unauthorised access. The EU's emphasis on data privacy, as exemplified

5. Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. Nature, 631(8022), 755-759.

6. https://www.nature.com/articles/sdata201618

7. Farrell, E., Minghini, M., Kotsev, A., Soler Garrido, J., Tapsall, B., Micheli, M., Posada Sanchez, M., Signorelli, S., Tartaro, A., Bernal Cereceda, J., Vespe, M., Di Leo, M., Carballa Smichowski, B., Smith, R., Schade, S., Pogorzelska, K., Gabrielli, L. and De Marchi, D., European Data Spaces - Scientific Insights into Data Sharing and Utilisation at Scale, EUR 31499 EN, Publications Office of the European Union, Luxembourg, 2023, ISBN 978-92-76-53522-5, doi:10.2760/400188, JRC129900.

by Regulations such as the GDPR, highlights the importance of addressing these concerns in the context of GenAI (see Section 5).

Privacy-preserving techniques, such as differential privacy and federated learning,[8] offer potential solutions to mitigate privacy risks while allowing data to be used effectively in model training. These approaches enable the extraction of insights from data without compromising individual privacy. Alternatively, GenAI can be used to produce synthetic data that can be used to train and improve models. Such data are created to mimic important statistical properties of the original data while ensuring the protection of user privacy, making them a useful tool for training models while protecting user privacy. While the generation and use of synthetic data require careful precautions to avoid capturing confidential information and to minimise any misrepresentation of original data, there are certainly advantages and potential in their use.

COMPUTING CAPACITY, NETWORK AND CONNECTIVITY CONSIDERATIONS

The computational demands of GenAI are substantial, necessitating advanced infrastructures capable of supporting the training and deployment of AI models. Scalable computational environments, high-speed connectivity, and efficient data storage solutions are all critical components of the GenAI infrastructure.

The success of GenAI relies on the availability of powerful hardware, such as GPUs and TPUs, which facilitate the intensive computations required for model training. The expansion and optimisation of data centre infrastructures is therefore essential to supporting the growth of GenAI. On the other hand, this increasing need for computational power has raised strong environmental concerns, given the vast amounts of energy required (see Section 4.5).

A recently published list of the top 500 AI supercomputers[9] indicates that the EU hosts around 50 of these machines, while the US hosts 134 and China a little more than 200. Altogether, they represent 80% of the total. However, the US dominates in terms of computational performance with recently deployed supercomputers, the most advanced of which, the xAI Colossus, incorporates up to 200,000 AI chips alone, representing more than the entire EU combined (with only 122,000 reported by the same source). The same source provides an estimate of the costs of deploying the hardware required for the operation of these supercomputers. There is a close relationship between performance and costs, suggesting that if the EU wants to reduce the gap in the computational performance vis-à-vis the US and China, significant investments will be required.

In addition to computational power, high-speed, low-latency networks are essential for real-time AI interactions and the seamless exchange of data between distributed systems. The deployment of Next Generation Access (NGA) – a fibre-based high-speed broadband infrastructure, 5G and the upcoming 6G networks, in particular – enhances the ability of GenAI applications to operate efficiently and effectively across diverse environments. The GÉANT network is a leading example of a high-bandwidth network interconnecting research and education networks to support, among others, AI development.

Finally, the movement and sharing of data between systems can be resource-intensive, necessitating strategic approaches to data management. Options such as edge computing and data condensation techniques can help mitigate the costs and inefficiencies associated with data transfer, enabling more efficient use of network resources.

8. European Commission, Joint Research Centre. Bacco, M., Kanellopoulos, S., Di Leo, M., Kotsev, A., Friis- Christensen, A., Technology Safeguards for the Re-Use of Confidential Data, European Commission, Ispra, 2025, JRC141298.

9. Investigating the trajectory of AI for the benefit of society https://epoch.ai/

MODEL COMPLEXITY AND SIZE

GenAI models have shown impressive capabilities at the cost of impressive computational intensity. Capabilities, or **model complexity**, which could be described as the potential to learn and represent increasingly complex relationships in data, typically grow with the number of (hyper) parameters that models have. The number of parameters (or **model size**) has been doubling every six months (Scaling Law), following an exponential trajectory. Nowadays, the size of available models is in the order of hundreds of trillions of parameters in the case of large models, and of billions for smaller ones.

The performance of a GenAI model does not depend entirely on its size. Its architecture, training techniques, and the quantity and quality of training data play a major role too. That is why the EU is investing in high performance computing (HPC) and gigafactories to support the development of AI models in hubs that can provide enough computational power and access to data enabled by the Common European Data Spaces currently in the making.

The EU is investing in Data Labs to make sure that high-quality data are available for AI training from diverse sources. Being able to minimise the amount of data needed for training is a key objective, and dataset condensation[10] represents a class of techniques used to generate a small synthetic training set from a large one. An advantage of condensation is that the confidentiality of original data is preserved because synthetic data are generated as output. However, condensation is extremely challenging in terms of computation, limiting its applicability at present when it comes to very large datasets. As a complement to condensation, it is worth highlighting that the paradigm of data visiting has the potential to minimise the need for large data transfers – which may also be needed for training purposes – because, instead of data, it is algorithms that are moved.

As an alternative, small models have attracted lots of attention because they have the advantage to require less computational resources, which is rather important when it comes to running models at the edge.[11] Edge devices have limited memory and computational capacity if compared to large cloud servers, thus small models have the potential to further spread the adoption and use of GenAI models. Distillation is the technique used to fine-tune (teach) a small model (student) using a large model (teacher) as a reference, so that tasks - often rather specific - can be carried out at a much lower cost.

OPEN SOURCE VS. PROPRIETARY MODELS

The choice between open source and proprietary models is a significant consideration for GenAI. By their very nature, open source models are customisable and adaptable, allowing developers to build on existing frameworks and tailor the models to specific needs. Open access to the underlying code ensures transparency and explainability, in turn promoting ethics, accountability and reproducibility, while minimising the time needed to identify and mitigate security risks. In addition, by removing licensing costs (as is often the case with open source), open source GenAI models are highly accessible (e.g. to individuals, research organisations and SMEs) and avoid vendor lock-in. These core features of open source resonate well with EU values, such as democracy, collaboration, inclusivity and transparency.[12] Not surprisingly, the potential of open source to power AI model development in the EU, contributing to its brand of open innovation, was highlighted in the recent AI Continent Action Plan.[13] While the open-source development model offers numerous advantages, there are important factors to consider regarding strategic autonomy and the need to ensure that open source businesses can expand as rapidly

10. Kim, Jang-Hyun, et al. "Dataset condensation via efficient synthetic-data parameterization." *International Conference on Machine Learning*. PMLR, 2022.

11. Meuser, Tobias, et al. "Revisiting edge ai: Opportunities and challenges." *IEEE Internet Computing* 28.4 (2024): 49-59.

12. Open Source Software Strategy 2020-2023: Think Open. C(2020) 7149 final

13. AI Continent Action Plan. COM(2025) 165 final

as their proprietary counterparts. These factors highlight the need for careful management and strategic planning to navigate the unique challenges and dynamics of open source growth.

The concept of openness is rather undefined in the context of GenAI. Technically, a GenAI model can be defined as open source when all its components (the source code used to train and run the model, the model weights, the model architecture and information on data usage)[14] are released; training data may also be shared, but this is not always possible due to legal and copyright restrictions. This has led to several discussions on what constitutes an open source model, but also to a general over-claiming of openness (a practice known as open washing), since in many cases the model weights are the only available component. Such models are usually identified as open-weights models.[15] Debates concern, among others, the family of models from the French Mistral AI (https://mistral.ai), those from the Chinese DeepSeek model (https://www.deepseek.com/en) and some US models such as BERT (https://github.com/google-research/bert) and Gemma (https://ai.google.dev/gemma) from Google AI. Some of these, however, also release some minimal training code and related documentation, but no access to the training data is provided. Due to the restrictions introduced through its licence, Meta's LlaMa models (https://www.llama.com) cannot be considered open source. An example of fully open source GenAI models is the Olmo family of models from Allen AI (https://allenai.org/olmo), where the underlying model code, weights, architecture, documentation, and information on data usage are available together with training data. Regarding the latter, in the playground environment at https://playground.allenai.org, users can see and access the exact documents used to generate the model replies. Such a varying degree of openness of GenAI models (in terms of code, weights, documentation, training data, hardware architecture, datasheets, licenses, etc.) can be assessed through the European Open Source AI Index.[16] [17]

In contrast, proprietary GenAI models do not offer the benefits of their open source alternatives, but they may provide a diverse set of competitive advantages through unique features and capabilities. These include dedicated commercial support and maintenance, enhanced performance, protection of intellectual property, user-friendly interfaces and access/interaction tools for non-technical users. Popular examples include all-in-one, general-purpose models like the GPT series from OpenAI (https://openai.com) and Google's Gemini (https://gemini.google.com).

Balancing these considerations involves assessing the trade-offs between innovation, control, and accessibility. The choice of the model (open source vs. proprietary) depends on the specific goals and needs of developers and users. Hybrid approaches are also possible, e.g. where an open source model is finetuned with proprietary algorithms enabling specific applications, and is deployed and scaled on secure private clouds so that sensitive data stay within the organisation's control and are not exposed to third-party vendors.

1.4 Why It Matters for EU Policymakers

KEY MESSAGES

- GenAI holds immense potential as a strategic asset for the European Union, offering opportunities for economic growth, innovation, and societal advancement.
- For EU policymakers, the task at hand is to navigate the complexities of the GenAI landscape, addressing challenges related to regulation, ethics, and skills while capitalising on opportunities for growth and competitiveness.
- The rapid evolution of GenAI poses challenges for policymakers to keep up with the latest developments and ensure that regulations are effective and up-to-date, calling for continuous

14. Recital 104 in the AI Act.
15. https://huggingface.co/blog/2023-in-llms
16. https://osai-index.eu/the-index
17. https://www.nature.com/articles/d41586-024-02012-5

research for policy collaboration to guarantee anticipatory evidence for policy.

GenAI is a technological breakthrough and potentially a strategic asset for the EU. To ensure that the EU harnesses the benefits of this transformative technology while aligning with European values, it needs to focus on a few critical areas, including an understanding of the socio-economic and ethical dimensions of GenAI.

STRATEGIC IMPORTANCE OF GENAI

GenAI represents a pivotal technological advancement with the potential to drive economic growth and innovation across various sectors – while protecting the rights of EU citizens and businesses. For EU policymakers, GenAI is a strategic tool with links to digital sovereignty and competitiveness. By leading in GenAI development and deployment, the EU can assert its position as a global technology leader, influencing the international standards and norms that govern AI technologies, models and systems. This is particularly important in sectors such as healthcare, robotics, and assistive technologies, where GenAI can bring about significant transformations.

The potential of GenAI to transform industries and the public sector, including healthcare, education and public administration, highlights its strategic importance. Policymakers can leverage GenAI to address pressing societal challenges, improve public services, and foster innovation ecosystems that support socio-economic resilience and job creation. Moreover, innovations in semiconductor technologies, such as neuromorphic chips and edge AI chips, can offer a path toward more sustainable AI deployment, which is crucial for the EU to balance climate neutrality targets with maintaining global competitiveness in AI.

POLICY AND REGULATORY LANDSCAPE IN THE EU

The EU has established a comprehensive policy and regulatory landscape for AI, including GenAI, with a focus on promoting innovation, trust, and protection of fundamental rights. The AI Act[18] establishes a risk-based approach, with stricter requirements for high-risk AI systems, including those used in critical infrastructure, healthcare, and law enforcement (see Section 5.1 for specific implications on GenAI).

Additionally, the Commission has established the European AI Office to oversee the implementation of the Regulation, provide guidance, and support the development and adoption of AI in the EU.[19] In the context of the AI Continent Action Plan, the European Commission announced investments of close to EUR 700 million in calls from Horizon Europe and the Digital Europe Programme as part of the GenAI4EU initiative for the development of advanced AI models and solutions in a wide range of sectors, as well as development of data infrastructures and skills.[20] Additionally, the EU is investing in AI Factories, which will enhance collaboration in AI across Europe and drive advances in AI applications. As part of the same initiative, the European Digital Innovation Hubs (see Section 3.2) will also support continuous learning by workers in SMEs, mid-caps, start-ups, and public-sector organisations. The EU is also supporting AI in Science (see Section 6.3 expanding on relevant impacts) through several initiatives, including the European AI Research Council, also known as Resource for AI Science in Europe (RAISE), which will pool resources to push the technological boundaries of AI and facilitate scientific breakthroughs.

18. 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) (Text with EEA relevance)

19. Commission Decision of 24 January 2024 establishing the European Artificial Intelligence Office

20. AI Continent Action Plan. COM(2025) 165 final.

The EU's regulatory framework for AI also includes the General Data Protection Regulation (GDPR), which applies to the processing of personal data in AI systems, and the Digital Services Act (DSA), which regulates online platforms and services, including those that use AI. Implications for GenAI are analysed further in Sections 5.2 and 5.3. The EU has also adopted horizontal data legislation (Section 5.5) to promote a competitive and trustworthy data economy. For GenAI developers, this framework provides mechanisms to access diverse, high-quality datasets, while ensuring compliance with data rights, enabling cross-sector and cross-border data sharing to support AI training and innovation. It comprises the Data Governance Act, designed to enhance trust in voluntary data sharing;[21] the Data Act which clarifies data access rights and facilitates business-to-business data sharing;[22] and the High-Value Dataset (HVD) Implementing Regulation to facilitate the reuse of high value datasets.[23] These horizontal measures are complemented by sector-specific Common European Data Spaces in strategic economic sectors and domains of public interest, and the European Health Data Space Regulation.[24]

CHALLENGES AND OPPORTUNITIES AT THE SCIENCE AND POLICY INTERFACE

As GenAI continues to evolve, EU policymakers face several challenges that must be addressed to maximise its benefits. One of the primary challenges is the need for comprehensive understanding of the techno-socio-economic aspects of GenAI, an anticipatory approach that requires establishing a continuous technology foresight process to scan for signals of change, analyse trends, and make sense of emerging developments and their implications through multi-stakeholder and multi-expertise collaboration. Such an anticipatory element will help inform policy recommendations, guide funding strategies, and foster coordination between European institutions and international partners. This proactive stance ensures that the EU remains adaptive, ready to tackle both the opportunities and risks that GenAI brings.

Policymakers, including both private and public decision-makers, must keep investing in education and training programmes that equip individuals with the skills needed to work alongside and with AI systems and harness their potential. As AI-generated content proliferates, the risk of over-reliance and dependence on such technologies in education, art, and public discourse raises new social and ethical concerns. While AI, and in particular GenAI, can augment human creativity and productivity, enabling faster workflows and alternative educational experiences, it can also dampen critical thinking, nuanced understanding, and the development of human skills while undermining our capacities to act in case such systems are not available. This tension between productivity gains and cognitive erosion demands that policymakers encourage judicious AI adoption, especially in sensitive domains like education and healthcare. AI should be a tool to augment human capabilities, not replace them. ■

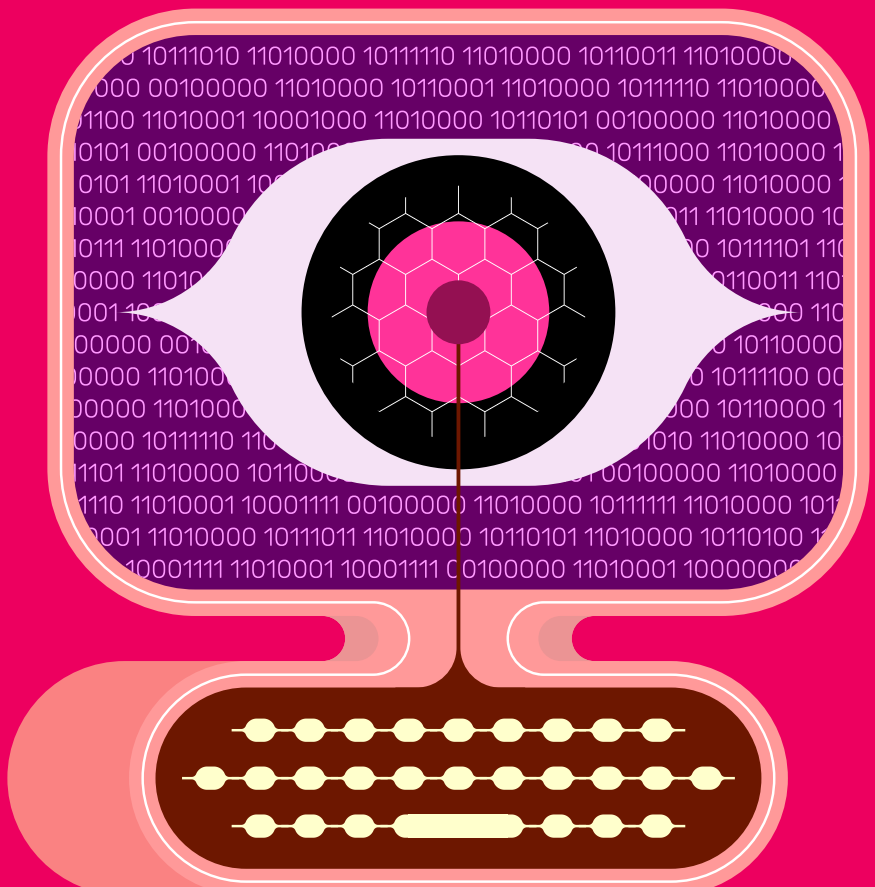
21. Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724.

22. 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.

23. Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use.

24. Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847.

2
TECHNOLOGICAL ASPECTS



This chapter focuses on the technological dimensions of GenAI, beginning with an evaluation of its capabilities and limitations. It highlights cybersecurity challenges that arise from the deployment of GenAI systems. Emerging technological trends are discussed, offering a forward-looking perspective for policymakers and stakeholders. Key issues include the evaluation paradigms needed to ensure the safety and reliability of GenAI applications, particularly in safety-critical domains.

2.1 Generative AI Evaluation

KEY MESSAGES

- The development of standardised evaluation methodologies and benchmarks is essential for understanding the potential risks and limitations of GenAI models and systems.
- Humans play a crucial role in GenAI evaluation, and their involvement is necessary for developing trustworthy benchmarks, improving explainability and predictability, and ensuring the safety of GenAI models and systems.

The rapid development and deployment of GenAI models and systems have raised significant concerns about their potential risks and limitations. Despite the efforts of the AI evaluation community, the capabilities and safety of GenAI in real-world scenarios are not yet fully understood, and risks have not been adequately identified and mitigated.[25] This situation is unacceptable in safety-critical domains, such as aviation, motor vehicles, and pharmaceuticals, where rigorous evaluation and testing are essential prerequisites for market release and general adoption.

25. L. Weidinger, et al. "Toward an Evaluation Science for Generative AI Systems", arXiv:2503.05336, 2025.

CURRENT EVALUATION PARADIGMS AND METHODOLOGIES

→ Among the different AI evaluation paradigms and methodologies,[26] **benchmarking** is currently the most widely used and adopted. Benchmarking enables standardised comparisons, reduces ambiguity in evaluation results, increases transparency, and facilitates performance tracking over time. However, benchmarking has several limitations, including problems related to data collection, annotation, and documentation, concerns regarding construct validity, sociocultural and sociotechnical gaps, limited diversity and scope, and risks associated with competitive and commercial influences. Additionally, benchmarking is vulnerable to issues such as rigging and gaming, questionable community vetting, saturation, complexity, and unknown unknowns.[27] The situation demands new ways of signalling which benchmarks to trust, especially if these benchmarks are to play a significant role in regulatory application contexts.

→ Another important evaluation methodology is **adversarial testing** through red teaming, where humans or automated agents interactively attempt to manipulate the GenAI model or system, eliciting undesirable responses. Red teaming can be used to assess capabilities, but its primary application is in evaluating potential harms.[28] Human evaluations, such as the "human uplift" study, involve assessing humans under two settings: one with access to traditional tools and another with

26. J. Burden, M. Tešić, L. Pacchiardi, J. Hernández-Orallo, "Paradigms of AI Evaluation: Mapping Goals, Methodologies and Culture", arXiv:2502.15620, 2025.

27. M. Eriksson, E. Purificato, A. Noroozian, J. Vinagre, G. Chaslot, E. Gomez, D. Fernandez-Llorca, "Can We Trust AI Benchmarks? An Interdisciplinary Review of Current Issues in AI Evaluation", arXiv:2502.06559, 2025.

28. Kurakin, Alexey & Goodfellow, Ian & Bengio, Samy. (2016). Adversarial Machine Learning at Scale. 10.48550/arXiv.1611.01236

additional access to GenAI models. The goal of human uplift studies is to determine whether GenAI models significantly enhance humans' capabilities to produce potential harms.[29] Although red teaming or human uplift studies offer a systematic way to identifying flaws in GenAI models and systems, a major limitation is that the results obtained can never be taken as absolute guarantees of safety[30] (absence of evidence is not evidence of absence). All methodologies involving human evaluations, including benchmarks, are significantly more costly to implement and less scalable. However, the role of humans in GenAI evaluation is crucial, including the need for human-level performance to improve the explainability and predictability of benchmarks.

→ **Trustworthy benchmarks** are expected to provide **human norms**, with capability profiles and difficulty levels to enhance their explanatory and predictability power.[31] However, it remains unclear what the most appropriate conditions are for generating human performance levels to compare with. For a particular task, and depending on the difficulty levels, should the target performance be that of the average human individual, the average human expert, the performance of any random human, or the collective ability of humanity as a whole? Should this be assessed individually or in groups, with or without the aid of supporting tools? How much time should be allocated for a specific task? These are just some of questions that must be addressed to create solid human-norm baselines. However, the human reference may become insufficient when the capabilities of GenAI substantially surpass those of humans.

→ The current **emergence of GenAI agents**, which enable autonomous systems to plan, reason, use tools, and maintain memory while interacting with dynamic environments, requires new benchmarks and evaluation methodologies. Multiple agent evaluation methods have been proposed, including final response, stepwise, trajectory-based, A/B comparisons, or gym-like approaches. Still, some important challenges remain, particularly regarding cost-efficiency, fine-grained evaluation, and the limited focus on safety.[32]

THE NEED FOR A SCIENCE OF EVALUATION

In recent years, there have been several calls for a "science of evals"[33] a new "model metrology" discipline, or an "evaluation science for GenAI". These initiatives advocate for the need for standardised evaluations of capabilities and safety for GenAI models and systems, as well as the creation of a specialised community that embodies the efforts and collaboration of multiple stakeholders, including academia, industry, users, and policymakers.

FUTURE PERSPECTIVES

In summary, we can identify the following trends that will constitute part of the necessary ongoing and future works in the field of GenAI evaluation:

→ New methodologies for benchmarking GenAI evaluation benchmarks, including transparency and clear assessment of what they are really measuring, will flourish in the coming years.

→ The focus on safety evaluation of agentic GenAI systems will be strengthened.

29. METR, "Evaluating AI Models for Critical Harms", URL: https://metr.org/evaluating-ai-models-for-critical-harms.pdf, 2024

30. John Burden, "Evaluating AI Evaluation: Perils and Prospects", arXiv:2407.09221, 2024.

31. Lexin Zhou, et al. "General Scales Unlock AI Evaluation with Explanatory and Predictive Power", arXiv:2503.06378, 2025

32. A. Yehudai, L. Eden, A. Li, G. Uziel, Y. Zhao, R. Bar-Haim, A. Cohan, M. Shmueli-Scheuer, "Survey on Evaluation of LLM-based Agents", arXiv:2503.16416, 2025.

33. Apollo Research, "We need a Science of Evals", URL: https://www.apolloresearch.ai/blog/we-need-a-science-of-evals, 2024.

→ Humans will continue to play a crucial role in GenAI evaluation, including red teaming, structured approaches, human-centred studies, A/B testing, and developing human-level performance references.

→ "Super-human evaluation", the need to evaluate capabilities that exceed those of humans and safety concerns that are unknowable or imperceptible to humans will become increasingly important.

→ We anticipate more calls and collaborative initiatives on evaluation science for GenAI as a collaborative effort involving multiple stakeholders and multiple disciplines.

2.2 Cybersecurity Challenges of Generative AI

KEY MESSAGES

- The cybersecurity and safety of GenAI systems are critical concerns that require a multifaceted approach, combining traditional software risk management with AI-specific strategies, and considering the full attack surface and AI artefacts, including data and models.
- GenAI systems are vulnerable to traditional cybersecurity threats and AI-specific vulnerabilities, such as data and model poisoning, adversarial attacks, and the misuse of generated content for malicious purposes. The significance of these challenges is increased by dependencies on third-party data and models.

In the rapidly evolving landscape of AI, GenAI has emerged as a transformative tool. However, the increased integration of GenAI components in software systems has introduced new risks and unique challenges that need to be properly addressed. GenAI systems are susceptible to the same cybersecurity risks associated with traditional digital systems operating in similar environments, as well as AI-specific vulnerabilities introduced by their GenAI components and assets, such as data and model poisoning, adversarial attacks, and the misuse of generated content for malicious purposes.

When discussing AI cybersecurity and safety, it is essential to acknowledge two key aspects. First, traditional cybersecurity practices and procedures, which are well-established and effective for securing conventional software systems, are limited in their capacity to address the broader range of vulnerabilities affecting GenAI systems. Second, AI cybersecurity encompasses the security and safety of AI systems, not those of single AI components, such as AI models. AI systems can be built on top of multiple AI components and tools, and their security encompasses their related assets, such as their training data.

SUPPLY CHAIN ATTACKS

AI systems inherit many vulnerabilities from traditional software supply chains, such as reliance on third-party dependencies, but they also introduce unique challenges due to their specific dependencies on data and third-party AI models. These vulnerabilities can compromise the integrity of training data, models, and deployment platforms, leading to biased outputs or security breaches. Addressing these risks requires a multifaceted approach that combines traditional software risk management with AI-specific strategies, such as using provenance information to track AI components and considering the full attack surface and AI artefacts, including data and models.[34]

34. Apruzzese et al. "Real Attackers Don't Compute Gradients": Bridging the Gap Between Adversarial ML Research and Practice". In: 2023 IEEE Conference on Secure and Trustworthy Machine Learning, SaTML 2023, Raleigh, NC, USA, February 8-10, 2023. IEEE, 2023, pp. 339–364. DOI: 10.1109/SATML54575.2023.00031. URL: https://doi.org/10.1109/SaTML54575.2023.00031.

The rise of open-weight models and new fine-tuning methods like LoRA[35] and other PEFT[36] methods, along with the emergence of on-device models, further complicate the AI supply chain, necessitating specialised security measures to mitigate potential poisoning attacks.[37]

→ **Data Poisoning:** the successful training of GenAI foundation models largely relies on the scale and diversity of the training data. The primary source of the massive datasets employed to pre-train GenAI models is the internet,[38] whose content is often unverified and sometimes potentially harmful. The lack of fine-grained control over the training data makes these datasets a large potential attack surface, which attackers may exploit by inserting adversarial samples. Attackers can leverage this opportunity to introduce vulnerabilities, backdoors, or biases, which may compromise the model's performance, thus degrading its capabilities, leading to harmful outputs such as spreading misinformation or introducing security risks by suggesting insecure code.[39] [40] [41]

→ **Model Poisoning:** GenAI systems are vulnerable to various model poisoning attacks, especially when developers rely on open-weight models. Models distributed through open-source platforms can carry hidden threats,[42] such as malware embedded in the models' source code and backdoors.[43] These threats can stay inactive and only trigger under specific conditions, such as when the model is loaded or for inputs containing specific words or phrases, making them challenging to detect and allowing the model to become a sleeper agent.[44] Recently, researchers have demonstrated that attackers can insert backdoors into pre-trained models, which may persist even after fine-tuning[45] or additional safety training,[46] raising significant concerns.

DIRECT PROMPT INJECTION

Direct Prompt Injection occurs when a user prompt alters the behaviour or output of a generative model in unintended ways, potentially causing it to violate guidelines, generate harmful content, enable unauthorised access, or influence critical decisions. When performing Direct Prompt Injection, a malicious user may pursue a variety of goals, such as enabling misuse, invading privacy, or violating integrity.[47] Direct Prompt Injection attacks can be roughly categorised as

35. E. J. Hu et al. "LoRA: Low-Rank Adaptation of Large Language Models". In: The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022. OpenReview.net, 2022. URL: https://openreview.net/forum?id=nZeVKeeFYf9.

36. Z. Han et al. Parameter-Efficient Fine-Tuning for Large Models: A Comprehensive Survey. 2024. arXiv: 2403. 14608 [cs.LG]. URL: https://arxiv.org/abs/2403.14608.

37. OWASP. OWASP Top 10 for LLM Applications 2025. Accessed: 2025-03-31. 2025. URL: https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/.

38. A. Radford et al. Language Models are Unsupervised Multitask Learners. 2018. URL: https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf.

39. C. W. Barrett et al. "Identifying and Mitigating the Security Risks of Generative AI". In: Found. Trends Priv. Secur. 6.1 (2023), pp. 1–52. DOI: 10.1561/3300000041. URL: https://doi.org/10.1561/3300000041.

40. E. Hubinger et al. Sleeper Agents: Training Deceptive LLMs that Persist Through Safety Training. 2024. arXiv: 2401.05566 [cs.CR]. URL: https://arxiv.org/abs/2401.05566.

41. K. Kurita, P. Michel, and G. Neubig. Weight Poisoning Attacks on Pre-trained Models. 2020. arXiv: 2004.06660 [cs.LG]. URL: https://arxiv.org/abs/2004.06660.

42. Mithril Security. MS Windows NT Kernel Description. Accessed: 2025-03-31. 2023. URL: https://blog.mithrilsecurity.io/poisongpt-how-we-hid-a-lobotomized-llm-on-hugging-face-to-spread-fake-news/.

43. K. Kurita, P. Michel, and G. Neubig. Weight Poisoning Attacks on Pre-trained Models. 2020. arXiv: 2004.06660 [cs.LG]. URL: https://arxiv.org/abs/2004.06660.

44. E. Hubinger et al. Sleeper Agents: Training Deceptive LLMs that Persist Through Safety Training. 2024. arXiv: 2401.05566 [cs.CR]. URL: https://arxiv.org/abs/2401.05566.

45. K. Kurita, P. Michel, and G. Neubig. Weight Poisoning Attacks on Pre-trained Models. 2020. arXiv: 2004.06660 [cs.LG]. URL: https://arxiv.org/abs/2004.06660.

46. E. Hubinger et al. Sleeper Agents: Training Deceptive LLMs that Persist Through Safety Training. 2024. arXiv: 2401.05566 [cs.CR]. URL: https://arxiv.org/abs/2401.05566.

47. A. Vassilev et al. "Adversarial machine learning: A taxonomy and terminology of attacks and mitigations". In: National Institute of Standards and Technology (2025). DOI: 10.6028/NIST.AI.100-2e2025. URL: https://doi.org/10.6028/NIST.AI.100-2e2025.

optimisation-based attacks, manual methods, and model-assisted attacks:

→ **Optimisation-based** attacks systematically refine adversarial prompts through algorithmic optimisation methods, aiming to maximise the probability of generating malicious or harmful responses. This can be achieved, for example, by optimising adversarial suffixes that allows the evasion of the safety alignment of GenAI models.[48] [49] [50] These suffixes can often be transferred between models, making open-weight models, which grant white-box access to malicious users, viable attack vectors for transferability attacks against closed systems that only offer API access.[51]

→ **Manual methods** seek to trigger two primary failure modes of GenAI models: competing objectives and mismatched generalisation.[52] Competing objectives arise when a model's capabilities and safety goals conflict, such as leveraging a model's willingness to follow user-provided instructions.[53] [54] [55] [56] [57] A prominent example is using role-playing strategies to push the model into a state of conflict with its original intent, thus compromising its safety protocols.[58] [59] [60] [61]

→ **Model-assisted** attacks employ auxiliary language models to generate and refine jailbreak prompts autonomously.[62]

48. A. Zou et al. Universal and Transferable Adversarial Attacks on Aligned Language Models. 2023. arXiv: 2307.15043 [cs.CL]. URL: https://arxiv.org/abs/2307.15043.

49. M. Andriushchenko, F. Croce, and N. Flammarion. Jailbreaking Leading Safety-Aligned LLMs with Simple Adaptive Attacks. 2024. arXiv: 2404.02151 [cs.CR]. URL: https://arxiv.org/abs/2404.02151.

50. Z. Liao and H. Sun. AmpleGCG: Learning a Universal and Transferable Generative Model of Adversarial Suffixes for Jailbreaking Both Open and Closed LLMs. 2024. arXiv: 2404.07921 [cs.CL]. URL: https: //arxiv.org/abs/2404.07921.

51. A. Zou et al. Universal and Transferable Adversarial Attacks on Aligned Language Models. 2023. arXiv: 2307.15043 [cs.CL]. URL: https://arxiv.org/abs/2307.15043.

52. A. Wei, N. Haghtalab, and J. Steinhardt. "Jailbroken: How Does LLM Safety Training Fail?" In: Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023. Ed. by A. Oh et al. 2023. URL: http://papers.nips.cc/paper%5C_files/paper/ 2023/hash/fd6613131889a4b656206c50a8bd7790-Abstract-Conference.html.

53. X. Shen et al. ""Do Anything Now": Characterizing and Evaluating In-The-Wild Jailbreak Prompts on Large Language Models". In: Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security, CCS 2024, Salt Lake City, UT, USA, October 14-18, 2024. Ed. by B. Luo et al. ACM, 2024, pp. 1671–1685. DOI: 10.1145/3658644.3670388. URL: https://doi.org/10.1145/3658644.3670388.

54. X. Liu et al. "AutoDAN: Generating Stealthy Jailbreak Prompts on Aligned Large Language Models". In: The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024. OpenReview.net, 2024. URL: https://openreview.net/forum?id=7Jwpw4qKkb.

55. X. Liu et al. "AutoDAN-Turbo: A Lifelong Agent for Strategy Self-Exploration to Jailbreak LLMs". In: CoRR abs/2410.05295 (2024). DOI: 10.48550/ARXIV.2410.05295. arXiv: 2410.05295. URL: https://doi.org/10. 48550/arXiv.2410.05295.

56. X. Li et al. DeepInception: Hypnotize Large Language Model to Be Jailbreaker. 2024. arXiv: 2311.03191 [cs.LG]. URL: https://arxiv.org/abs/2311.03191.

57. N. Xu et al. "Cognitive Overload: Jailbreaking Large Language Models with Overloaded Logical Thinking". In: Findings of the Association for Computational Linguistics: NAACL 2024, Mexico City, Mexico, June 16-21, 2024. Ed. by K. Duh, H. Gómez-Adorno, and S. Bethard. Association for Computational Linguistics, 2024, pp. 3526–3548. DOI: 10.18653/V1/2024.FINDINGS-NAACL.224. URL: https://doi.org/10.18653/v1/ 2024.findings-naacl.224.

58. H. Lv et al. CodeChameleon: Personalized Encryption Framework for Jailbreaking Large Language Models. 2024. arXiv: 2402.16717 [cs.CL]. URL: https://arxiv.org/abs/2402.16717.

59. Á. Huertas-García et al. "Camouflage is all you need: Evaluating and Enhancing Language Model Robustness Against Camouflage Adversarial Attacks". In: CoRR abs/2402.09874 (2024). DOI: 10.48550/ARXIV.2402.09874. arXiv: 2402.09874. URL: https://doi.org/10.48550/arXiv.2402.09874.

60. Yuan et al. "GPT-4 Is Too Smart To Be Safe: Stealthy Chat with LLMs via Cipher". In: The Twelfth Interna- tional Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024. OpenReview.net, 2024. URL: https://openreview.net/forum?id=MbfAK4s61A.

61. Y. Deng et al. "Multilingual Jailbreak Challenges in Large Language Models". In: The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024. OpenReview.net, 2024. URL: https://openreview.net/forum?id=vESNKdEMGp.

62. P. Chao et al. "Jailbreaking Black Box Large Language Models in Twenty Queries". In: CoRR abs/2310.08419 (2023). DOI: 10.48550/ARXIV.2310.08419. arXiv: 2310.08419. URL: https://doi.org/10.48550/arXiv. 2310.08419.

[63] [64] For example, an attacker model, a target model, and a judge model may be employed to train a generative model (the attacker) to generate jailbreaks for another generative model (the target) relying on a reward function derived from the judge model evaluating whether the target model's output is harmful.

INFORMATION EXTRACTION

During their life cycle, GenAI models are exposed to a wide range of information that may be of interest to attackers. For example, their training data may contain personally identifying information that has not been properly anonymised, or sensitive information may become part of their input when a retrieval augmented generation pipeline is employed by the system. Additionally, assets of the system itself, such as the model weights or architecture and the system prompt, may be valuable targets.

→ **Data Leakage and Membership Inference:** data leakage occurs when sensitive or confidential information is unintentionally exposed to unauthorised parties. In the context of GenAI, this information encompasses various types of data, such as confidential training data, personal identifiable information, and copyrighted material. Leaking sensitive data can lead to legal actions and fines against GenAI system providers, as well as harming their reputation and resulting in a loss of competitive advantage. To cause data leakage, attackers usually rely on membership inference attacks.Those attacks try to uncover whether an input sample was part of the training data of a GenAI model,[65] potentially exposing confidential or sensitive data memorised by the model during training, such as copyrighted material[66] or credit card numbers.[67]

→ **Model Inversion** occurs when attackers attempt to reconstruct training data or infer sensitive information from the model's outputs.[68] [69] [70] In this type of attack, the attacker typically has access to the model.

63. A. Mehrotra et al. "Tree of Attacks: Jailbreaking Black-Box LLMs Automatically". In: Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024. Ed. by A. Globersons et al. 2024. URL: http://papers.nips.cc/paper%5C_files/paper/2024/hash/70702e8cbb4890b4a467b984ae59828a-Abstract-Conference.html.

64. E. Perez et al. "Red Teaming Language Models with Language Models". In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Ed. by Y. Goldberg, Z. Kozareva, and Y. Zhang. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, Dec. 2022, pp. 3419–3448. DOI: 10.18653/v1/2022.emnlp-main.225. URL: https://aclanthology.org/2022.emnlp-main.225/.

65. R. Shokri et al. "Membership Inference Attacks Against Machine Learning Models". In: 2017 IEEE Symposium on Security and Privacy, SP 2017, San Jose, CA, USA, May 22-26, 2017. IEEE Computer Society, 2017, pp. 3–18. DOI: 10.1109/SP.2017.41. URL: https://doi.org/10.1109/SP.2017.41.

66. E. Su et al. Extracting Memorized Training Data via Decomposition. 2024. arXiv: 2409.12367 [cs.LG]. URL:https://arxiv.org/abs/2409.12367.

67. N. Carlini et al. "The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks". In: 28th USENIX Security Symposium, USENIX Security 2019, Santa Clara, CA, USA, August 14-16, 2019. Ed. by N. Heninger and P. Traynor. USENIX Association, 2019, pp. 267–284. URL: https://www.usenix. org/conference/usenixsecurity19/presentation/carlini.

68. M. Fredrikson et al. "Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing". In: Proceedings of the 23rd USENIX Security Symposium, San Diego, CA, USA, August 20-22, 2014. Ed. by K. Fu and J. Jung. USENIX Association, 2014, pp. 17–32. URL: https://www.usenix.org/conference/ usenixsecurity14/technical-sessions/presentation/fredrikson%5C_matthew.

69. M. Fredrikson, S. Jha, and T. Ristenpart. "Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures". In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, October 12-16, 2015. Ed. by I. Ray, N. Li, and C. Kruegel. ACM, 2015, pp. 1322–1333. DOI: 10.1145/2810103.2813677. URL: https://doi.org/10.1145/2810103.2813677.

70. Y. Zhang et al. "The Secret Revealer: Generative Model-Inversion Attacks Against Deep Neural Networks". In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020. Computer Vision Foundation / IEEE, 2020, pp. 250–258. DOI: 10.1109/CVPR42600.2020.00033. URL: https://openaccess.thecvf.com/content%5C_CVPR%5C_2020/html/Zhang%5C_The%5C_Secret%5C_Revealer%5C_Generative%5C_Model-Inversion%5C_Attacks%5C_Against%5C_Deep%5C_Neural%5C_Networks%5C_CVPR%5C_2020%5C_paper.html.

By analysing the model's outputs, the attacker tries to reverse-engineer the model to extract information about the original training data. If the model was trained on sensitive data, such as medical records or personal information, model inversion could lead to privacy breaches.

→ **Model Extraction** occurs when attackers attempt to extract the parameters from a remote model, so that they can have their own copy.[71] Additionally, model extraction is related to training data extraction. While the two attacks share some similarities, they have different goals: model extraction aims to steal the parameters of the remote model, whereas training data extraction seeks to extract the training data that were used to generate those parameters. Researchers have recently shown that model-stealing attacks can extract precise, nontrivial information from black-box AI models in production systems.[72]

INDIRECT PROMPT INJECTION

Indirect Prompt Injection occurs when a GenAI model accepts input from external sources, such as websites or files, which can alter its behaviour or output in unintended ways. This type of attack is carried out by a malicious third party, without direct interaction with the underlying model, and can affect the system operations. The primary user of the model often suffers the consequences of an indirect prompt injection attack, which can compromise the integrity, availability, or privacy of the GenAI system. Indirect prompt injection can cause several issues,[73] including:

→ Disrupting availability by making the model perform time-consuming operations, instructing it not to use certain APIs or tools, or corrupting its output.

→ Compromising integrity by instructing the model to respond with attacker-specified information, such as spreading misleading information, recommending fraudulent products or services, suppressing or hiding certain information, or redirecting users to malicious websites or content.

→ Compromising privacy: by causing the leakage of sensitive information, for example, by persuading primary users to provide information that is then leaked to the attacker.

2.3 Emerging Technological Trends: a Future-Looking Perspective for Policy-makers

KEY MESSAGES

- New trends in technology developments include Agentic AI, which involves autonomous systems making independent decisions and learning from outcomes; Multi-modal AI, which integrates diverse data formats, enhancing versatility but posing bias challenges; Advanced AI Reasoning, which enhances decision-making by analysing complex information and drawing logical conclusions; and Explainability in AI, which highlights the increasing need for AI systems to provide understandable justifications.
- Despite being technologically revolutionary, these developments

71. F. Tramèr et al. "Stealing Machine Learning Models via Prediction APIs". In: 25th USENIX Security Symposium, USENIX Security 16, Austin, TX, USA, August 10-12, 2016. Ed. by T. Holz and S. Savage. USENIX Associ- ation, 2016, pp. 601–618. URL: https://www.usenix.org/conference/usenixsecurity16/technical- sessions/presentation/tramer.

72. N. Carlini et al. "Stealing part of a production language model". In: Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024. OpenReview.net, 2024. URL: https://openreview.net/forum?id=VE3yWXt3KB.

73. S. Abdelnabi et al. "Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection". In: Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security, AISec 2023, Copenhagen, Denmark, 30 November 2023. Ed. by M. Pintor, X. Chen, and F. Tramèr. ACM, 2023, pp. 79–90. DOI: 10.1145/3605764.3623985. URL: https://doi.org/10.1145/3605764.3623985.

call for policymakers to reflect on the possible need to review policy initiatives when it comes to copyright, and to keep implementing ethical oversight while prioritising AI literacy, as well as enforcing standards for transparency and explicability in AI systems and considering the sustainable use of resources.

In the beginning, LLMs were able to generate fluent and contextually appropriate language, but without engaging in genuine comprehension or logical deliberation; nicknamed "stochastic parrots",[74] they repeated the most probable answer without understanding. The core technology of GenAI has been unchanged: it uses the transformer architecture.[75] However, the pace of GenAI development has been unprecedented and is expected to continue in the short term. The developments have been dictated by scaling laws: increasing model size, dataset volume, and computational power lead to significant performance improvements.[76]

However, recent advancements in architectures and training paradigms have paved the way for the emergence of models exhibiting four key functionalities: agentic AI, multi-modal systems, reasoning, and explicability. These advances enable autonomous action, integrate diverse data, enhance decision-making, and ensure transparency—showing significant potential to reshape GenAI use and its impacts and raising critical policy questions around accountability, fairness, and governance in its deployment.

74. Bender, Emily M., et al. "On the dangers of stochastic parrots." *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Mar. 2021, pp. 610–623, https://doi.org/10.1145/3442188.3445922.

75. Vaswani, Ashish, et al. "Attention Is All You Need." *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5998–6008. https://doi.org/10.5555/3295222.3295349.

76. Kaplan, Jared, et al. "Scaling Laws for Neural Language Models." arXiv, 2020, arXiv:2001.08361. https://arxiv.org/abs/2001.08361.

AGENTIC AI

Agentic AI refers to systems that do more than a one-off response to prompts - they make autonomous decisions, initiate actions in pursuit of goals, and learn from outcomes. Unlike conventional AI, which relies on external prompts, agentic AI displays a form of computational agency - exhibiting traits of intentionality and initiative. These AI agents are becoming increasingly autonomous, capable of navigating digital environments, managing complex tasks, and even correcting their own behaviour through feedback mechanisms. For example, Agent-R introduces a self-correcting framework;[77] the agent and model spiking neural network learns while "dreaming" (living new experiences in a model-based simulated environment);[78] and Meta's collaborative Reasoner (Coral) trains and evaluates AI agents on collaborative reasoning,[79] allowing LLMs to revise their outputs and learn from feedback loops, improving performance over time. This pushes the boundaries of AI from passive tools toward semi-autonomous collaborators.

Such developments carry significant implications for the future of work and knowledge production. In scientific research, for instance, AI co-scientists like those developed by Google DeepMind are autonomously generating hypotheses and designing experiments.[80] This shift challenges traditional notions of expertise, authorship,

77. Yuan, Siyu, et al. "Agent-R: Training Language Model Agents to Reflect via Iterative Self-Training." arXiv, 24 Mar. 2025, arxiv.org/abs/2501.11425

78. Capone, Cristiano, and Pier Stanislao Paolucci. "Towards Biologically Plausible Model-Based Reinforcement Learning in Recurrent Spiking Networks by Dreaming New Experiences." *Nature News*, Nature Publishing Group, 25 June 2024, www.nature.com/articles/s41598-024-65631-y.

79. "Collaborative Reasoner: Self-Improving Social Agents with Synthetic Conversations." *Collaborative Reasoner: Self-Improving Social Agents with Synthetic Conversations | Research - AI at Meta*, ai.meta.com/research/publications/collaborative-reasoner-self-improving-social-agents-with-synthetic-conversations/. Accessed 7 May 2025.

80. Gottweis, Juraj, and Vivek Natarajan. "Accelerating Scientific Breakthroughs with an AI Co-Scientist." Google Research Blog, 19 Feb. 2025, research.google/blog/accelerating-scientific-breakthroughs-with-an-ai-co-scientist/

and accountability in research, as the line blurs between human and machine-driven discovery. On the business side, Microsoft created Agent Store,[81] which enables the use of agents in everyday enterprise settings, even as personal assistants. These digital co-workers are now capable of executing complex workflows independently adapted to needs.

Beyond specialised fields, agentic AI is also emerging in everyday internet use, like OpenAI's Operator AI agent[82] or Google's Mariner.[83] BrowseComp[84] introduces benchmarks for AI agents that can autonomously navigate the web, a foundational step toward machine-first digital ecosystems. These systems raise new issues about data privacy, cybersecurity, and governance, especially as AI begins to make decisions in online environments without direct human intervention. As this trend grows, standard-setting and regulatory clarity will be essential to manage risks while enabling innovation.

The rise of agentic AI opens new territories in Human-AI collaboration, personalisation of models and adaptive learning by these models. It also raises fundamental policy questions. If an AI system can act independently, who is responsible when something goes wrong? What does it mean to be an "employee" in a world where tasks are carried out by autonomous systems? And how do we define authorship or ownership when machines co-create or even originate content? How do we interact with AI systems that become more autonomous? The EU should address these challenges by assessing and potentially by rethinking liability frameworks, employment law, and intellectual property rights to reflect the changing nature of agency in the digital age.

81. "Boosting HR and IT Services at Microsoft with Our New Employee Self-Service Agent in Microsoft 365 Copilot." *Microsoft Inside Track, Microsoft*, 6 May 2025.

82. https://openai.com/index/introducing-operator/

83. https://deepmind.google/technologies/project-mariner/

84. Wei, Jason, et al. "BrowseComp: A Simple Yet Challenging Benchmark for Browsing Agents." arXiv, 16 Apr. 2025, arxiv.org/abs/2504.12516

MULTI-MODAL GENAI

Multi-modal GenAI represents a significant step forward in the evolution of GenAI. By integrating multiple data formats - text, images, audio, data (including genetic or clinical data) and even tactile or olfactory (e-nose) input[85]- these systems offer richer, more versatile applications. The GPT models have increasing multimodal capacities, starting from GPT4o ("o" for "omni"). GPT-5, expected in 2025,[86] significantly increases multi-modal reasoning; Meta AI's Multimodal Iterative LLM Solver (MILS) introduces a multimodal framework without training.[87] Yet this advancement also amplifies existing risks. The scale and diversity of training data increase potential for bias, misinformation, and high energy consumption, complicating regulatory oversight and societal trust. Under-represented languages, cultures and dialects may receive subpar services. Ensuring equitable access without reinforcing digital divides remains a central policy concern.

Aya Vision[88] introduces multilingual and multimodal capabilities designed to improve global accessibility, especially for low resourced languages. By seamlessly combining language processing with visual and auditory inputs, Aya Vision supports inclusive AI experiences across diverse user groups, enhancing access to education, healthcare, and communication services.

The democratisation of (artistic) expression comes with additional policy questions. There is

85. Lim, Hyeongtae, et al. "Intelligent olfactory system utilizing in situ ceria nanoparticle-integrated laser-induced graphene." *ACS Nano*, 21 Apr. 2025, https://doi.org/10.1021/acsnano.5c03601.

86. Edwards, Benj. "Sam Altman Lays Out Roadmap for OpenAI's Long-Awaited GPT-5 Model." Ars Technica, 12 Feb. 2025, https://arstechnica.com/ai/2025/02/sam-altman-lays-out-roadmap-for-openais-long-awaited-gpt-5-model/

87. Girdhar, Rohit. "LLMs can see and hear without any training." arXiv, 30 Jan. 2025, arxiv.org/abs/2501.18096v1

88. Dash, Saurabh, et al. "A Deepdive into Aya Vision: Advancing the Frontier of Multilingual Multimodality." Hugging Face Blog, 4 Mar. 2025, https://huggingface.co/blog/aya-vision.

a risk that these systems, by constantly recycling existing styles and materials, may reduce novelty in creative work (see Section 3.2). This "self-bias" effect could lead to a saturation of mediocre and derivative content, dampening genuine artistic innovation. Policymakers must grapple with how to balance access to creative tools with incentives for originality. The ease of generating high-quality visual or audio content, often based on existing works, poses serious challenges for copyright enforcement, especially when systems are trained on large datasets without explicit permission (see Section 5.4). The EU must address the ethical, regulatory, and technical complexities associated with these systems, from safeguarding data privacy to ensuring fairness and inclusivity. Improving copyright and data governance frameworks will be essential, as will promoting quality assurance, ethical uses across languages, and modalities to foster trust in these powerful tools. Additionally, fostering support for diverse, original content creation could counteract the risk of homogenisation and bias in AI-generated media.

REASONING

Reasoning in GenAI has evolved significantly, extending from enhanced decision-making capabilities to simulating human cognition. Building on innovative architectures, training paradigms, knowledge representations, and auxiliary systems (calculators, specialised APIs) designed to emulate human-level deliberation and structured problem-solving. AI is moving beyond narrow, task-specific systems toward scalable frameworks that integrate reasoning and adaptability across sectors such as manufacturing, logistics, education, and public services. This mainstreaming of AI suggests a trajectory toward generalised intelligence, but also raises concerns around infrastructure, interoperability, and governance.[89]

Emerging developments in brain-inspired cognition seek to simulate human reasoning, memory retention, and adaptive learning in AI systems, enhancing their natural interaction capabilities.[90] The deliberation process slows down the response but improves its quality. Major LLMs have introduced a pro search/research option for a more thoughtful, step-by-step reasoning. However, this human-level emulation raises ethical and feasibility concerns, especially around the increasingly blurred lines between human and machine cognition, and increased energy consumption. These advancements make it critical for policymakers to address the ethical boundaries of AI development.

The introduction of Large Concept Models (LCM)[91] [92] further expands reasoning by integrating vast conceptual knowledge into language models; "AI Stem Cell" technology uses a rule-based scientific framework for decisions.[93] This enhances decision-making capacities in complex contexts, such as scientific research, legal analysis, and policy development. However, the depth of conceptual integration raises concerns about model interpretability, computational costs, and data privacy. As these systems grow in scale, ensuring transparency and mitigating associated risks will be crucial for responsible AI deployment.

Finally, as AI takes on more reasoning tasks, there is a risk of human technical skills erosion;

89. Office of the Director of National Intelligence. "Technology." Global Trends 2040: A More Contested World, Mar. 2021, https://www.dni.gov/index.php/gt2040-home/gt2040-structural-forces/technology

90. Peel, Michael. "Microsoft Teams up with AI Start-up to Simulate Brain Reasoning." @FinancialTimes, Financial Times, 18 Mar. 2025, www.ft.com/content/37e44758-04a6-450b-abe3-f51f1d7d972a

91. Team, Lcm, et al. *Large Concept Models: Language Modeling in a Sentence Representation Space*. 2024, arxiv.org/pdf/2412.08821.

92. Chih-Hsuan. "SONAR: Sentence-Level Multimodal and Language-Agnostic Representations." Medium, 27 Dec. 2024, https://medium.com/@chs.li.work/sonar-sentence-level-multimodal-and-language-agnostic-representations-73a81d3f5913. Accessed 8 May 2025

93. Partsol. "World's First AI Stem Cell-Engineered Cognitive AI Platform Launched by Irish-Based Partsol." *PR Newswire: Press Release Distribution, Targeting, Monitoring and Marketing*, Cision PR Newswire, 31 Mar. 2025, www.prnewswire.com/news-releases/worlds-first-ai-stem-cell-engineered-cognitive-ai-platform-launched-by-irish-based-partsol-302415793.html.

gaps in AI literacy within education systems exacerbate this issue.[94] [95] The EU must prioritise skills development and AI literacy to ensure human expertise remains relevant. Towards this end, the European Commission in collaboration with the OECD has recently published the first draft of an AI Literacy Framework for primary and secondary schools (see Section 6.2).[96] Balancing the benefits of reasoning-capable AI with the need for explainability, ethical oversight, and human skill retention is vital for sustainable integration.

EXPLICABILITY

Explicability, or the capacity for AI systems to provide understandable justifications for their decisions, is becoming essential as GenAI models take on more complex roles across sensitive sectors. Explainable AI (XAI)[97] [98] enhances trust by ensuring that AI systems can explain their outputs in ways that are interpretable and understandable for humans. In sectors like security, healthcare, finance, and manufacturing, understanding how AI systems arrive at conclusions is vital for user confidence, regulatory compliance, and effective human-AI collaboration. XAI distinguishes itself from traditional black-box models by incorporating post hoc explanation techniques such as LIME and SHAP, saliency maps and attention visualisation, and introducing rules-based decisions alongside human-centred approaches from philosophy and cognitive science, to provide interpretable insights.

Trustworthy AI[99] [100] [101] further expands this concept by focusing on making AI systems reliable, besides transparent and explainable, using methods like neurosymbolic computing to combine machine learning with logical reasoning. It also introduces tools for risk assessment, such as Key AI Risk Indicators (KAIRI),[102] helping manage the potential downsides of AI systems. This approach strengthens fairness and accountability in areas like security, finance, and public services, but standardising how trust is measured and managed across sectors remains a key challenge.

Explicability also intersects with fairness, as seen in fair machine learning.[103] By embedding fairness mechanisms into algorithms, these systems work to prevent discriminatory outcomes across demographic groups. A bias detection system introduces innovative methodologies such as counterfactual reasoning and automated frameworks. Yet, balancing fairness with accuracy, and standardising fairness definitions, remains a complex challenge, particularly in high-stakes domains like HR, tech, healthcare,[104]

94. O'Sullivan, James. *The Case for AI Illiteracy*. Substack, 29 Mar. 2025, https://substack.com/inbox/post/160133422

95. Bush, Stephen. "Anime Lessons in the Limits of AI." *@FinancialTimes*, Financial Times, Apr. 2025, on.ft.com/4iNW6Wl. Accessed 8 May 2025.

96. New AI Literacy Framework to Equip Youth in an Age of AI – OECD Education and Skills Today https://oecdedutoday.com/new-ai-literacy-framework-to-equip-youth-in-an-age-of-ai/

97. Miller, Tim. "Explanation in Artificial Intelligence: Insights from the Social Sciences." *Artificial Intelligence*, vol. 267, Feb. 2019, pp. 1–38, https://doi.org/10.1016/j.artint.2018.07.007

98. Holzinger, A., Saranti, A., Molnar, C., Biecek, P., Samek, W. (2022). Explainable AI Methods - A Brief Overview. In: Holzinger, A., Goebel, R., Fong, R., Moon, T., Müller, KR., Samek, W. (eds) xxAI - Beyond Explainable AI. xxAI 2020. Lecture Notes in Computer Science(), vol 13200. Springer, Cham. https://doi.org/10.1007/978-3-031-04083-2_2

99. Choung, Hyesun, et al. "Trust in AI and Its Role in the Acceptance of AI Technologies." *International Journal of Human–Computer Interaction*, vol. 39, no. 9, Apr. 2022, pp. 1–13, https://doi.org/10.1080/10447318.2022.2050543

100. Laux, Johann, et al. "Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk." Regulation & Governance, vol. 18, no. 1, Feb. 2023, https://doi.org/10.1111/rego.12512

101. Ali, Sajid, et al. "Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence." *Information Fusion*, vol. 99, no. 101805, Apr. 2023, p. 101805, https://doi.org/10.1016/j.inffus.2023.101805

102. Giudici, Paolo, et al. "Artificial Intelligence Risk Measurement." Expert Systems with Applications, vol. 235, 1 Jan. 2024, pp. 121220–121220, https://doi.org/10.1016/j.eswa.2023.121220

103. TNO. "Fair Machine Learning Combats Biases." TNO, 2025, https://www.tno.nl/en/technology-science/technologies/fair-machine-learning/

104. Sollini, Martina, et al. "Towards Clinical Application of Image Mining: A Systematic Review on Artificial Intelligence and Radiomics." *European Journal of Nuclear Medicine and Molecular Imaging*, June 2019, https://doi.org/10.1007/s00259-019-04372-x

[105] and criminal justice. Similarly, sectors such as legal decision-making and healthcare[106] demand highly interpretable AI outputs. In these areas, explainability supports compliance, ethical accountability, and clinical validation, ensuring that decisions informed by AI remain transparent and trustworthy.

Other concepts like epistemic AI,[107] counterfactual explainable AI and concise reasoning via reinforcement learning[108] continue to push explicability forward. These approaches embed uncertainty quantification, sensitivity to input and interpretability constraints directly into AI models, ensuring not only that AI outputs can be explained but that their confidence levels and potential risks are clearly communicated.

A new frontier in explainability is the use of attribution graphs.[109] These graphs provide a visual representation of AI decision-making pathways, allowing users to trace the specific factors and data points that influenced a model's output. This increases transparency and helps build trust; however, these methods introduce added complexity and require significant computational resources and will need to be integrated into existing workflows and standardised across different industries.

For policymakers, explicability is no longer a technical debate but an ethical dimension not to say a legal requirement to be considered. The EU must ensure that AI systems deployed across the EU remain transparent, auditable, and aligned with societal values, balancing model complexity, computational feasibility, and the right to explanation. However, it is essential to acknowledge that current explainability and interpretability tools face challenges, such as ensuring the explainability, accuracy and reliability of their outputs, which must be carefully considered in the development and deployment of these systems. Transparency is expected to remain a challenge in the future.

FROM SEARCH ENGINE TO GENAI

GenAI has the potential to transform the way humans interact with the digital world, as exemplified by the rapid pace with which GenAI features are being integrated into the digital services used by many European citizens.

Prominent examples are search engines. The way users look for information on the internet is evolving from a keyword-based information search and aggregation strategy, parsing through lists of results, into a conversational search using LLMs equipped with internet searching capabilities.[110] Since the initial ChatGPT revolution in 2023, large search engines have responded with their own GenAI features. Among the first adopters in this space was Microsoft's Bing,[111] providing natural language search services able to search for information on the web, presenting it to users as AI-generated summarizations and chat responses.

The revolution of internet search has brought both exciting opportunities and significant uncertainties for users, publishers, and the tech

105. Bohr, Adam, and Kaveh Memarzadeh. "Chapter 2 - the Rise of Artificial Intelligence in Healthcare Applications." *ScienceDirect*, edited by Adam Bohr and Kaveh Memarzadeh, Academic Press, 1 Jan. 2020, pp. 25–60, www.sciencedirect.com/science/article/pii/B9780128184387000022?via%3Dihub

106. Saraswat, Deepti, et al. "Explainable AI for Healthcare 5.0: Opportunities and Challenges." *IEEE Access*, vol. 10, 2022, pp. 1–0, https://doi.org/10.1109/access.2022.3197671

107. Alvarado, Ramón. "AI as an Epistemic Technology." *Science and Engineering Ethics*, vol. 29, article no. 32, 21 Aug. 2023, https://doi.org/10.1007/s11948-023-00451-3

108. Fatemi, Mehdi, et al. "Concise Reasoning via Reinforcement Learning." *ArXiv.org*, 2025, arxiv.org/abs/2504.05185. Accessed 8 May 2025

109. Lindsey, Jack, et al. "On the Biology of a Large Language Model." Transformer Circuits, 27 Mar. 2025, https://transformer-circuits.pub/2025/attribution-graphs/biology.html

110. AI is weaving itself into the fabric of the internet with generative search | MIT Technology Review https://www.technologyreview.com/2025/01/06/1108679/ai-generative-search-internet-breakthroughs/

111. What is Copilot (formerly Bing Chat), and How Can You Use It? | Microsoft Copilot https://www.microsoft.com/en-us/microsoft-copilot/for-individuals/do-more-with-ai/general-ai/what-is-copilot?msockid=34e529e478816c9f13203c5879c96dae&form=MA13KP

industry. Google is using AI to provide quick summaries of search results,[112] which may reduce clicks to publishers' sites, while new search engines like Perplexity and ChatGPT (OpenAI) are offering conversational search experiences with deep, comprehensive answers. However, this raises concerns among publishers about "zero-click searches", where users do not click through to original sources, threatening their traffic and revenue with potential copyright or unfair competition considerations. The new search experience is becoming more dynamic, with AI tools able to provide real-time data, multimedia, and even perform tasks, moving towards an "agentic" future where AI acts on users' behalf.[113]

However, the ongoing transformation is not limited to search engines. Online marketplaces, such as Amazon[114] and Zalando,[115] have started to equip their users with personal shopping assistants based on GenAI systems able to search through vast product catalogues and provide product recommendations to user questions. Similarly, social media platforms such as Snapchat,[116] Meta[117] and X,[118] have started to provide access to their own chatbots as well. The integration of GenAI into social media interfaces enables features that go beyond conversational use cases, for example by providing the users with possibilities to create or modify content that can directly be published on the platforms, blurring the line between user and AI-generated content, and potentially resulting in a new set of challenges and risks to address (see Section 5.2). Current legislation in place in the EU, notably the Digital Services Act (DSA), requires designated online platforms to closely monitor and address existing and emerging risks exacerbated by the use of GenAI in digital services, particularly when minors are among the users of these services.

Intersecting these opportunities of the evolution of virtual worlds, which let users experience virtual and real information with different levels of immersiveness and interaction, will continuously push the boundaries of human machine interaction.[119] In support of the EC Strategy for Virtual Worlds,[120] the Joint Research Centre (JRC) also investigates also the possible futures and related socio-economic impacts of Virtual Worlds,[121] including human wellbeing, skills and competences, and supporting industrial ecosystems. Ongoing work also pays particular attention to the combined use of Virtual World technologies with GenAI systems, which leads to particularly disruptive opportunities but also challenges. ■

112. Google I/O 2024: New generative AI experiences in Search https://blog.google/products/search/generative-ai-google-search-may-2024/

113. MIT Technology Review article, "AI generative search is the internet's biggest breakthrough in years" (January 2025)

114. Amazon Rufus: How We Built an AI-Powered Shopping Assistant - IEEE Spectrum https://spectrum.ieee.org/amazon-rufus

115. Zalando: Inspiring and empowering customers with AI-powered experiences | Zalando Corporate https://corporate.zalando.com/en/technology/inspiring-and-empowering-customers-ai-powered-experiences

116. Say Hi to My AI https://newsroom.snap.com/say-hi-to-my-ai

117. Europe, Meet Your Newest Assistant: Meta AI | Meta https://about.fb.com/news/2025/03/europe-meet-your-newest-assistant-meta-ai/

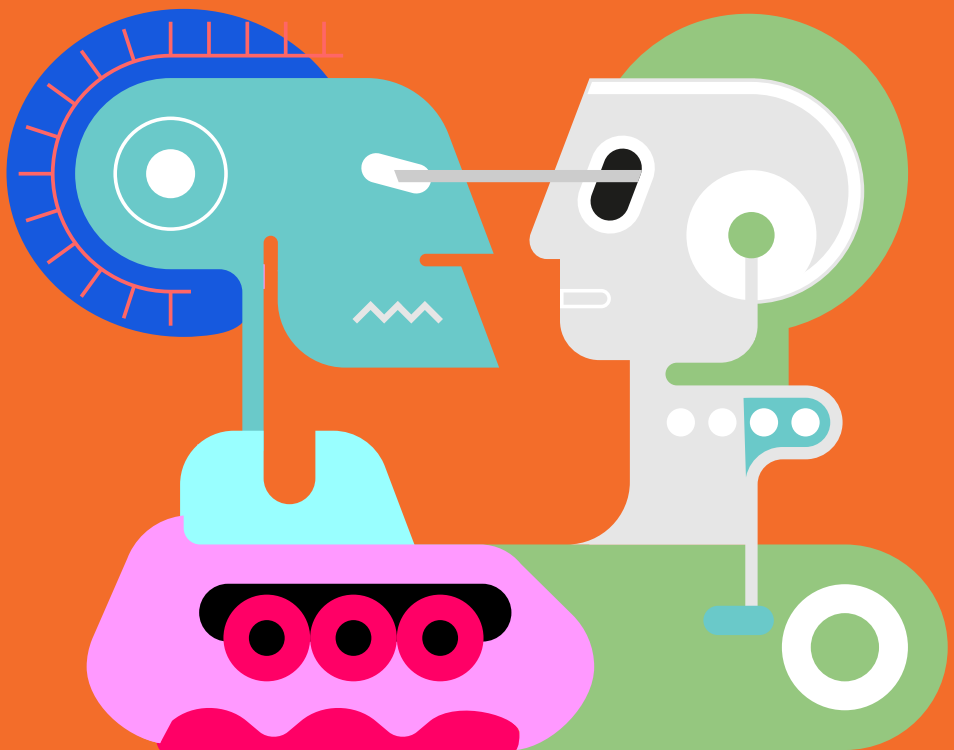
118. About Grok https://help.x.com/en/using-x/about-grok

119. https://joint-research-centre.ec.europa.eu/projects-and-activities/next-generation-virtual-worlds_en

120. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52023DC0442

121. https://publications.jrc.ec.europa.eu/repository/handle/JRC133757

3
ECONOMIC IMPLICATIONS



This Section builds on the foundational insights provided in Section 1, expanding on the economic ramifications of GenAI, and focusing on the EU's competitive position in the global landscape. It explores industry transformations and the emergence of new business models driven by GenAI. A detailed analysis of market share and trends, particularly on Conversational AI, provides insights into the competitive dynamics of the consumer market within Europe. The chapter also examines the impact of GenAI on the labour market and employment, addressing both opportunities for job creation, potential workforce displacement, and productivity considerations. Key questions include how to harness GenAI for economic growth while addressing the challenges it poses to employment stability.

3.1 EU's Competitive Position in the Global GenAI Landscape

KEY MESSAGES

- As the GenAI market continues to evolve, the EU must remain agile and proactive in its approach, ensuring that it remains at the forefront of this transformative technology.
- The EU faces significant competition. To reduce dependencies and build its technological sovereignty, it needs to invest in its vibrant research community and boost its innovation capacities.

This subchapter delves into the EU's position in the global GenAI market, examining its current competitive standing, exploring strategic opportunities for growth and leadership, and identifying existing challenges and barriers.

CURRENT EU STANDING

The EU has established itself as a significant player in the GenAI landscape. However, major global players such as the United States and China are constantly pushing the technological frontier with massive investments (see Section 1.2). The current standing of the EU can be assessed through several key indicators, including research output, innovation capacity, and market presence.

→ **Research Output and Innovation Capacity** The EU's research institutions and universities are pivotal in advancing GenAI technology, contributing significantly to the global knowledge pool. The EU ranks second globally in terms of GenAI-related academic publications (see Section 1.2).

→ **Market Presence and Industrial Impact** European companies are actively engaged in the development and deployment of GenAI technologies. Notable actors include Mistral AI and LightOn, which exemplify successful GenAI start-ups in the EU. The EU's diverse industrial landscape, particularly in sectors such as automotive, pharmaceuticals, and finance, offers numerous opportunities for GenAI integration. Despite these opportunities, the EU's market share in the global GenAI industry lags behind that of the US and China. Strengthening the commercialisation of research, driving strategic investment and funding, supporting unified regulatory sandboxes, attracting and retaining talent, forging strategic adoption in key sectors, and facilitating market access for EU-based companies are crucial steps toward improving the EU's market presence (see Section 1.2).

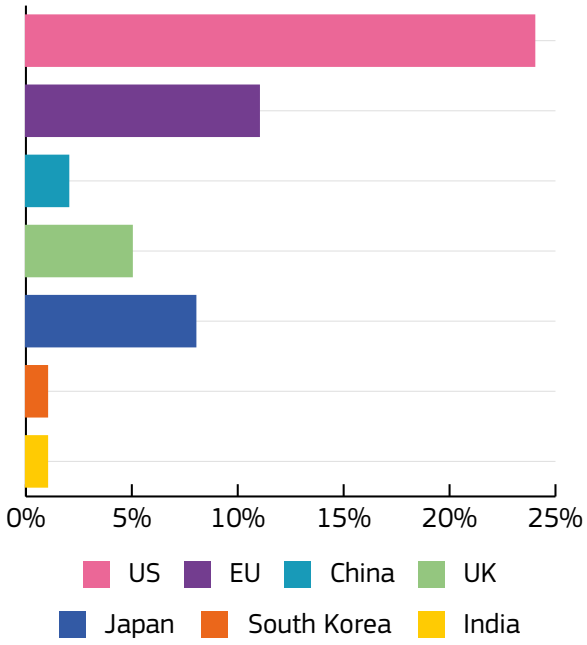
OWNERSHIP AND DEPENDENCIES

The growing foreign ownership of GenAI players raises important policy considerations related to national security, technological sovereignty, and economic competitiveness. As control over AI innovation and infrastructure increasingly shapes global influence, the extent to which domestic players are ultimately controlled by foreign interests may impact data governance and strategic decision-making. Foreign ownership may affect the development and deployment of GenAI technologies and existing and emerging regulatory frameworks, such as foreign investment reviews and export controls, which in turn can impact national security interests and approaches to international collaboration. Therefore, data on the ownership of GenAI players worldwide reveal key insights into the global landscape of GenAI development and control.[122]

This Section explores this topic by focusing on the global ultimate owner of all identified GenAI players, where a global ultimate owner is defined as a controlling shareholder holding more than 50% of the firm's shares. Therefore, foreign-owned EU players are those entities with a global ultimate owner located in countries outside the EU. Likewise, foreign-owned US players are owned by entities outside the US.

The US has a significant lead in the ownership of foreign GenAI players, holding 24% of foreign-owned GenAI players, pointing to its strong AI research and development ecosystem (see Figure 5). The EU holds the second largest share of foreign-owned GenAI players, signalling investment in GenAI enterprises abroad. Japan follows the EU (8%). China, despite being a major player in the global AI economy, has a relatively small share of foreign GenAI players with only 2%. Finally, 12% of all EU GenAI players are foreign-owned. China owns only 5% of foreign-owned EU GenAI players. In comparison, 14% of EU-owned foreign GenAI players are Chinese.[123]

Figure 5. Control of foreign GenAI players.



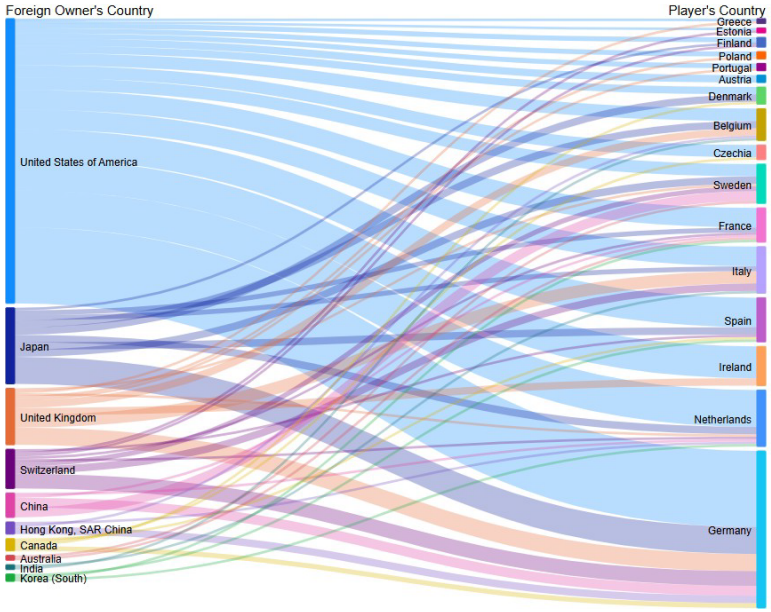
Source: JRC DGTES Dataset.

Results from a JRC data survey provide information about which foreign countries control players in EU countries (considering a firm located in the EU is foreign-owned when the owner is located outside the EU). Figure 6 shows that the US own the largest share of foreign-owned EU players (49%) followed by Japan (13%), the UK (11%), Switzerland (7%) and China (5%) of foreign-owned EU GenAI players.

122. Ownership is defined as a global ultimate owner holding more than 50% of shares of a firm. From a country perspective, analysing both foreign ownership of local players and local ownership of foreign players is useful to highlight access to key knowledge and possible dependencies.

123. This analysis also distinguishes between **foreign**, **domestic**, and **local** players from the EU perspective. **Foreign-owned** EU players are EU-based entities owned by investors located in countries outside the EU. **Domestic-ownership** refers to EU players that are owned by entities located within the EU (e.g. a French player owned by a German investor). **Locally-owned** players are those owned nationally (e.g. an Italian player that has Italian ownership). **Foreign ownership** is defined as a player receiving direct investment by an entity located in a country different from the country of residence of the player. The foreign investor exerts **control** by holding **more than 50%** of shares (IMF, 2009).

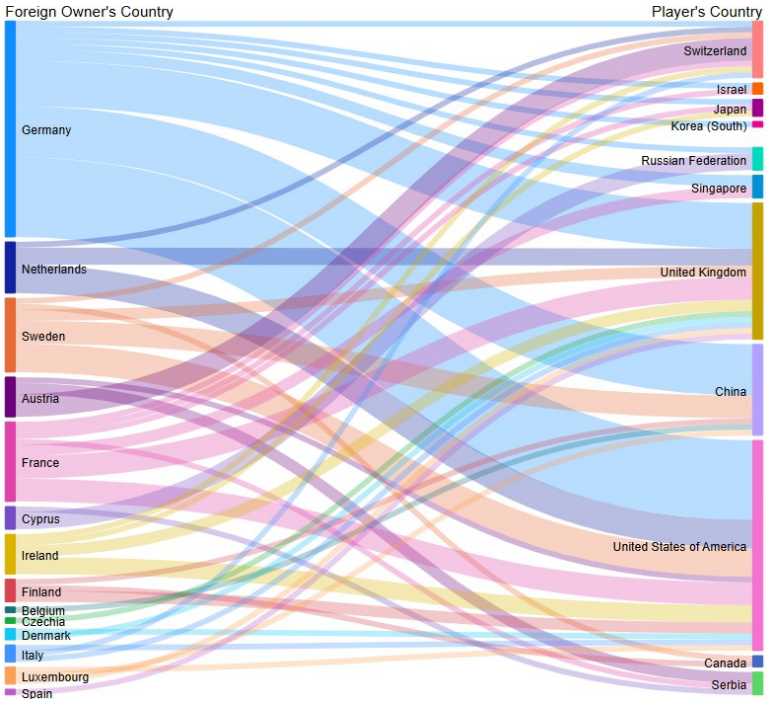
Figure 6. Foreign ownership of EU players.



Source: JRC DGTES Dataset.

Zooming in on the EU, Germany owns the highest proportion of players abroad, followed by France, Sweden, and the Netherlands.[124] Similarly, Figure 7 shows where in the EU foreign-owned players are located. Germany owns a majority of foreign-owned players, followed by Sweden and France.

Figure 7. Which EU countries own foreign players?



Source: JRC DGTES Dataset.

124. Abendroth-Dias et al. (2025) *DGTES Handbook: A Snapshot of EU Digital Competitiveness and Dependencies*, Publications Office of the European Union.

STRATEGIC OPPORTUNITIES

The EU can enhance its positioning in the GenAI domain by capitalising on several strategic opportunities.

- → **Ethical and Trustworthy AI**
  One of the EU's key differentiators is its commitment to ethical and trustworthy AI. The region has been proactive in developing regulatory frameworks that prioritise data privacy, security, and ethical considerations. By positioning itself as a leader in ethical GenAI, the EU can attract global partners and customers who value responsible AI practices, as well as domestic and foreign investment. This approach not only enhances the EU's reputation but also provides a competitive edge in a market increasingly concerned with AI ethics.

- → **Collaboration and Ecosystem Building**
  The EU's emphasis on collaboration and ecosystem building presents significant growth opportunities. Initiatives such as cross-border research collaborations and public-private partnerships can facilitate knowledge exchange and resource sharing, driving the EU's GenAI capabilities forward.

- → **Focus on Specific Markets and Applications**
  The EU can leverage its diverse industrial base to focus on niche markets and applications where GenAI can have a transformative impact. By identifying sectors with high potential for GenAI integration, such as healthcare, agriculture, and energy, as well as media and audiovisual creative sectors, the EU can tailor its strategies to address specific industry needs. This targeted approach allows the EU to develop specialised expertise and solutions, enhancing its competitiveness in these domains.

CHALLENGES AND BARRIERS

While the EU has significant potential in the GenAI landscape, it must address several challenges and barriers to maintain and improve its competitive standing. The European Commission recently published the AI Continent Action Plan, outlining a set of initiatives aiming to accelerate the adoption of AI, focusing on 5 main pillars: computing, data, sectoral approach to boost new industrial uses of AI and improve the delivery of a variety of public services, skills and regulatory simplification.

- → **Fragmented Market and Regulation**
  While the EU's single market should be the reference area for GenAI development, remaining internal barriers may pose challenges for its adoption and scaling. Efforts to identify and remove these barriers, along with legal instruments, such as the AI Act, providing a harmonised framework at EU level, push for a unified market for GenAI, enabling cross-border collaboration and innovation.

- → **Investment and Funding Gaps**
  As mentioned in Section 1.2, compared to the US and China, the EU faces challenges in attracting investment and funding for GenAI initiatives. Limited access to venture capital can impede the growth of GenAI enterprises. To address this issue, the EU must keep providing targeted funding programmes and incentives for GenAI start-ups and scale-ups while promoting private investment.

- → **Talent and Skills Shortage**
  A vibrant environment that attracts talent in the GenAI field is a significant element that can facilitate the EU's competitive standing. Developing robust AI talent through education, training, and upskilling initiatives is also critical to meet the growing demand for GenAI expertise. Collaborative efforts among education and training institutions, academia, industry, and governments can help bridge the skills gap and ensure a skilled workforce capable of driving GenAI innovation.

→ **Market Dynamics and Contestability**
The GenAI sector includes upstream and downstream activities associated with the provision of generative AI models. This emerging industry is quite dynamic, with an active value ecosystem and relevant R&D investments. However, some current market dynamics may influence the structure of GenAI markets while modifying its future competitive landscape. As mentioned before, the development and deployment of Gen AI systems include data, infrastructure, and algorithms, along with technical expertise. Depending on economic conditions, issues such as resource constraints, access to markets and technological developments may emerge and may reduce the presence of active competitors.[125]

3.2 Industry Transformation, New Business Models and Adoption

KEY MESSAGES

- GenAI is a catalyst for industry transformation, driving the emergence of innovative business models. At the same time, as in the case of creative industry, it requires careful consideration of the potential benefits and drawbacks.
- Digital maturity is a critical factor in GenAI adoption: SMEs need to develop a certain level of digital maturity, including digital skills, business processes, and infrastructure, to fully leverage the potential of GenAI.
- The uptake of AI technologies, including GenAI, is higher in larger enterprises in the EU, potentially widening the gap as smaller enterprises face challenges in adoption due to possibly limited resources and capacity.

125. For more information see https://competition-policy.ec.europa.eu/document/download/c86d461f-062e-4dde-a662-15228d6ca385_en

GenAI has the potential to transform industries across the EU, acting as a critical driver of innovation and economic transformation. By leveraging advanced algorithms and data-driven insights, GenAI could help by not only optimising existing processes but also creating new business models that may challenge traditional paradigms. This subchapter explores the impact of GenAI on various industrial sectors. Illustrative case studies and deep dives are presented in Section 6.

IMPACT OF GENAI ON TRADITIONAL INDUSTRIES – BENEFITS AND RISKS

The integration of GenAI could have a profound impact on traditional industries, revolutionising processes, and encouraging the emergence of novel business models.

GenAI – especially in its role as a component of emerging Agentic AI – is expected to transform the **manufacturing sector** (automotive, electronics, consumer goods, etc.) by enabling smart production lines, relying on advanced data analytics.[126] Agentic AI will also have a disruptive impact on predictive maintenance through autonomous and adaptive decision-making. Manufacturers can optimise supply chains, reduce waste, enhance product design and automate processes. This transformation is leading to the development of interconnected systems that can autonomously manage production tasks, resulting in increased efficiency and reduced downtime. These applications are already observed in the data, as 1.7% and 1.6% of EU activities in GenAI (as identified in Section 1.2) also belong to the mobility and electronics industrial ecosystems, respectively (DGTES).[127]

126. For more information on advanced manufacturing in the EU see Strategic Insights into the EU's Advanced Manufacturing Industry: Trends and Comparative Analysis (Fabiani et al, 2024).

127. GenAI activities identified in Section 1.2 can be tightly connected to other industrial ecosystems, either in the form of applications or innovating in technologies which are relevant for the industry. The co-occurrence of each industrial ecosystem-related keywords and GenAI technologies in academic publications, patent applications, and business descriptions constitute the overlap. More information in the upcoming report ATLAS: An Analytical Tool for Linking and Assessing industrial ecoSystems (Signorelli et al., forthcoming).

In the **retail** sector, GenAI is reshaping consumer experiences by personalising interactions and optimising inventory management. Retailers are using GenAI to analyse consumer behaviour, predict trends, and customise marketing strategies. This technology enables dynamic pricing models and automated customer service solutions, creating a customer-centric approach that can enhance loyalty and satisfaction. In the EU, 1.1% of its GenAI activities also have applications for the retail industry, above the global average (DGTES). GenAI is playing a pivotal role in **healthcare** by improving diagnostic accuracy and personalising patient care, as well as by enabling the analysis of vast datasets to detect patterns and predict disease progression, aiding in early diagnosis and treatment planning (see Section 6.1). However, the impact can be even more disruptive, as this is the industrial ecosystem with the highest activity overlap with GenAI: almost 10% of GenAI research, innovation, and business activities in the EU relate to the healthcare ecosystem (DGTES).

In **creative industries**, GenAI is revolutionising content creation and design processes. It enables artists and designers to generate innovative works by analysing audience preferences and trends. These industries are relevant for GenAI in the EU, with over 3% of GenAI activities being related to creative industries (DGTES). AI-generated content, such as music, video, and art, is becoming increasingly popular, leading to new business models focused on digital and interactive experiences, as well as the successful integration of GenAI in existing processes in the audiovisual and media industries (e.g. virtual movie production) At the same time, GenAI has sparked concerns about its potential detrimental impact. For instance, one of the main issues is that training data for AI models may include the work of creators, which raises copyright concerns[128] (Section 5.4). This could lead to a serious modification of the incentives for innovation and creativity, as AI-generated adaptations of original works could potentially displace the latter in commercial settings. Finally, the rapid adoption of GenAI in the creative industry could also lead to a homogenisation of styles, as AI models may rely on existing trends and styles rather than create something entirely new (see Section 2.3), and the generated content may become part of the training data for subsequent models. The case of GenAI in creative industries highlights how in some cases the impact is complex and requires careful consideration.

ROLE OF SMES AND DIGITAL MATURITY

Small and medium-sized enterprises (SMEs) represent 99% of all businesses in the EU, and are beginning to explore the potential of GenAI. Networks such as the European Digital Innovation Hubs (EDIHs) play a critical role in supporting SMEs in unlocking the value of GenAI. Through case studies of SMEs supported by EDIHs, it is evident that GenAI can deliver tangible benefits across diverse sectors, including healthcare, food technology, manufacturing, and education.

128. https://www.theintrinsicperspective.com/p/welcome-to-the-semantic-apocalypse

Table 1. Applications of Generative AI by SMEs and Support received from EDIHs

Customer	Sector	Interest in GenAI	Challenge	Support provided by EDIH
FreezerData BV	Energy	Improving business processes	Calibrating and building the technology. (R&D)	Test before invest
Alpha-Protein GmbH	Agricultural biotechnology and food biotechnology	Improving business processes	Calibrating the technology and exploring possibilities. (R&D)	Test before invest
Aqualeg	Health care	Improving the final product/service	Calibrating the technology. (R&D)	Test before invest
MultiSkript Verlag	Cultural and creative economy	Improving the final product/service	Exploring possibilities. (R&D)	Test before invest
Confidential-Mind Oy	Security	Gen AI-enabled product.	Funding. (Implementation)	Support to find investment

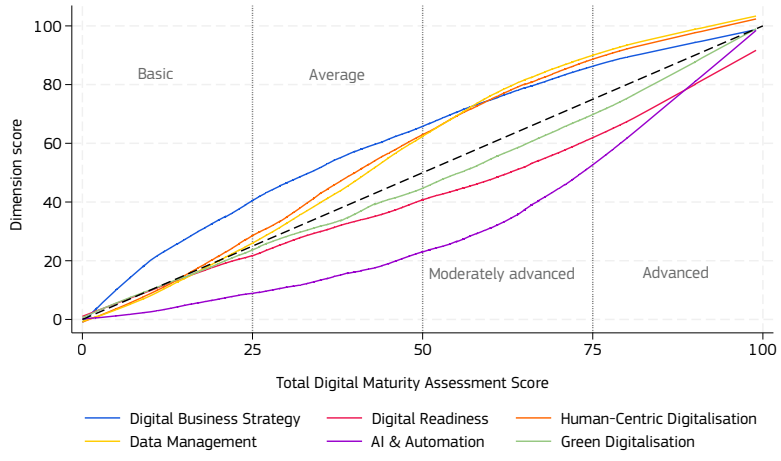
Customer	Sector	Interest in GenAI	Challenge	Support provided by EDIH
Miðeind	Cultural and creative economy	Gen AI-enabled product.	Computing and human resources. (Implementation)	Support to find investment, test before invest
Multiple customers	Multiple sectors	Exploring possibilities, raising digital awareness.	Lack of knowledge and awareness.	
(R&D)	Networking and access to innovation ecosystems; Training and skills development			

Source: JRC.

The analysis of 176 SMEs case studies (see Table 1) reveals that while only a few SMEs are currently adopting GenAI, they have found various applications for the technology. These applications range from optimising business processes to enhancing products or services, and developing GenAI-enabled products. For instance, FreezerData, a Dutch company, explored the use of GenAI to implement a virtual service mechanic to address labour shortages. Similarly, Alpha-Protein, a German start-up, aims to leverage GenAI to optimise its production process. Other SMEs, such as Aqualeg and MultiSkript, are using GenAI to enhance their customer service and product offerings.

The adoption of GenAI requires a certain level of digital maturity, including digital skills, business processes, and infrastructure. Digital maturity assessments such as the one conducted by EDIHs[129] reveal that firms are generally moderately advanced in their digital transformation when they begin adopting AI technologies like GenAI. The assessments also indicate that achieving a certain level of development in strategy, data management, and digital skills is essential for fully leveraging AI and other advanced technologies,[130] as illustrated in Figure 8.

Figure 8. Development of dimensions in relation to total digital maturity assessment scores.[131]



Source: JRC elaboration.

129. Kalpaka, A., Rissola, G., De Nigris, S., & Nepelski, D. (2023). Digital Maturity Assessment (DMA) Framework and Questionnaires for SMEs/PSOs: A guidance document for EDIHs. European Commission, JRC133234.

130. Carpentier, E., D'Adda, D., Nepelski, D. and Stake, J., European Digital Innovation Hubs Network's activities and customers, Publications Office of the European Union, Luxembourg, 2025, https://data.europa.eu/doi/10.2760/7784020, JRC140547.

131. The figure shows locally weighted regression lines of dimension scores on total digital maturity assessment score for 13,668 assessments of EU firms performed by the EDIHs. The black dashed reference line represents the average scores of the dimensions if they would all contribute equally to the total score. When the dimension lines are above the reference line they contribute relatively more to the total digital maturity assessment score, while contributing relatively less when below.

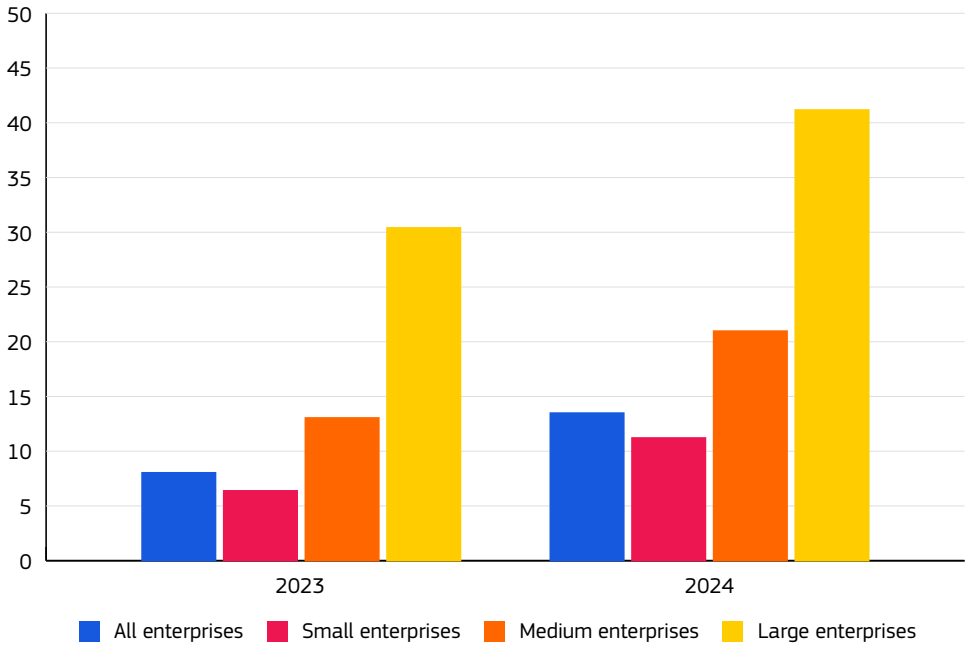
Figure 8 illustrates the average dimension scores in the Digital Maturity Assessment Tool in relation to the total digital maturity score. In general, strategy contributes relatively more to the total score for firms with a basic digital maturity score, and AI & Automation only starts to develop when firms have reached a moderately advanced level of digital maturity. It also shows that firms need a certain level of Data Management and Human-Centric Digitisation, including skills and employee empowerment, to be able to fully integrate AI technologies such as GenAI.

This is further confirmed by the aforementioned case studies, where SMEs exploring and adopting GenAI generally exhibit higher digital maturity, with a solid digital infrastructure, including data management and governance systems, and skilled employees.

In 2024, Eurostat data[132] showed that the use of AI technologies among EU enterprises increased over the previous year, with substantial differences based on company size. This is reflected in all AI technologies, therefore including GenAI as well as in the different economic activities. Large companies, where 41% are using AI, lead the uptake compared to about 21% of medium-sized enterprises and 11% of small enterprises (see Figure 9). Large companies could gain a competitive advantage, improving their efficiency and decision-making, while smaller companies may find it harder to keep up, potentially increasing the gap between them in the market.

132. Use of artificial intelligence in enterprises, Statistics Explained, Eurostat 2025, https://ec.europa.eu/eurostat/statistics-explained/index.php?oldid=568530

Figure 9. Enterprises in the EU using AI technologies by size, EU, 2023 and 2024 (% if enterprises).



Source: Eurostat.

Understanding the regional dynamics of AI adoption in general, and GenAI in particular, is crucial for harnessing its full potential across the European Union. The transformative nature of these technologies can exacerbate existing territorial divides in Europe if not properly managed. As GenAI continues to transform industries and societies, it is essential to examine how EU investments in AI are allocated at the regional level in order to identify areas of strength and weakness, address potential disparities, and develop targeted strategies to support regional growth and development, ensuring that no territory

is left behind. This, in turn, can help promote digital cohesion, foster innovation, and create new opportunities for businesses and citizens across the EU, ultimately contributing to a more competitive and prosperous economy. The JRC has conducted an analysis[133] of the geographical distribution of EU investments related to artificial intelligence (AI) during the 2014-2020 programming period. The analysis, which covers the EU27 regions at the NUTS 2 level, reveals the following key findings:

→ Approximately €8 billion of EU funds (from Horizon 2020 and cohesion policy) were allocated to AI investments in European regions during the 2014-2020 period, accounting for an annual average of 7% of total AI investment in the EU.

→ The share of EU funding in the total AI investment tends to be higher in Central and Eastern European countries and, to a lesser extent, in Southern Europe.

→ More developed regions have a higher specialization in AI EU-funded investments, which generates spillover effects that enhance similar patterns in neighbouring regions.

→ AI-related investments are more concentrated in regions with a higher concentration of ICT activities and that are more innovative, highlighting the importance of agglomeration effects.

→ Regions that have selected AI as an innovation priority for their Smart Specialization Strategies are also more likely to have a higher funding specialization in AI.

133. Anabela Marques Santos, Francesco Molica, Carlos Torrecilla-Salinas, EU-funded investment in Artificial Intelligence and regional specialization, Regional Science Policy & Practice,Volume 17, Issue 7, 2025.

3.3 Market Shares, Trends, and Competitive Analysis: the Case of Conversational AI in Europe

KEY MESSAGES

- The Generalist Conversational AI (GCAI) market in the EU is dominated by a few key players, with ChatGPT by OpenAI emerging as the clear market leader, although other players, such as ChatOn and NovaAI, are also competitive.
- The competitive landscape of conversational GenAI tools varies at the country level, with different dynamics shaping the competitive positioning of tools across EU member states, and local actors displaying greater prominence within specific countries.
- Incumbent technology companies with established assets in adjacent market, notably messaging platforms, cloud services, and search engines, are increasingly integrating GCAI functionalities into their existing ecosystems.

Generalist Conversational AI systems (GCAI), such as ChatGPT (OpenAI), Gemini (Google), Claude (Anthropic), and DeepSeek Chat (DeepSeek), are designed to engage in open-ended, human-like dialogue across a wide range of topics. Unsurprisingly, GCAI tools have attracted considerable public and media attention, highlighting their societal relevance and eliciting closer analytical scrutiny.

The competitive dynamics and innovation trajectories within the GCAI ecosystem depend on the relationship between user services and the foundational LLMs on which they are built. On the one hand, vertically integrated systems such as ChatGPT, Claude, and DeepSeek Chat are developed and deployed by the same organisations that train the underlying

models. These **first-party applications** benefit from full-stack control, facilitating a tighter alignment between model capabilities and user interface design. On the other hand, **third-party applications** such as ChatOn and NovaAI rely on external LLMs - typically accessed via Application Programming Interfaces (API) from providers like OpenAI and DeepSeek - to build user-facing services without developing their own foundational models. This reliance on third-party players via API or other means, in turn, leads to an overall lack of control, mainly over the end-product behaviour and on the supply of the underlying model. This Section focuses on GCAI services within the European Union (EU), aiming to analyse the structure and evolution of this emerging market segment and trying to identify the key operators in the EU and how their services and market penetration differ across EU Member States.

MAIN PLAYERS

The market for GCAI tools is complex due to the wide range of applications and specialised tools that cater to specific tasks. To maintain clarity and provide information about the current state of play of the adoption of GenAI solutions, the focus of this Section is on relevant GCAI tools that provide a variety of services in the business-to-consumer (B2C) market, excluding niche players like text editing or image generation tools.[134]

The emphasis is on understanding the impact and reach of these players in the EU, offering insights into the competitive environment of GCAI tools in the consumer space.

MAIN PLAYERS AT EU LEVEL

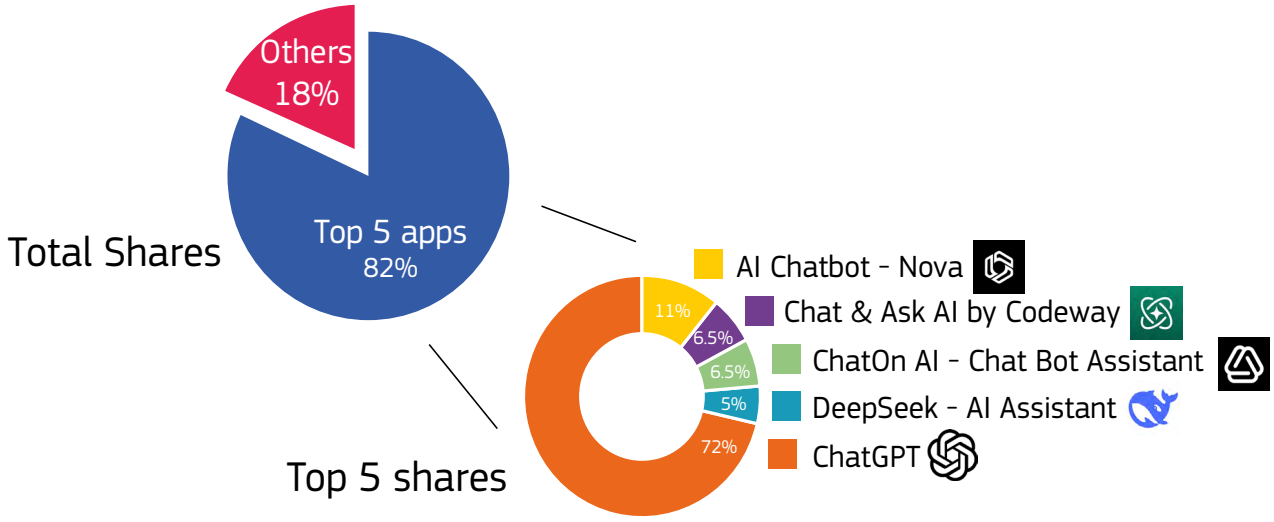
By using monthly app downloads and website traffic metrics, significant players that meet the established criteria can be identified to provide a snapshot of the current market dynamics. The identified apps and services include a mix of native LLM-based assistants and LLM-powered services. The top five players, predominantly from the US and China, dominate the market, accounting for 82% of MAU and downloads. ChatGPT by OpenAI is the clear market leader in both categories. Despite the dominance of big tech companies, interface-only solutions like ChatOn have gained significant traction, indicating a competitive dynamic between vertically integrated players and those focusing solely on consumer interfaces.[135]

While the top players include well-known tech giants like Google and Microsoft, their apps do not rank among the most used (Figure 10), suggesting that brand recognition alone does not ensure sustained user engagement. Interestingly, interface-only players have successfully penetrated the market, showing that companies not primarily focused on AI development can achieve significant success. The success of non-AI specialist companies highlights the complexity and competitiveness of the GCAI market landscape, with new entrants challenging established players for market share.

134. The focus is on companies with publicly accessible conversational interfaces, avoiding those focused solely on business-to-business (B2B) markets or foundational LLM models without consumer interfaces. This is driven by the availability of data, as B2C markets allow for analysis based on web traffic and app usage, offering insights into market shares and usage patterns.
Main players in the EU market are identified by analysing app downloads and web traffic. Companies with apps exceeding 500,000 worldwide downloads or websites with over 100,000 EU visits in the past six months are considered significant players. This threshold ensures that only companies with a meaningful presence and impact in the EU market are included. The focus remains on consumer-facing services, acknowledging that companies with both B2B and B2C operations are not evaluated for their B2B market share, nor is there a distinction between consumer and business-generated usage data.

135. The data examines market shares using metrics like total traffic, visits, downloads, and app users from April 2024 to March 2025.

Figure 10. Market share of top 5 apps by MAU.

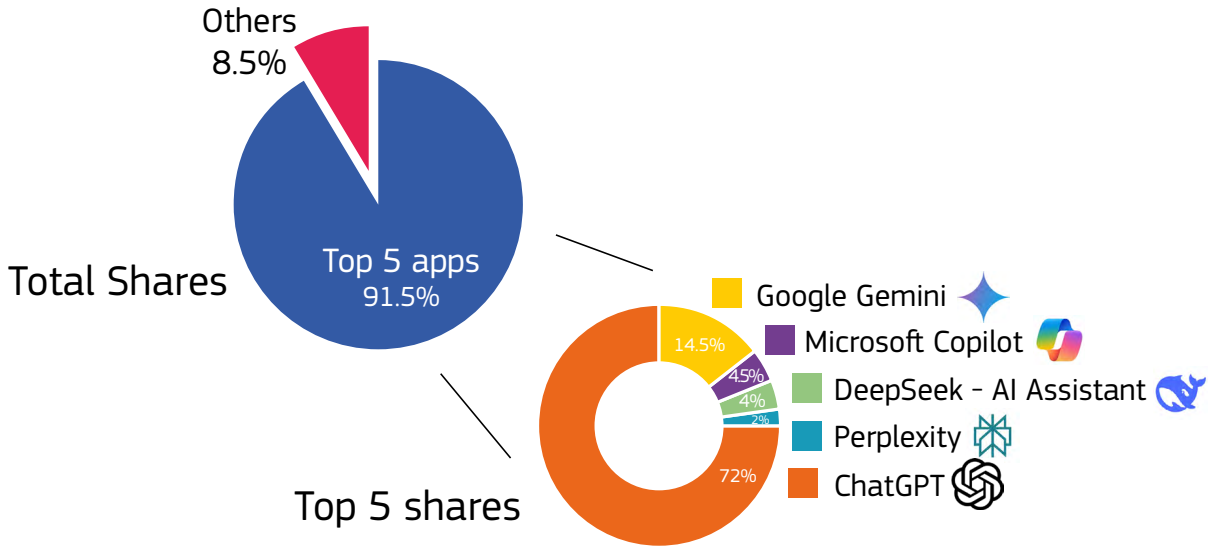


Source: JRC elaboration based on average monthly active users according to Sensor Tower, from 2024-04-01 to 2025-03-31. Users are counted at the device level8, adding all iOS and Android devices.

The website market is similarly concentrated (see Figure 11). The top 5 firms account for 92% of total unique visitors in the EU, reflecting an even higher concentration rate than observed in the app market. However, in contrast to the app market, interface-only GCAI firms are notably less prominent in the website landscape. Several factors may help explain this divergence. First, leading GCAI tools such as Copilot and Gemini are tightly integrated into browser environments or operating systems (e.g. Edge, Chrome), positioning websites as the default access point for users rather than standalone conversational AI interfaces. Interface-only firms, lacking these integration advantages, may find mobile apps a more effective channel to engage users directly.

Moreover, beyond integration factors, interface-only players appear to face additional challenges in establishing visibility and trust on the web. One possible explanation is that accessing AI tools via websites often requires users to navigate directly to specific URLs or rely on discovery through search engines - pathways where established firms such as Google and Microsoft benefit from strong brand recognition and well-developed search engine optimisation (SEO) capabilities. In contrast, app stores may offer a more advantageous environment for newer entrants, providing centralised, high-traffic platforms with features such as rankings, user reviews, and categorisation that can enhance visibility among consumers. Additionally, app-based marketing strategies - particularly those leveraging social media platforms like TikTok and YouTube - may play a key role in enabling interface-only GCAI apps to attract users quickly, potentially reducing their reliance on a strong web presence.

Figure 11. Market share of top 5 websites.



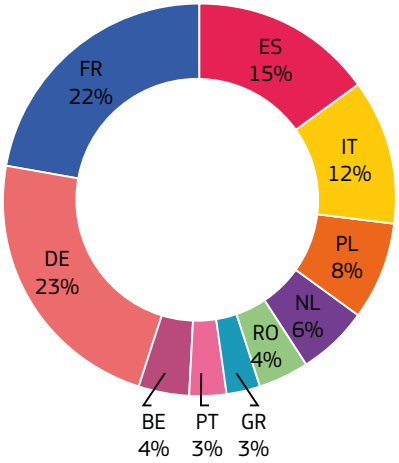
Source: JRC elaboration based on unique visitors (desktop and mobile devices) by Similarweb, from April 1, 2024, to March 31, 2025.

At the country level, the competitive landscape of GCAI tools varies significantly. Germany, France, Spain, Italy, and Poland lead in GCAI tool adoption based on downloads and MAUs. Together, they represent a substantial portion of total downloads and MAUs in the EU, with Germany being the largest market. While major players maintain a presence across all Member States, local actors like Mistral in France show greater prominence in specific countries. The rankings of apps differ based on downloads and active user engagement, reflecting varying user preferences and engagement levels across different regions.

Figure 12. GCAI App Market Shares by EU Member State (Downloads and MAU, in Millions).

country	downloads	%	mau	%
DE	38.80	24.51	15.80	20.0
FR	28.84	18.22	14.70	18.6
ES	19.50	12.32	10.20	12.9
IT	16.44	10.38	8.02	1.01
PL	10.27	6.49	5.56	7.0
NL	7.13	4.51	3.90	4.9
RO	4.89	3.09	2.78	3.5
GR	4.57	2.89	1.79	2.3
PT	3.90	2.46	2.22	2.8
BE	3.87	2.45	2.45	3.4
SE	3.08	1.95	2.18	2.8
CZ	2.76	1.75	1.47	1.9
AT	2.74	1.73	1.65	2.1
HU	2.52	1.59	1.11	1.4
BG	1.85	1.17	0.83	1.1
FI	1.63	1.03	0.92	1.2
SK	1.55	0.98	0.67	0.8
DK	1.51	0.95	1.06	1.3
IE	1.27	0.80	0.91	1.1
HR	1.17	0.74	0.67	0.9
Total	**158.29**	**100**	**79.11**	**100**

Main countries based on mau (%)



Source: Sensor Tower. The table presents the sum of downloads and monthly active users across selected apps from April 1, 2024, to March 31, 2025. Figures are not adjusted for multi-homing.

The results for website usage shows a similar picture. Based on monthly unique visitor data, the five EU Member States exhibiting the highest levels of GCAI website usage are Germany (DE), France (FR), Spain (ES), Italy (IT), and the Netherlands (NL). As shown in Figure 12, each of these countries recorded over 5 million visitors during the study period. Together, they account for approximately 64% of the total EU market, with Germany leading at 18%.

A second group of countries includes Poland (PL), Sweden (SE), Belgium (BE), Portugal (PT) and Denmark (DK). Poland closely follows the Netherlands in terms of number of visitors (4.2 million and 4.6, respectively), while the remaining countries each account for less than 3.5% market share.

Overall, the website segment is characterised by a more concentrated and homogeneous competitive structure than the app market. The five leading providers - ChatGPT, Gemini, Copilot, DeepSeek, and Perplexity - account for the majority of website-based GCAI usage across the EU. In contrast to the app segment, Mistral, which demonstrated notable traction - particularly in France - does not appear among the top website providers, suggesting that certain players may exhibit platform-specific engagement patterns.

The ability of EU companies to compete in the Generative AI innovation race has far-reaching implications, spanning economic growth, talent retention, and even national security. Therefore, it is essential to grasp the current state of play, as well as the strengths and weaknesses of the EU's ecosystem. While analysing the B2C market for Generative Conversational AIs (GCAIs) may not fully capture the cutting-edge innovations emerging in the EU's B2B sector, or in the entire EU GenAI ecosystem, it still offers valuable insights due to the significant media attention and public prominence that B2C interfaces receive.

3.4 Impact on the Labour Market: Employment and Productivity

KEY MESSAGES

- Employment policies may need to consider the labour market dynamics brought about by GenAI. These include potential impacts on income and inequalities, occupational restructuring, and potential transformation of occupations.
- GenAI advancement is also bringing a noticeable shift in demand toward skills needed to navigate and engage with AI, such as critical thinking and emotional intelligence, leading to a possible divide between high-skill and low-skill jobs. These skills include AI literacy but also broader understanding of the ethical and even regulatory implications of the technology in work practices.
- Encouraging workforce resilience and adaptability will help in addressing the changing nature of labour market needs caused by the advent of GenAI.
- Overall, while GenAI poses challenges related to job displacement and inequality, it also offers significant potential for productivity enhancements and employment stability. Policymakers and organisations must navigate these dynamics carefully to maximise the benefits of GenAI while mitigating its risks.

Anticipating the impact of GenAI on the labour market and employment, it is crucial to prepare the workforce, foresee potential job displacements, and adapt educational systems and curricula to address current and future needs. One approach to do so is to analyse which occupations are most impacted by the advances of GenAI. For instance, does GenAI equally impact engineers, cooks, teachers, and cleaners? Understanding which occupations are most and least exposed to GenAI

can help policymakers design employment and education policies to ensure a smoother transition.

There is widespread concern about job displacement due to the advent of GenAI.[136] Nevertheless, recent studies show that the launch of ChatGPT has not significantly reduced employment in large US companies. Instead, these firms have seen increases in labour productivity, particularly those with higher GenAI exposure, indicating that it can enhance efficiency and productivity without necessarily cutting jobs.[137]

For example, in Germany, automation technologies like industrial robots have led to job creation in service sectors, balancing the loss in manufacturing, and suggesting potential pathways to offset the disruptive effects of GenAI.[138]

GenAI might also exacerbate labour market inequalities by increasing the productivity of occupations requiring cognitive abilities that align with GenAI capabilities, and workers who are empowered to use such tools, potentially leading to more precarious conditions for low-wage workers. Despite these concerns, GenAI offers opportunities for reducing inequality by enabling task substitution in high-wage roles and democratising skills access, helping lower-performing workers catch up.

A JRC research project[139] [140] tackled this issue by mapping the intensity of AI research, the corresponding cognitive abilities and work tasks for over 100+ occupations. For instance, there has been a substantial amount of (generative) AI research related to comprehension and expression (cognitive ability), which is required to train others (task), which in turn is an important task for teachers (occupation). In contrast, there has been relatively less AI research dealing with sensorimotor interaction (cognitive ability), the ability needed to carry and move objects (task), which cleaners (occupation) particularly perform to do their job. This explains why (generative) AI has less of an impact on cleaners than on teachers. The research did not take into account new and emerging trends such as Agentic AI, which is expected to have yet another level of impact across occupations.

AI RESEARCH, COGNITIVE ABILITIES AND TASKS

Over the past decade, AI research has mainly focused on cognitive abilities linked to understanding and generating **ideas**, which are fundamental for GenAI in particular:

- → **Comprehension and expression:** processing natural language, summarising the main messages, and expressing ideas and positions – for instance, reading a report and answering a question regarding its content.

- → **Attention and search:** seeking relevant information within a text or an image according to particular criteria – for instance, within a large report, finding the most important bits related to a specific question, or classifying the nature of documents (e.g. resume, scientific report).

- → **Conceptualisation, learning and abstraction:** generalising from examples, learning from demonstration, and accumulating (abstract) knowledge – for instance, storing the information acquired by answering several questions on a given report.

136. Cedefop. (2025). *Skills empower workers in the AI revolution first findings from Cedefop's AI skills survey.* Publication Office of the European Union. Policy brief. DOI: 10.2801/6372704. https://www.cedefop.europa.eu/en/publications/9201

137. Yu, Jason, and Cheryl Qi. "The impact of generative AI on employment and labor productivity." *Review of Business* 44.1 (2024): 53-67.

138. Dauth, Wolfgang, et al. "The adjustment of labor markets to robots." Journal of the European Economic Association 19.6 (2021): 3104-3153.

139. Dessart, F., Fernández Macías, E., & Gómez E. (2025). Anticipating the impact of AI on occupations: a JRC methodology. JRC Science for Policy Brief. JRC142580

140. Tolan, S., Pesole, A., Martínez-Plumed, F., Fernández-Macías, E., Hernández-Orallo, J., & Gómez, E. (2021). Measuring the occupational impact of AI: tasks, cognitive abilities and AI benchmarks. Journal of Artificial Intelligence Research, 71, 191-236.

These abilities, in turn, allow AI to perform certain tasks. For instance:

- → Comprehension and expression are needed to instruct, train and teach people, and write letters, memos and emails, but also to resolve conflicts and negotiate with people. LLMs have significantly impacted our ability to understand natural language and generate expressive, coherent text.
- → Attention and search are needed to perform mathematical and statistical tasks, but also to filter through large amounts of information on the internet (as search engines do).
- → Conceptualisation, learning and abstraction: with machine learning, AI systems can learn from data and recognise patterns which are not explicitly programmed. AI can also transfer knowledge from one domain to another.

EXPOSURE OF OCCUPATIONS TO AI

Mapping the tasks involved in each occupation together with the tasks that AI can perform enables computing an exposure score for each occupation. This exposure score is not absolute, but relative – that is, it shows the extent of exposure of occupations to AI with respect to one another.

The occupations most exposed to (generative) AI analysed by a JRC study were electrotechnology engineers, software developers, teachers, office clerks, and secretaries. For instance, teachers were more exposed to AI than 90% of workers.[141] [142] The impact that AI had on these occupations was mainly driven by AI-enabled "ideas"-related abilities that are required for these occupations, such as comprehension and expression, attention and search, and conceptualisation, learning and abstraction. Conversely, AI research impacted relatively less occupations such as cleaners and helpers, waiters and bartenders, and shop salespersons. This is because AI research in abilities required for these occupations – for instance, sensorimotor interaction and navigation – was still scarce.

The potential **impact of GenAI seems different** from previous waves of technological progress. Doctors, teachers and engineers, for instance, were not particularly exposed to previous waves of technological outbreaks (e.g. robotisation) as much as cashiers, machine operators and assemblers, to name a few.[143] [144] GenAI has the potential to revert this pattern, by impacting high-income occupations more than low-income occupations. This is because AI has so far made progress on abilities related to ideas, such as conceptualising, learning, abstracting, comprehending and searching information, etc., abilities that doctors, engineers and teachers particularly need to perform their job tasks. It is important to note that other, non-AI based technological progresses may affect low-income occupations (for instance, self-checkout machines).

GENAI AND PRODUCTIVITY

Recent studies highlight substantial productivity gains facilitated by LLMs such as ChatGPT. For instance, a study focusing on the customer service industry shows that the introduction of AI-based tools led to a 14% increase in productivity on average, with novice and low-skilled workers experiencing a 34% boost.[145] Similarly, in

141. Dessart, F., Fernández Macías, E., & Gómez E. (2025). Anticipating the impact of AI on occupations: a JRC methodology. JRC Science for Policy Brief. JRC142580

142. Tolan, S., Pesole, A., Martínez-Plumed, F., Fernández-Macías, E., Hernández-Orallo, J., & Gómez, E. (2021). Measuring the occupational impact of AI: tasks, cognitive abilities and AI benchmarks. Journal of Artificial Intelligence Research, 71, 191-236.

143. Dessart, F., Fernández Macías, E., & Gómez E. (2025). Anticipating the impact of AI on occupations: a JRC methodology. JRC Science for Policy Brief. JRC142580

144. Tolan, S., Pesole, A., Martínez-Plumed, F., Fernández-Macías, E., Hernández-Orallo, J., & Gómez, E. (2021). Measuring the occupational impact of AI: tasks, cognitive abilities and AI benchmarks. Journal of Artificial Intelligence Research, 71, 191-236.

145. Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond (2025). "Generative AI at Work". The Quarterly Journal of Economics, Volume 140, Issue 2, Pages 889–942.

professional writing tasks, ChatGPT is found to have significantly increased average productivity (tasks are performed faster and quality is higher), primarily by substituting for worker effort rather than complementing worker skills.[146] In line with this, another study examines how ChatGPT has changed the demand for freelancers by dividing over 3 million job postings into 116 fine-grained skill clusters, labelling them as substitutable by, complementary to or unaffected by LLMs.[147] The results indicate that labour demand increased after the launch of ChatGPT, but only in skills clusters that were complementary to or unaffected by the AI tool. Overall, the results suggest a shift toward more specialised expertise for freelancers rather than uniform growth across all complementary areas. LLMs have crossed the threshold to become useful across a wide range of cognitive tasks[148] and the key is to identify the comparative advantages GenAI tools have in generating content.

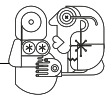
GenAI holds promise for enhancing productivity and quality in various domains. However, the efficacy of these systems depends on the task complexity and user skill level, with potential disparities in performance across different demographic groups. While GenAI can reduce intra-occupational performance gaps, it may exacerbate inequalities between educational and occupational groups, necessitating a nuanced approach to mitigate adverse effects and maximise the technology's potential. ■

146. Noy, Shakked and Whiney Zhang (2023). "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence". Science 381, Issue 6654, pp. 187-192.

147. Teutloff, Ole, Johanna Einsiedler, Otto Kässi, Fabian Braesemann, Pamela Mishkin, and R. Maria del Rio-Chanona (2025). "Winners and losers of generative AI: Early Evidence of Shifts in Freelancer Demand". Journal of Economic Behavior & Organization, 106845.

148. Korinek, Anton (2023). "Generative AI for Economic Research: Use Cases and Implications for Economists". *Journal of Economic Literature* 61.4, pp. 1281–1317. For example, for doing research in economics, the author describes use cases in six main areas: ideation, writing, background research, coding, data analysis and mathematical derivations.

4
SOCIETAL IMPACTS AND CHALLENGES



This chapter investigates the societal implications of GenAI, emphasising the skills gap and the need for AI literacy among citizens and the workforce. It discusses digital commons and the intersection of AI with environmental concerns, as well as Gen AI in the media and the overall perception, public discourse and narrative around its developments. The chapter also addresses the rights of children and issues of gender bias within AI systems, and the potential to generate false or misleading content. A behavioural approach to GenAI policy analysis is proposed as a means to navigate privacy and data protection challenges. The chapter raises critical questions about how to ensure inclusive and ethical AI applications that align with societal values.

4.1 Skills Gap and AI Literacy for Citizens and the Workforce

KEY MESSAGES

- Addressing the skills gap and enhancing AI literacy in the workforce is a complex challenge that requires coordinated efforts from multiple stakeholders, including businesses, educational institutions, and policymakers.
- By adopting comprehensive strategies that focus on upskilling, reskilling, and fostering AI literacy from an early age through education and continuous through lifelong learning, societies can help ensure that their workforces and citizens are better prepared to harness the potential of GenAI.

INTRODUCTION TO SKILLS GAP AND DIGITAL SKILLS

GenAI has the potential to bring a substantial transformation in the workforce landscape across various industries (see Section 3.4). This shift highlights a potentially significantly evolving skills gap, primarily in digital competencies, which is becoming increasingly evident as GenAI technologies proliferate. While the propensity for AI and GenAI to replicate some human skills will lead to skills displacement, their widespread diffusion is expected to create a growing demand for digital and data science skills related to the development and maintenance of AI systems, as well as complementary cognitive and transversal skills to enable workers to use and interact with GenAI systems.[149] Such skills, as well as analytical thinking, resilience, flexibility, agility, along with leadership and social influence are the skills required by employers.[150]

This skills gap is not merely about understanding how to use GenAI tools but extends to comprehending the broader implications of AI technologies, including their potential to automate tasks and enhance human capabilities, which addresses the economic implications of GenAI. The focus here is on the responses at societal and training levels needed to equip workers, and citizens in general, with the necessary skills.

Digital skills remain high on the EU policy agenda, and existing evidence supports a need for ongoing and sustained efforts. The Digital Decade Policy Programme has a target that at least 80% of persons aged 15–74 should have at least basic digital skills by 2030.[151] In 2023, 56% of that population group had basic or above basic digital skills, which is far from the target. The EU also has a target[152] that the share of low-achieving students in computer and information literacy should be less than 15% in 2030. Yet, in 2023,

149. https://www.oecd.org/en/publications/oecd-employment-outlook-2023_08785bba-en/full-report/skill-needs-and-policies-in-the-age-of-artificial-intelligence_fe530fbf.html

150. 3. Skills outlook - The Future of Jobs Report 2025 | World Economic Forum https://www.weforum.org/publications/the-future-of-jobs-report-2025/in-full/3-skills-outlook/#3-1-expected-disruptions-to-skills

151. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en

152. As agreed in the Council Resolution on a strategic framework for European cooperation in education and training towards the European Education Area and beyond (2021-2030). https://op.europa.eu/en/publication-detail/-/publication/b004d247-77d4-11eb-9ac9-01aa75ed71a1

43% of students did not reach the basic level of digital skills.[153] In 2023, the Council adopted a Recommendation on improving the provision of digital skills and competencies in education and training,[154] recognising the importance to promoting a quality inclusive and consistent approach to the development of digital skills at all levels of education and training.

The 2024–2029 European Commission continues to place a high emphasis on skills (including digital skills) through, for example, its overarching Union of Skills.[155] As part of this policy action, digital skills have been recognised as basic skills in the recently adopted Action Plan on Basic Skills. A forthcoming European Strategy for Vocational Education and Training (VET).[156] The Future of European Competitiveness Report[157] highlights that addressing skills gaps at all stages is crucial to Europe's long-term success and that there is a need for increased efforts in education, training and lifelong learning. It warns that while technology, including AI, is pivotal to preserving Europe's social model, it could exacerbate inequalities if not accompanied by strong investments in education and skills development. The EU Competitiveness compass[158] addresses this need for the domain of AI through the Apply AI Strategy, while the recently published AI Continent Action Plan underlines the need to reinforce AI skills, including basic AI literacy through further developing excellence in AI education, training and research.

- → Digital legislation including the AI Act,[159] the Cybersecurity Act[160] and the Digital Services Act[161] has implications for citizens in terms of awareness of their rights and responsibilities and understanding of ethical and social implications of digital technologies. Article 4 of the AI Act requires providers and deployers of AI systems to ensure a sufficient level of AI literacy of their staff and other persons dealing with AI systems on their behalf. To do so, providers and deployers should consider staff technical knowledge, experience, education and training and the context in which the AI systems are to be used, including the persons targeted by such AI systems.[162]

- → AI literacy is increasingly recognised as a critical component of workforce development and one of the increasing skills demanded by employers.[163] It involves not only the ability to use AI tools but also an understanding of AI's fundamental concepts, ethical considerations, and societal impacts. As AI systems become more integrated into workplace processes, developing AI literacy is essential for empowering workers to make informed decisions and use AI responsibly and effectively.

The importance of AI literacy needs to be treated not as a standalone skill but as embedded within the broader context of digital competence (see next paragraph on DigComp 3.0). By fostering a workforce that is knowledgeable about AI, organisations can enhance productivity and innovation while mitigating risks associated with AI adoption.

153. International Computer and Information Literacy Study, ICILS https://op.europa.eu/en/publication-detail/-/publication/59721dc6-a0aa-11ef-85f0-01aa75ed71a1/language-en

154. https://eur-lex.europa.eu/eli/C/2024/1030/oj/eng

155. https://commission.europa.eu/topics/eu-competitiveness/union-skills_en

156. https://commission.europa.eu/document/e6cd4328-673c-4e7a-8683-f63ffb2cf648_en

157. https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en

158. Competitiveness compass - European Commission

159. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

160. https://eur-lex.europa.eu/EN/legal-content/summary/the-eu-cybersecurity-act.html

161. https://eur-lex.europa.eu/legal-content/EN/TXT/?toc=OJ%3AL%3A2022%3A277%3ATOC&uri=uriserv%3AOJ.L_.2022.277.01.0001.01.ENG

162. See also Q&A on AI literacy: https://digital-strategy.ec.europa.eu/en/faqs/ai-literacy-questions-answers

163. The Future of Jobs Report 2025 | World Economic Forum https://www.weforum.org/publications/the-future-of-jobs-report-2025/digest/

DIGITAL COMPETENCE FRAMEWORK 3.0

The European Digital Competence Framework for Citizens, commonly known as DigComp,[164] aims to address the digital skills gap by providing a structured approach to enhancing digital competences among citizens. The forthcoming DigComp 3.0 iteration, due to be published by the end of 2025, will further integrate AI-related competences, reflecting the need for citizens to understand and interact with AI systems ethically and effectively. This framework is central in supporting the development of critical digital skills that are necessary for navigating and benefitting from the evolving digital landscape. A systematic and transversal integration of AI, including GenAI, will be included in DigComp 3.0 across all five competence areas of the framework with a focus on conceptual understanding, ethical, and societal implications. The emphasis is on encouraging a critical, reflective, and balanced use of AI systems and their outputs. This integration will align with other key initiatives such as the AI Act and the forthcoming AI Literacy Framework for primary and secondary schools co-developed by the European Commission and OECD (see Section 6.2).[165] Specifically, GenAI competences will be considered as follows:

- **Citizen cybersecurity:** There is an increased focus on citizen cybersecurity, including the role of AI systems in both cyberattacks and cybersecurity. This is reflected in three out of the 21 competencies: managing digital identity, protecting devices, and protecting personal data and privacy.

- **Rights, choice, and responsibility:** The competence engaging in digital citizenship will be expanded to include more explicit references to consumer rights, choice, active participation, and influencing. This includes awareness and assertion of rights in relation to recent digital legislation such as the Digital Services Act and AI Act.

- **Misinformation, disinformation, and threats to democracy:** The competence evaluating digital content focuses on dealing with misinformation and disinformation, recognising the role of GenAI and the potential speed at which digital information can spread. This includes identifying biased sources, fact-checking, flagging, and reporting misinformation and disinformation.

- **Twin transition:** The competence protecting the environment includes a consideration of how recent technologies and trends, such as GenAI and social media, are resource-intensive and the role of individual practices in mitigating this impact.

Overall, the updated DigComp 3.0 framework emphasises the importance of digital literacy, critical thinking, and responsible behaviour in the digital age, with a focus on the societal implications of emerging technologies, such as AI and GenAI.

THE ACADEMIC OFFER ON AI

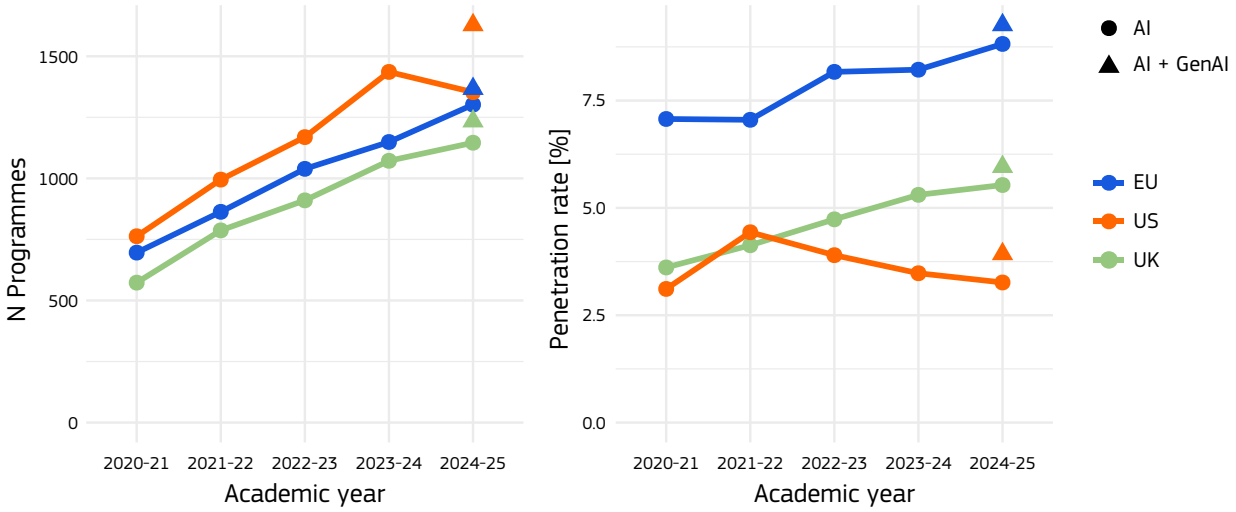
The EU has emphasised the importance of developing a digitally skilled population and workforce to remain competitive in the digital economy. As mentioned above, several policy initiatives, including the Digital Decade Policy Programme and the AI Continent Action plan, aim to strengthen AI capabilities in the workforce. The AI Continent Action plan proposes supporting an increase in EU bachelor's and master's degrees and PhD programmes in key technologies to enlarge the pool of AI specialists. Higher educational institutions play a critical role in shaping the future workforce, and their academic offerings provide valuable insight into

164. Digital Competence Framework for Citizens (DigComp) - European Commission https://joint-research-centre.ec.europa.eu/projects-and-activities/education-and-training/digital-transformation-education/digital-competence-framework-citizens-digcomp_en

165. Empowering learners for the age of AI: draft AI literacy framework launch - European Education Area https://education.ec.europa.eu/event/empowering-learners-for-the-age-of-ai-draft-ai-literacy-framework-launch

the skills that will define tomorrow's human capital. Monitoring trends in academic programmes is therefore essential for understanding how education systems are responding to emerging technological developments. Through the use of text mining, capturing the inclusion of advanced digital technologies in the programmes' syllabus, and drawing on data from StudyPortals, the JRC monitors the availability of English-taught master's programmes, bachelor's programmes, and short professional courses, and studies their characteristics.[166] [167] [168] The analysis below focuses on master's degrees, especially relevant for anticipating skills development as they represent the final academic stage before entering the workforce.

Figure 13. AI (and GenAI) related master's degrees by geographic area and academic year, 2020-25.



The penetration rate (right) is defined as the share of AI (and GenAI) master's degrees relative to the total number of masters offered in the geographic area.

Source: JRC elaboration.

Figure 13 shows the trends in the number of programmes by geographical area. For the period 2020–2024, the focus is on AI-related programmes. In 2024–25, the scope is expanded to include programmes explicitly addressing GenAI, as such data were not available in earlier years. The latest set of results confirms a consistent increase in the number of AI master's degrees offered in the EU and the UK, complemented with an additional number of degrees including GenAI. The US shows a marked decline in 2024–25, which is only offset by considering GenAI, suggesting a strategic shift towards replacing traditional AI programmes with GenAI offerings. Figure 13 (right) shows the penetration rates – the proportion of AI (and GenAI) master's degrees relative to the total number of master's offered in the geographic area –, reflecting the relative importance of AI across all academic offerings. This plot highlights the EU's strong and sustained leadership in AI adoption in academic offerings between 2020–21 and 2024–25. The EU consistently records significantly higher penetration rates

166. López Cobo M., et al., Academic offer and demand for advanced profiles in the EU. Artificial Intelligence, High Performance Computing and Cybersecurity, EUR 29629 EN, Publications Office of the European Union, Luxembourg, 2019.

167. Righi, R., et al. Academic offer of advanced digital skills in 2019-20. International comparison. Focus on Artificial Intelligence, High Performance Computing, Cybersecurity and Data Science, EUR 30351 EN, Publications Office of the European Union, Luxembourg, 2020.

168. https://joint-research-centre.ec.europa.eu/predict/academic-offer-advanced-digital-technologies_en

compared to the UK and the US. Interestingly, the inclusion of GenAI-related master's degrees leads to an increase of 0.4 percentage points in the penetration rate, compared to traditional AI programmes, highlighting the increased interest generated by this emerging field.

STRATEGIES FOR UPSKILLING AND TRAINING

→ **Workplace training programmes:** Organisations are increasingly implementing in-house training programmes focused on GenAI and digital skills. These programmes often include workshops, online courses, and hands-on projects that allow employees to gain practical experience with AI tools. Such initiatives are crucial for keeping the existing workforce competitive and capable of adapting to technological changes, and to understand remaining capacity building needs (see AI-empowered JRC box). In addition, the Commission has created an AI literacy repository collecting practices from organisations that provide and deploy AI systems[169] in order to support the implementation of Article 4 of the AI Act.

AI-empowered JRC
As European Commission's in-house science service, the JRC has launched an initiative to explore the potential of AI in the workplace. With 75% of colleagues using AI tools like GPT@JRC (in-house secure platform for staff to access and experiment with Large Language Models), the organisation aims at upskilling staff to harness the full potential of AI. The "AI-Empowered JRC" project aims to support staff in using AI tools effectively, addressing concerns around limitations and risks, and developing the skills needed to work with AI systems. As part of this initiative, the JRC is focusing on upskilling and training, particularly for managers, to ensure that AI is integrated into daily tasks in a way that enhances productivity and efficiency. As an example of specific application at the Science and Policy interface, experiments to deploy "AI Research Assistants" powered by GenAI and agentic AI are ongoing. The project's approach emphasises experimentation, collaboration, and mutual learning, and may offer useful insights and best practices for other organisations looking to adopt AI in their own workplaces. The purposeful use of Generative AI is given particular attention.

→ **Partnerships with educational institutions:** Collaborations between businesses, public sector organisations and educational institutions can facilitate the development of curricula that align with industry needs. This will also be important in the phase of highlighting what training needs businesses and organisations have. Higher educational institutions and technical schools can offer specialised courses and certifications in AI literacy, ensuring that graduates enter the workforce with the skills required to thrive in a GenAI-driven environment.

→ **Policy-backed initiatives:** Policymakers have a vital role in supporting the workforce through public initiatives aimed at enhancing AI literacy. Policy programmes can provide funding for training initiatives, create incentives for companies to invest in employee development, and establish national standards for digital literacy.

ROLE OF EDUCATIONAL INSTITUTIONS AND POLICYMAKERS

Educational institutions are at the forefront of bridging the skills gap. By integrating AI literacy into their curricula, education and training institutions at all levels can prepare students for a future where AI is ubiquitous. This includes not only teaching technical skills but also fostering critical thinking and problem-solving abilities, which are essential in an AI-driven world.

169. See https://digital-strategy.ec.europa.eu/en/library/living-repository-foster-learning-and-exchange-ai-literacy

Policymakers, on the other hand, must create an enabling environment that supports lifelong learning and continuous skill development. This involves crafting policies that encourage educational innovation, support digital infrastructure development, and promote equitable access to learning opportunities for all segments of the population.

ADDRESSING THE GAP

To address the skills gap, a multifaceted approach is necessary, involving formal education and training, and also upskilling and reskilling initiatives. Upskilling refers to enhancing existing skills to meet the demands of new technologies, while reskilling involves training workers in entirely new skills that are relevant to the current job market. A study conducted with 2,307 GenAI decision-makers show that two out of 3 respondents acknowledge that their employees do not have the skills to work with GenAI. About half are planning employee education and training to increase GenAI adoption.[170]

4.2 GenAI and Information Manipulation

KEY MESSAGES

- Generative AI models are revolutionising content creation by making it remarkably easy to produce highly convincing manipulative content rapidly and at scale. This capability can be exploited to dominate social media discussions, mimic authoritative news outlets, and create realistic images, videos, and audio. The potential for mis/disinformation is vast, as AI-generated content can be used to mislead the public, erode trust in media, and distort the overall information landscape. Such manipulation can have profound effects, including altering the outcomes of elections, influencing political processes, and shifting public support on critical issues like climate change. By flooding the digital space with misleading information, generative AI can significantly skew public perception and decision-making.
- The rapid generation of mis/disinformation often outpaces the ability to rebut it effectively, as crafting accurate and verifiable responses takes time. This discrepancy means that even well-intentioned use of AI for communication can struggle against the tide of falsehoods. To maintain a commitment to fact-based trustworthy communication to uphold the integrity of information is often an uphill battle in the face of AI-generated mis/disinformation.
- While technical solutions like watermarking AI-generated content and verifying information with trusted sources are valuable, they are not sufficient on their own. The most effective defence against mis/disinformation lies in promoting media and AI literacy among citizens, including policy makers. By equipping individuals with the skills to critically engage with AI-generated content, they can better discern false or inaccurate information. Educating the public to question and analyse the content they encounter empowers them to make informed decisions and resist manipulation. This foundational approach, combined with technical measures, is vital to maintaining a well-informed society resilient to the challenges posed by generative AI.

Various facets of mis/disinformation appear in many sections in this report, but the topic also deserves dedicated consideration when examining the implications of GenAI and how it can be used to create or to fight false or misleading content.

170. NTT Data (2025). Global GenAI Report How organizations are mastering their GenAI destiny in 2025. Global GenAI Report | NTT DATA https://services.global.ntt/en-us/campaigns/global-genai-report

Unintentional bias already introduces the risk of misleading content when AI is used to generate or even summarise content (see also Section 4.8). For instance, using generative AI to obtain information, answers can be heavily biased based on the dataset on which the AI was trained. When AI is used intentionally for manipulating information, the use of GenAI leads to unprecedented challenges by enabling the creation of deep fakes of both audio and visual content at large scale and speed, distorting media integrity.

Independently of the content creation, GenAI can also be used for systemic data poisoning at scale, for example, to pollute free and open knowledge repositories (see also Section 2.2) or to amplify the dissemination in online news and social media. As an example, bias in AI models can significantly impact attitudes towards climate change. A worldwide change of attitude towards climate change may have a compound effect on global emissions, amplifying or reducing them significantly. Recommender AI systems can spread GenAI-generated disinformation and sensationalist content, exacerbating polarisation on climate change [171] [172] [173] (see also Section 4.5).

GenAI-based influence operations directed at electoral processes have been monitored closely. Studies reveal a significant increase in artificially generated content with the rise of GenAI capabilities, unveiling Foreign Information Manipulation and Interference (FIMI) attempts.[174]

[175] [176] [177] One notable interference applying AI based impersonating techniques was the 'Doppelganger' campaign that consists of the use of fake clones of legitimate websites – both from media organisations and public institutions.[178]

Especially when it comes to social media, GenAI-powered bots boost the spreading of misleading narratives and manipulated information, frequently adding negative sentiment to disrupt public discourse, provoke targeted communities and foster radicalisation (see also Section 4.3).

Addressing issues caused by manipulated information requires AI literacy among policy-makers and citizens to ensure they can critically engage with AI-generated content and identify false or manipulative information, including deep fakes (see also Section 4.1). Furthermore, the Regulation on transparency and targeting of political advertising has a specific obligation to ensure transparency of the use of AI to target or deliver political advertisings, and a dedicated Code of Conduct on Disinformation has been adopted by the European Commission and European Board for Digital Services in the context of the DSA along with Guidelines on elections-related risks (see also Section 5.2). It represents an EU-wide effort to address the issues raised above.

At the same time, it should be underlined that GenAI can also help in combatting information manipulation and its dissemination (see also Section 6.4 for the deep dive on GenAI and cybersecurity). Opportunities include the use of GenAI for clear and targeted communication, fact checking, identifying AI-generated content, and more.

171. Meyerson, E. (2012). YouTube Now: Why We Focus on Watch Time. YouTube Official Blog.

172. Chaslot, G. (2017). How YouTube's A.I. boosts alternative facts. *Medium.*

173. Falkenberg, M., Galeazzi, A., Torricelli, M. et al. Growing polarization around climate change on social media. Nat. Clim. Chang. 12, 1114–1121 (2022). https://doi.org/10.1038/s41558-022-01527-x

174. https://edmo.eu/wp-content/uploads/2024/03/EDMO_TFEU2024-Narratives_Report-National_Elections-2nd-edition-1.pdf

175. https://www.eeas.europa.eu/eeas/1st-eeas-report-foreign-information-manipulation-and-interference-threats_en

176. https://www.eeas.europa.eu/eeas/2nd-eeas-report-foreign-information-manipulation-and-interference-threats_en

177. https://www.eeas.europa.eu/eeas/3rd-eeas-report-foreign-information-manipulation-and-interference-threats-0_en

178. https://euvsdisinfo.eu/uploads/2024/06/EEAS-TechnicalReport-DoppelgangerEE24_June2024.pdf

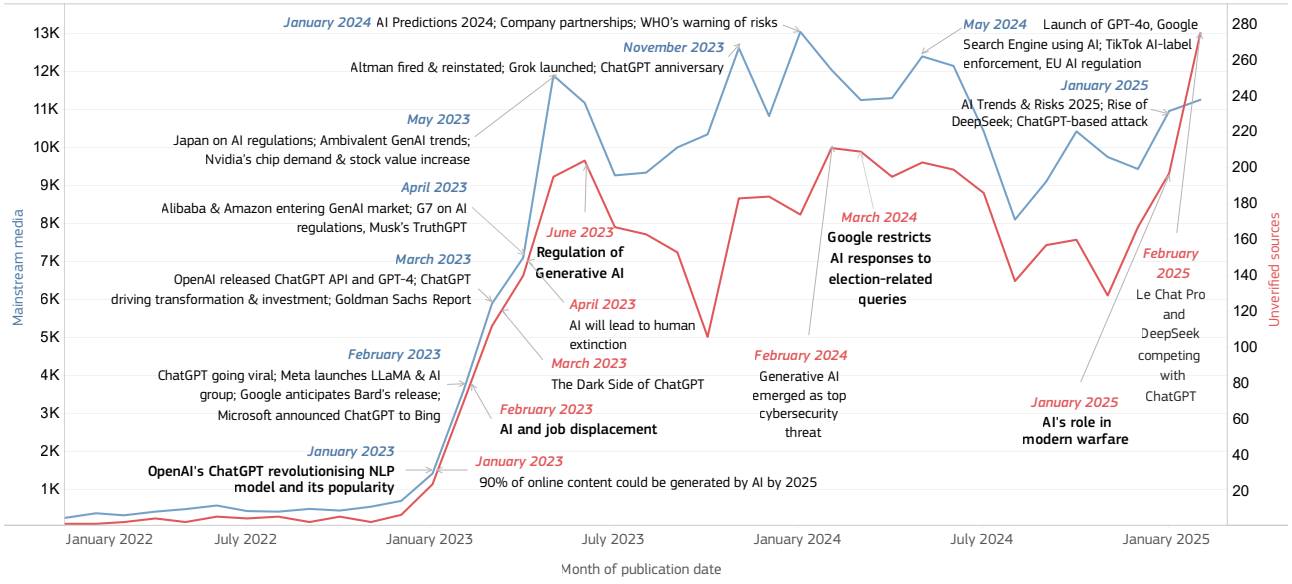
4.3 Generative AI portrayal in the Media

KEY MESSAGES

- The media portrayal of GenAI is often polarised, with utopian visions highlighting its transformative potential and dystopian concerns warning about ethical implications, job displacement, and privacy issues. This dual narrative shapes public discourse and might influence policy perceptions globally.
- Media coverage of GenAI has seen a significant increase since late 2022, particularly after the public release of ChatGPT. Subsequently, reporting peaks followed key events such as the release of new GenAI models, their integration into commercial products, and political regulatory actions from numerous stakeholders, including the active participation of the European Commission in the discourse.
- An analysis of news articles shows that the media convey the high relevance of international governance to manage GenAI and mitigate associated risks.

Understanding the media portrayal of GenAI is crucial to apprehend its perceived role and impact on society and the economy. The perception of GenAI in online media news articles is often framed within a spectrum of narratives, mostly portraying extreme scenarios. On the one hand, utopian visions highlight the transformative potential of GenAI, suggesting how it could revolutionise industries, enhance creativity, and solve complex global challenges, leading to unprecedented innovation and prosperity. On the other hand, dystopian concerns are widespread, with articles warning about the ethical implications, job displacement, and the erosion of privacy and security that could accompany AI's rise. Speculations about strong GenAI fear characterise the debate, with some envisioning a future where machines surpass human intelligence, raising existential questions about control, autonomy, and even potential leading to human extinction. Meanwhile, the take-up of GenAI as an integral part of modern everyday life is increasingly evident, as it becomes embedded in routine activities, prompting discussions about dependency and the reshaping of human capabilities.

Figure 14. Evolution of reporting volume on Generative AI in mainstream media and unverified sources.



Monthly timeline distribution of news articles on Generative AI with annotations. Source: EMM full index. Period: 01-01-2022 to 28-02-2025

Source: JRC elaboration.

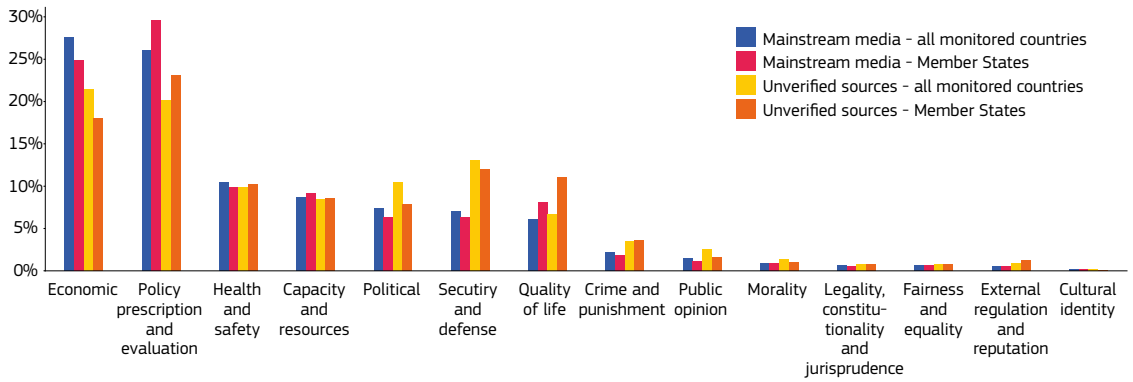
Figure 14 shows how both mainstream and unverified sources, i.e. sources that were indicated by independent fact-checkers as often spreading mis/disinformation, have covered GenAI, based on the reporting trends and peaks, as well as the narratives related to GenAI that influence policy and public perception across the globe.[179] The analysis shows a similar distribution in the reporting trends both for the mainstream media and unverified sources around main GenAI-related events. Nevertheless, unverified sources often distort information by misinterpreting real events and statements. The main reported topics related to GenAI were identified by assessing the key clusters that were computed for each month based on the article sentences' semantic similarity topics, covering technological and corporate developments, regulatory and ethical considerations, economic and industrial impact, global competition and collaboration, and cultural and societal impact. These themes persist in both examined datasets of articles from the mainstream media and from unverified sources. However, the articles in the latter, focus readers' attention on a different perspective within each theme.

FRAMING DIMENSIONS IN NEWS ARTICLES RELATED TO GENERATIVE AI

The examination of media coverage related to GenAI offers additional insights through the analysis of framings detected in the articles. These framing dimensions correspond to specific aspects mentioned in the context of the main topic, thereby revealing the perspective from which issues or news pieces are presented. Figure 15 compares framing dimensions detected in news articles from the mainstream media and unverified sources, colour-coded by source type, including *economic, policy prescription and evaluation, health and safety, capacity and resources, political issues, security and defence, crime and punishment, public opinion, morality, legislation, fairness and equality, regulation, and cultural identity.* The figure reveals differences between mainstream and unverified sources, as well as between articles coming from EU Member State sources and global reporting. We quantitatively analysed framing dimensions and their distribution between the mainstream media and unverified sources. Framing dimensions are assigned to over 75% of the articles from both mainstream and unverified sources across the observed period.

179. The analysis is based on 265 306 online news articles collected via the Europe Media Monitor from both mainstream media and unverified sources[1] between 1 January 2022 and 28 February 2025, selected by searching for a set of relevant keywords or keyword combinations within the text. These keywords are translated into all 24 EU languages, as well as Russian, Chinese (traditional and simplified), Arabic, Hindi, Turkish, Norwegian, Georgian, and Japanese. The keyword combinations[3] have been selected carefully to include articles of interest but to avoid irrelevant 'noise' in the news collection. The data-driven analysis used automated techniques including AI, machine learning and large language models (LLMs) to discover trends, patterns and shifts in reporting around GenAI, identify the main stories of mis/disinformation, compare specific trends, discover the associated sentiment and the most common framing dimensions.

Figure 15. Framings of GenAI news by media type and target.



Framing dimensions related to Generative AI news by media type and source country. Period: 01-01-2022 to 28-02-2025

Source: JRC elaboration.

Although *Economic* and *Policy prescription and evaluation* framings were prevalent in both mainstream and unverified sources, the mainstream media exhibited a higher share of these framings. The *Economic* framing stood out due to the substantial number of news articles highlighting the transformative impact of GenAI. This is not only revolutionising the tech sector but also reshaping various other industries, spurring major investment and intensifying competition. The *Policy prescription and evaluation* framing was particularly stressed in the EU mainstream media, addressing the need for and calls for policies to regulate the development and deployment of AI and GenAI technologies, ensuring ethical standards and mitigating potential risks. Both media types frequently used *Health and safety* as well as *Capacity and resources* framings. The news items provide instances on how technological advancements can have a positive impact on health care but also raise concerns about increased risks for the public such as job security. Moreover, the rapid development requires a significant amount of financial resources and capacities to run AI systems.

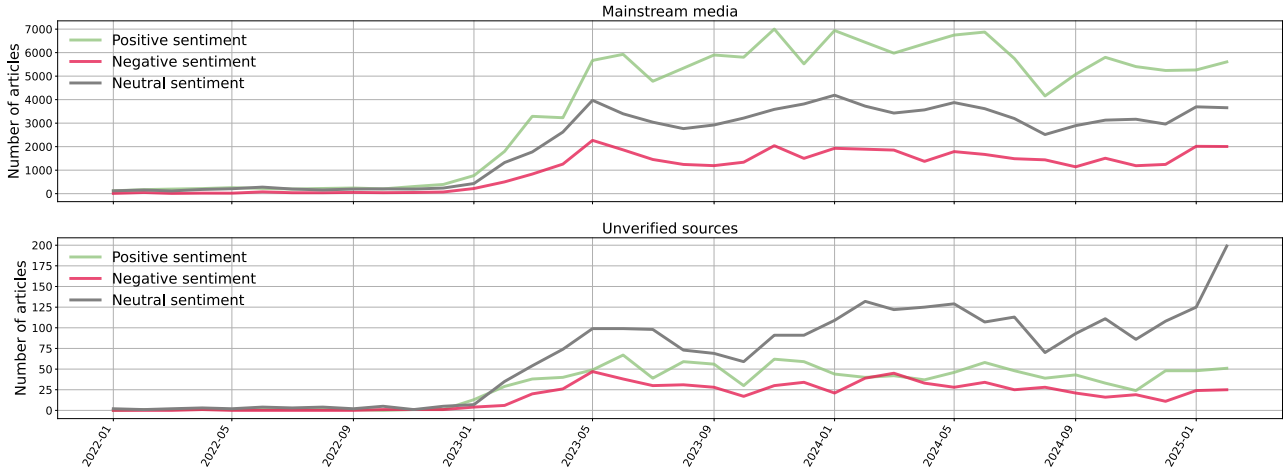
SENTIMENT ANALYSIS OF NEWS ARTICLES RELATED TO GENERATIVE AI

In addition, a pre-trained machine-learning model was applied to the news articles for sentiment analysis to identify the sentiment expressed in these articles.[180] The model assigns positive, neutral or negative sentiment to each article.

Figure 16 presents the evolution of news reporting by sentiment (positive, neutral, negative) divided between mainstream media and unverified sources.

180. Di Nuovo, E., Cartier, E., De Longueville, B. (2024). Meet XLM-RLnews-8: Not Just Another Sentiment Analysis Model. In Natural Language Processing and Information Systems, 28th International Conference on Applications of Natural Language to Information Systems, NLDB 2024, Turin, Italy, June 25–27, 2024, Proceedings (pp. 1). Springer Science and Business Media Deutschland GmbH.

Figure 16. Sentiment evolution related to news on GenAI.



The timeline distribution of GenAI-related articles divided by source type. Source: EMM full and Disinfo index. Period: 01-01-2022 to 28-02-2025.

Source: JRC elaboration.

The **predominant sentiment of the articles from the mainstream media related to GenAI is positive**, prevailing throughout the total observation period. The next prominent sentiment is neutral. Roughly 30% of the total news articles from the mainstream media are portrayed in a negative light, addressing the concerns about the risks that AI technologies may pose.

Most of the positive headlines refer to economic growth and associated opportunities for multiple industries, featuring the significant advancement

and major investments in start-ups focusing on key activities associated with progress, such as technology, healthcare and sustainability. Media also shed light positively on start-ups gaining awards and recognition at international competitions.

In contrast, the negatively connoted headlines refer to start-ups and tech companies facing significant challenges across various industries, including insolvencies, and a lack of investment in start-ups applying GenAI in various domains, lobbying for favourable regulations and more support from governments to address these issues. These negative headlines also capture the threats and risks of the new technologies connoted with a **negative** sentiment.

Sentiment analysis conducted on articles from **unverified sources** reveals the prevalence of the neutral sentiment, closely followed by the positive and the negative sentiments alike. Unverified sources were prone to spreading misleading narratives and disinformation often presenting GenAI with a sensationalist and alarmist sentiment. On the one hand, the headlines with positive sentiment exaggerate the capabilities of GenAI, portraying it as a near-magical technology able to solve all societal problems while, on the other hand, negative headlines precipitate a catastrophic dystopia. Such articles exploit fears of the unknown, suggesting that GenAI could imminently replace human decision-making or even replace humans themselves. Unverified sources often downplay the existing challenges and ethical considerations, oversimplifying complex debates to fit more dramatic narratives. By focusing on extreme scenarios and often ignoring expert opinions, they contribute to a polarised discourse, where the potential benefits of GenAI are either overstated or surpassed by exaggerated risks.

Overall, the analysis shows that the **media conveys the high relevance of international governance to manage Generative AI** and mitigate associated risks. This section is relevant for policymakers as it highlights the need for action. The results show that applying various text mining techniques on media articles related to this topic provides important insights into the discourse on GenAI in general and can highlight trends in the reporting tonality. This is of particular interest to media analysts, communication experts and policymakers.

The approach described can be used to **inform outgoing communication strategies**, based on the understanding of the way in which issues have been framed in different media. They can also be used to assess the impact of outgoing communication from the EU institutions, by observing whether there is an influence on the discourse in the media around specific issues related to GenAI.

4.4 Digital Commons

KEY MESSAGES

- The digital commons are a crucial element in AI training datasets, and their fate is closely tied to the development of AI technologies.
- As AI continues to evolve, it is essential to protect and support the digital commons to ensure the long-term health and diversity of both since protecting the digital commons is essential for fair and advanced AI.

The digital commons refer to online content and resources – such as code, software, images, texts, and other forms of knowledge – that are shared with free and open licences and are therefore accessible to anyone with an internet connection. They are the online equivalent of common public goods and spaces such as public parks or libraries, and, for example, include various online encyclopaedias and wikis, open source software repositories, digitised cultural heritage archives, openly licensed art and media collections, and online forums where people document and seek answers to questions. The digital commons

comprise a wealth of code, knowledge, and cultural content that is unique because it is diverse, collaboratively made, and free of charge. This has not escaped AI developers and, today, content belonging to the digital commons constitutes a crucial element in practically all AI training datasets. It is therefore not a question if the fate of the digital commons and AI technologies are closely intertwined, but rather *how and with what effects*. While the digital commons are generally considered non-rival – meaning their value and usability do not decrease due to overuse – they are known to be threatened by issues such as pollution, undersupply, and a lack of findability.[181] These threats are all reinforced by the introduction of GenAI technologies, opening up new areas of possibility and concern.

OPPORTUNITIES

On the positive side, GenAI may provide new abilities to search and navigate open databases, help fact-check and improve metadata in shared knowledge repositories, and assist in the translation of knowledge and information, thereby improving the findability, accessibility, and quality of content belonging to the digital commons. It is important to emphasise that such uses of AI are not new to the field of the digital commons, which has long relied on machine learning to manage and develop open datasets. Building on such practices, knowledge related to the digital commons could also feed back into and play an important role in improving current AI practices – especially regarding the collection and curation of high-quality AI training datasets. Based on consent and full transparency, projects such as Public Diffusion are, for example, re-envisioning current AI training practices by using content from the commons and the public domain to develop AI models that are fully based on principles of openness and approval,[182] showcasing a radical alternative to commercial AI training practices that are often based on closed, non-consensual datasets that may violate the rights of copyright holders. With decades of experience in curating text and image databases to address issues concerning bias, discrimination, and cultural gaps, experts from fields relating to the digital commons (including archives, museums, and libraries) could also play an important role in making AI training datasets, models, and systems fairer and more diverse, collaborative, and inclusive.[183]

RISKS AND CHALLENGES

The use of GenAI may have negative effects on the digital commons, including:

→ Enclosure and privatisation of free and open knowledge: Restrictive measures to prevent data scraping may inadvertently limit access to online data for public good archives.[184]

→ Decreased voluntary contributions: As people rely on closed chatbot services, they may be less likely to contribute to shared knowledge databases.[185]

→ Pollution of free and open knowledge repositories: AI-generated data with errors, hallucinations, and disinformation may enter databases like Wikipedia,[186] requiring costly fact-checking and correction.

→ Financial strain on digital commons organisations: Providing open data

181. Dulong De Rosnay, Mélanie, and Felix Stalder. "Digital Commons." Internet Policy Review 9, no. 4 (December 17, 2020). https://doi.org/10.14763/2020.4.1530.

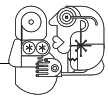
182. See Public Diffusion beta, https://source.plus/public-diffusion-private-beta

183. Bunz, Mercedes. "The Role of Culture in the Intelligence of AI." In AI in Museums, edited by Sonja Thiel and Johannes C. Bernhardt. transcript Verlag, 2023

184. Longpre, Shayne, and et.al. "Consent in Crisis: The Rapid Decline of the AI Data Commons," ArXiv, 2024. https://doi.org/10.48550/arXiv.2407.14933

185. Del Rio-Chanona, R Maria, Nadzeya Laurentsyeva, and Johannes Wachs. "Large Language Models Reduce Public Knowledge Sharing on Online Q&A Platforms." PNAS Nexus 3, no. 9 (September 2, 2024): 1-12. https://doi.org/10.1093/pnasnexus/pgae400.

186. Brooks, Creston, Samuel Eggert, and Denis Peskoff. "The Rise of AI-Generated Content in Wikipedia." arXiv, October 10, 2024. https://doi.org/10.48550/arXiv.2410.08044.



to AI web crawlers incurs significant infrastructural costs (more than half of website traffic is due to automated crawlers compared to human users), with little return, and may lead to financial burdens on organisations hosting open-source content.[187]

These risks may ultimately restrict access to and expansion of shared knowledge and place a significant economic burden on organisations that host content belonging to the digital commons.

FUTURE PERSPECTIVES

The future of the digital commons is uncertain and depends on the level of support it receives. There are two possible scenarios:

Best-case scenario:

→ The digital commons continue to thrive with financial and infrastructural support from public institutions and commercial AI developers.

→ The digital commons remain a diverse, high-quality source of information, promoting culturally balanced AI models and global access to knowledge, information, and cultural history.

→ Generative AI technologies provide creative support to communities around the world and thereby facilitate the production of new content that will make the commons grow. They are also successfully adopted by organizations hosting content belonging to the commons, making information management more efficient and accessible.

Worst-case scenario:

→ The digital commons deteriorate due to pollution and lack of new data, becoming an untrustworthy source of culture and information.

→ They wither away, reducing the diversity of governance models for creating and managing knowledge.

→ This would threaten global access to free and open culture, history, and knowledge, and damage AI development by relying on low-quality or outdated data.

While the GenAI has a potential to enrich cultural practices and support free and open knowledge and information repositories, there is a need to reflect on the potential negative consequences it may bring. Importantly, protecting the digital commons in the age of GenAI is not just a matter of securing open, diverse, and decentralised access to high-quality information for citizens across the globe, but a matter of securing the development of AI technologies that are fairer, more diverse, and therefore also more advanced and useful.

4.5 Environmental Implications of Generative AI

KEY MESSAGES

- The direct impact of AI on the environment is considerable and this can pose problems in particular if data centres are deployed in water poor regions, and recycling of components is an issue.
- Specific measures such as mandatory tracking of energy consumption, better planning of data centre in water - rich areas, and recycling programmes could help reduce the impact of AI on the environment.
- The EU has been pioneering efforts against climate change and integrating environmental protection into policies. Notably, the EU introduced multiple measures to address the impact of data

187. https://arstechnica.com/ai/2025/03/devs-say-ai-crawlers-dominate-traffic-forcing-blocks-on-entire-countries/

centres, including the Energy Efficiency Directive and its sustainability rating as well as the Energy Labelling and Ecodesign Directives, and is preparing an upcoming Cloud and AI Development Act focusing on sustainable data centres as well. The EU AI Act includes environmental provisions on AI models and systems transparency, adhesion to codes of practice and information disclosure.

- Emerging technologies in AI, such as energy-efficient transistors, neuromorphic chips, and specialised edge AI chips, are enabling on-device intelligence and reducing energy consumption, but challenges must be addressed to fully realise their potential for sustainable, real-time applications.

DIRECT ENVIRONMENTAL IMPACT OF AI

The increasing use of AI models, particularly GenAI and LLMs, has significant environmental implications and this is expected to grow. These models require massive computational resources, resulting in high energy consumption, water usage, and mineral extraction. The estimated 5,000 data centres in the US and 2000 in Europe[188] are expected to increase their power demand by 50% by 2027 and between 103% and 165% by 2030 according to different estimates.[189] [190]

Based on conservative estimates, data centres' energy demand increased from 194 TWh in 2010 to 204 TWh in 2018.[191] The global electricity consumption of data centres increased to 460 TWh in 2022. According to recent JRC estimates Data centres in the EU used an estimated 45–65 TWh of electricity in 2022 (1.8–2.6% of total EU electricity use), while telecommunication networks used an estimated 25–30 TWh of electricity (1–1.2% of total EU electricity use)[192] IEA estimated that globally, data centres consumed around 1.5% of electricity consumption in 2024.[193]

While it is challenging to identify the exact share of energy consumption attributable to AI, it is estimated that it currently accounts for 14% of data centre energy consumption, projected to rise to 27% by 2027.[194]

The environmental impacts of AI infrastructure include:

→ **Energy consumption:** High electricity demand for training and running models contribute to greenhouse gas (GHG) emissions, especially if the energy mix has a low share of renewable sources.[195] For example, different estimates rate that global energy demand for AI infrastructure will reach between 1% and 1.5% of total global energy consumption by 2027.[196] This has led AI companies to invest in nuclear energy and gas power generation.[197]

188. Statista. Leading countries by number of data centers as of March 2024. Blog post: https://www.statista.com/statistics/1228433/data-centers-worldwide-by-country/

189. IEA (2025), Energy and AI, IEA, Paris https://www.iea.org/reports/energy-and-ai

190. Goldman Sachs: AI to drive 165% increase in data center power demand by 2030

191. Masanet, Eric, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey. (2020). "Recalibrating Global Data Center Energy-Use Estimates." *Science* 367, no. 6481: 984–986.

192. Kamiya, G. and Bertoldi, P., Energy Consumption in Data Centres and Broadband Communication Networks in the EU, Publications Office of the European Union, Luxembourg, 2024, doi:10.2760/706491, JRC135926.

193. IEA (2025), Energy and AI, IEA, Paris https://www.iea.org/reports/energy-and-ai, Licence: CC BY 4.0

194. Goldman Sachs: AI to drive 165% increase in data center power demand by 2030 https://www.goldmansachs.com/insights/articles/ai-to-drive-165-increase-in-data-center-power-demand-by-2030

195. Ramachandran, K., Stewart, D., Hardin, K., & Crossan, G. (2024). As generative AI asks for more power, data centers seek more reliable, cleaner energy solutions. Deloitte Center for Technology Media & Telecommunications.

196. Morgan Stanley (2024). Powering the AI Revolution. Blog post: https://www.morganstanley.com/ideas/ai-energy-demand-infrastructure

197. CNBC Why Big Tech is turning to nuclear to power its energy-intensive AI ambitions https://www.cnbc.com/2024/10/15/big-tech-turns-to-nuclear-energy-to-fuel-power-intensive-ai-ambitions.html

It should be noted that there is high uncertainty surrounding the numbers provided above, for several reasons. On the one hand, these estimates may not be taking into account the possible widespread adoption of reasoning models such as OpenAI's o1, which yield better capabilities at the expense of an increased demand in computational resources. On the other hand, these estimates were made before the release of DeepSeek-R1, which showed that it was possible to compete with OpenAI for a fraction of the energy consumption. At the same time, the observed pace of energy consumption growth can also be under-reflective of the actual demand, given limitations and bottlenecks, for instance on the availability of AI chips, multi-year-long lead times for equipment, and power availability constraints.[198]

Obligations to release the full consumption of each data centre used for AI could help evaluate and reduce the total footprint.[199] [200]

→ **Water footprint:** Projected to account for 4.2-6.6 billion cubic meters of water withdrawal in 2027, this exceeds the total annual water withdrawal of half of the United Kingdom.[201] This is a significant concern for building data centres in countries with water shortages. Adding to the challenge, other available cooling mechanisms are less efficient, producing waste heat and potentially utilising harmful cooling agents like fluorinated-gases.

→ **Mineral resources:** Extraction of raw materials like silicon, aluminium, copper, tin, tantalum, lithium, gallium, germanium, palladium, cobalt, and tungsten for manufacturing chips has significant environmental costs.[202] Hundreds of tonnes of ore are required to be excavated and processed for just one ton of material.[203]

→ **E-waste:** Data centre hardware has a short lifespan of around 3.5 years,[204] resulting in a potential accumulation of 1.2-5.0 million tons of e-waste during 2020-2030.[205] Implementing a circular economy could increase the longevity of data centre hardware and reduce e-waste by 16-86% according to the same sources.

However, AI models can also help mitigate climate change by supporting applications like pollution tracking, weather monitoring, and energy optimisation.[206] The impact of these applications is still to be quantified and their potential may be hindered by a number of factors, such as interoperability concerns, critical shortage of skills, and limitations in the digital infrastructure. It is worth noticing that data centres tend to be highly concentrated in spatial terms and, given their substantial power and water draw, this poses significant challenges to local reservoirs and to the power transmission

198. Bashir, Noman, Priya Donti, James Cuff, Sydney Sroka, Marija Ilic, Vivienne Sze, Christina Delimitrou, and Elsa Olivetti. (2024). The Climate and Sustainability Implications of Generative AI. *An MIT Exploration of Generative AI*, March. https://doi.org/10.21428/e4baedd9.9070dfe7.

199. Crawford, K. (2024). Generative AI's Environmental Costs Are Soaring–and Mostly Secret. Nature, 20 February, 2024.

200. Luccioni, A. S., Strubell, E., & Crawford, K. (2025). From Efficiency Gains to Rebound Effects: The Problem of Jevons' Paradox in AI's Polarized Environmental Debate. arXiv preprint arXiv:2501.16548.

201. Pengfei Li et al, Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models https://arxiv.org/pdf/2304.03271

202. Brito, Griffin and Koski (2022), "Nvidia GPU — Design Life-Cycle" https://www.designlife-cycle.com/nvidia-gpu

203. Mills (2020), " Mines, Minerals, and "Green" Energy: A Reality Check" https://manhattan.institute/article/mines-minerals-and-green-energy-a-reality-check

204. Veau et al, Navigating E-waste for datacenters https://sustainability-ai.de/static/6d0446a876b6ccb79b58bbaeffcf8b62/Project-Report.pdf

205. Wang, P., Zhang, LY., Tzachor, A. et al. (2024). E-waste challenges of generative artificial intelligence. Nat Comput Sci 4, 818–823.

206. Vinuesa, R., Azizpour, H., Leite, I. et al. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun 11, 233. https://doi.org/10.1038/s41467-019-14108-y

grid. In Ireland, for example, data centres consume around 20% of the metered electricity supply. Localised environmental impacts of AI are likely to affect arable regions, exacerbating environmental inequalities.[207]

VOLUNTARY PROGRAMS AND REGULATION FOR DATA CENTRES

The negative environmental consequences of data centres' energy consumption and environmental footprint have been successful addressed by voluntary programs, such as the EU Code of Conduct for Energy Efficiency in Data Centres.[208] Through the commitment on the adoption of agreed and yearly updated best practices more than 400 data centre operators realised significant efficiency gains. The data collected within the program shows that the average power usage effectiveness (PUE) of participating companies fell from 1.8 to less than 1.3 over the last 15 years. Other similar initiatives demonstrate that the industry is actively working on the mitigation of their environmental sustainability with ambitious goals of being carbon neutral or water-positive. The AI community is also calling for more informed and sustainable use of AI.[209]

Regulators have also recently included sustainability of data centres among their priorities. In March 2024 the European Commission adopted the Delegated Act on a common rating scheme for data centres in the European Union. The Delegated Act implements the recast Energy Efficiency Directive (EED) and details the energy key performance indicators (KPI) that large data centre operators (with a power demand of the installed information technology of at least 500kW) must submit on regular basis to the European database on data centres.

The EU Taxonomy Regulation[210] and, in particular, its Climate Delegated Act, also deal with data centres: Section 8.1 of the EU Taxonomy Climate Delegated Act indeed addresses the economic activity "Data processing, hosting and related activities" and sets out the technical screening criteria to identify whether a data centre contributes substantially to, respectively, climate change mitigation and to climate change adaptation, while doing no significant harm to other environmental objectives.[211]

EMERGING TECHNOLOGIES AND BREAKTHROUGH INNOVATIONS

Emerging technologies like energy-efficient transistors made from materials such as molybdenum disulfide promise to allow AI processing in small devices, such as smartwatches, enhancing on-device intelligence while reducing reliance on cloud computing.[212] Similarly, but on a different scale, nanomagnetic computing offers ultra-low-energy computation by leveraging magnetic properties at the nanoscale, which could dramatically reduce energy consumption across AI-powered devices.[213]

The development of spike neural networks is another promising innovation in the pursuit of energy-efficient AI. By mimicking the brain's

207. Rein S. and Wierman A. (2024), The Uneven Distribution of AI's Environmental Impacts Harvard Business Review https://hbr.org/2024/07/the-uneven-distribution-of-ais-environmental-impacts

208. https://e3p.jrc.ec.europa.eu/en/groups/data-centres-code-conduct

209. See tools to compare the energy consumption of AI models, e.g. https://huggingface.co/blog/sasha/announcing-ai-energy-score

210. Regulation (EU) 2020/852 of the European Parliament and of the Council of 18 June 2020 on the establishment of a framework to facilitate sustainable investment, and amending Regulation (EU) 2019/2088 (OJ L 198 22.06.2020)

211. Bertoldi P., Assessment Framework for Data Centres in the Context of Activity 8.1 in the Taxonomy Climate Delegated Act, European Commission, Ispra, 2023, JRC131733

212. Hsu, Jeremy. "Energy-Efficient Transistor Could Allow Smartwatches to Use AI." New Scientist, 12 Oct. 2023, www.newscientist.com/article/2397235-energy-efficient-transistor-could-allow-smartwatches-to-use-ai/. Accessed 7 May 2025.

213. Imperial College London. "'Nanomagnetic' Computing Can Provide Low-Energy AI." ScienceDaily, 5 May 2022, www.sciencedaily.com/releases/2022/05/220505114646.htm

signalling process, SNNs provide event-driven processing, drastically reducing idle energy consumption. These advancements are crucial for edge computing, IoT devices, and mobile AI, where energy constraints are critical.[214] [215] [216] Furthermore, neuromorphic chips and optoelectronic neurons are optimising energy efficiency by emulating brain-like processing capabilities, which is particularly beneficial for applications in autonomous systems, healthcare monitoring, and wearables.

New specialised chips for edge AI are demonstrating enhanced inference speed and energy efficiency on local devices, minimising the need for data centre dependency. These chips, using technologies such as Binary Neural Networks (BNNs) and memristive crossbar arrays, allow for AI computations directly on devices, reducing latency and bandwidth requirements in smart cities, wearables, and industrial IoT applications.[217] [218]

As the use of GenAI increases, semiconductor innovations such as those exemplified will be key to enabling its sustainable and long-term development and uptake. Policy makers must therefore consider measures to foster RDI, as well as industrial scalability, hardware standardisation and supply chain security, ensuring a true twin transition where energy-efficient technologies meet growing adoption and performance demand.

ENVIRONMENTAL IMPACT OF AI VIA ITS SOCIETAL IMPACT

As mentioned in Section 4.2, biased AI models can influence attitudes, also on climate change. GenAI can also be a source of critical disinformation on climate, with models potentially biased due to hallucinations, training data issues, or reinforcement learning phase biases.[219] [220] [221]

As described in Section 5.1, the AI Act includes environmental provisions on AI models and systems transparency, adhesion to codes of practice and information disclosure. The environmental requirements extend to any provider, including those based outside of Europe placing GenAI on the European market. While this constraint is a natural consequence of European legislation, it limits the capability of addressing environmental considerations at a global scale. Some US-based providers of AI have threatened not to serve Europe if the regulation is against their interest.[222] [223] In parallel, models developed in China are becoming competitive, but they may also fall short on compliance with the European AI rules that prevent the spread of biased information about climate or address direct environmental impacts of models.[224] [225]

214. Ward-Foxton, Sally. (4 Dec. 2023). "What Is Holding Back Neuromorphic Computing?" EE Times. www.eetimes.com/what-is-holding-back-neuromorphic-computing/.

215. Zhu, Rui-Jie, et al. (27 Feb. 2023). SpikeGPT: Generative Pre-Trained Language Model with Spiking Neural Networks. https://doi.org/10.48550/arxiv.2302.13939.

216. Cerf, Emily. 7 Mar. (2023). "SpikeGPT: Researcher Releases Code for Largest-Ever Spiking Neural Network for Language Generation." News. news.ucsc.edu/2023/03/eshraghian-spikegpt/. Accessed 7 May 2025.

217. Tang, Baoshan, et al. (2022). "Wafer-scale solution-processed 2D material analog resistive memory array for memory-based computing." Nature Communications, vol. 13, no. 1. Article number: 3037. https://www.nature.com/articles/s41467-022-30519-w.

218. Chen, Yu-Hsin, et al. (Jan. 2017). "Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Networks." IEEE Journal of Solid-State Circuits, vol. 52, no. 1, pp. 127–138. IEEE Xplore, https://ieeexplore.ieee.org/document/7738524

219. Gartner website: Experts Answer the Top Generative AI Questions for Your Enterprise https://www.gartner.com/en/topics/generative-ai

220. Turing.com: Top Generative AI Industry Applications: An In-Depth Look https://www.turing.com/resources/generative-ai-applications

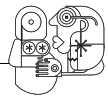
221. CNN Analysis: DeepSeek's AI is giving the world a window into Chinese censorship and information control | CNN

222. DW: Europe's AI bosses sound warning on soaring compliance costs https://www.dw.com/en/europes-ai-bosses-sound-warning-on-soaring-compliance-costs/a-70243489

223. Brookings: Are tariffs Big Tech's new tool against EU regulation? https://www.brookings.edu/articles/are-tariffs-big-techs-new-tool-against-eu-regulation/

224. Euronews: 'Harmful and toxic output': DeepSeek has 'major security and safety gaps,' study warns https://www.euronews.com/next/2025/01/31/harmful-and-toxic-output-deepseek-has-major-security-and-safety-gaps-study-warns

225. Wired: Here's How DeepSeek Censorship Actually Works—and How to Get Around It https://www.wired.com/story/deepseek-censorship/



To address these challenges, the EU has been pioneering efforts against climate change and integrating environmental protection into policies like the EU Charter of Fundamental Rights [Article 37]. The EU is an ideal region to develop AI technology with strong environmental guarantees.

Climate action in the age of AI requires:

- → **Energy-efficient computation:** Advancing innovation in semiconductors and promoting edge computing to reduce overall reliance on energy-intensive data centres (see examples in previous sub-section).
- → **Renewable energy for data centres:** Reducing GHG emissions and promoting water-rich regions for data centre placement.
- → **Efficient data centres:** Minimising energy and heat consumption as well as e-waste, while considering potential rebound effects.
- → **Transparent and trustworthy AI models:** Encouraging citizens to choose models that provide reliable information on climate change.
- → **International cooperation:** Establishing global standards to monitor and minimise AI's environmental impacts, considering the indirect effects of policies from main AI powerhouses like the USA and China.
- → **Sustainable material use in semiconductors:** Ensuring responsible sourcing of raw materials and fostering innovation in semiconductor design to reduce material demand and enhance circularity through reuse and recycling.

Promoting **Green, EU-centric AI models and systems** and setting up international standards can help ensure that AI development aligns with environmental protection and climate action goals. Having state-of-the-art AIs developed in the EU as "trustworthy by design" or "compliant by design" is a way to ensure that they comply with EU regulations. Potential strategies on AI and climate in the long term include the promotion of Green, EU-centric AI models and systems and the setup of international standards to monitor and minimise the environmental impacts of AI.

4.6 Generative AI and Children's Rights

KEY MESSAGES

- GenAI presents both opportunities and challenges for children's rights. While it has the potential to enhance education, creativity, communication, and information access, it also poses significant risks, such as deceptive manipulation, AI-based harmful content, and lack of privacy and safety.
- To ensure that GenAI is developed and used in a way that respects children's rights and promotes their well-being, it is essential to design trustworthy GenAI systems, prioritise transparency and accountability, and invest in education and training programmes that promote critical thinking and media literacy skills.
- Child safeguards need to be embedded in GenAI design and must go beyond general privacy and ethics frameworks in order to address specific vulnerabilities of children. In this context, a rights-based, age-appropriate, and inclusive approach is essential to ensure safe, empowering, and equitable AI experiences for all children.

Children's rights are a fundamental aspect of ensuring the well-being and development of young individuals, and recent advancements in GenAI have significant implications for this area, with children as disproportionately early adopters of technology. This Section reflects on the opportunities and challenges presented by GenAI in relation to children's rights, with a focus on some relevant aspects, including education, creativity

and mental health. Additional relevant input is also reported in Sections 5.2 and 6.2.

OPPORTUNITIES

GenAI presents several opportunities for enhancing the lives of children. One key area is education, where personalised and adaptive learning experiences can be tailored to the needs and specificities of individual learners as mentioned in Section 6.2. This can be achieved by using AI-based systems that provide real-time feedback and support, allowing teachers to shift their role to that of scaffolders, as introduced by Vygotsky's concept of a zone of proximal development (ZPD) that represents the gap between what a learner masters and is capable of doing with support from a teacher or a peer with more knowledge or expertise.[226] For instance, AI-powered adaptive learning systems can adjust the difficulty level of educational content based on a child's performance and interests, providing a more effective and engaging learning experience.

GenAI also has the potential to support creativity in children, enabling them to develop new forms of expression using new tools. For example, children can use GenAI to create video games, animations, or other forms of digital content based on their ideas. This can facilitate access to technology and promote the development of creative and problem-solving skills, which are essential for success in the digital age. Moreover, GenAI can support communication by allowing children to efficiently document and sustain their communication, ensuring that their voices are heard. This can be particularly beneficial for children with disabilities or language barriers, who may face challenges in expressing themselves.

In healthcare, generative AI tools can support early detection of health and developmental issue or provide insights into medical data.[227]

Finally, GenAI can support information access by translating complex content into age-adapted text, images, and other formats, making it more accessible to children. This can also be done in non-dominant languages, promoting inclusivity and diversity.

BARRIERS AND CHALLENGES

Despite the opportunities presented by GenAI, there are several barriers and challenges that need to be addressed. One significant risk is deceptive manipulation which is a form of influence where someone is led to believe something false or is persuaded to act in a way that is not in their best interest. Such deception or misleading tactics are prohibited under the AI Act. Because children's cognitive capacities are still in development, children are especially susceptible to the dangers of misinformation and disinformation. Additionally, AI-based harmful content, such as child sexual abuse material, poses a significant threat to children's safety and well-being. Another source of risk is the use of GenAI in the form of AI companions, which can be particularly detrimental for children, with risks of impersonation or deception in addition to tampering with children's emotional development (see also Section 4.7) and acquisition of interpersonal skills. The Article 28 of the DSA emphasises the need for minors' protection, highlighting the importance for providers of online platforms of ensuring high levels of privacy, safety and security for children (see Section 5.2) particularly regarding non-consensual image generation, which can have severe consequences for children's mental and physical health.

The risks of biases in training datasets, constructed and analysed by dominant actors, can also lead to unfair discrimination and to a loss of diversity and richness in GenAI outcomes. This can further result in a form of globalisation of GenAI outcomes, where the perspectives, experiences and needs of marginalised groups are excluded. Furthermore, like many emerging technologies, GenAI can worsen existing inequalities, particularly for

226. Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.

227. Generative AI: Risks and opportunities for children | Innocenti Global Office of Research and Foresight https://www.unicef.org/innocenti/generative-ai-risks-and-opportunities-children

children from marginalized communities who may face greater risks and have less access to its benefits. To be effective, generative AI must be inclusive, equitable, and responsible, catering to the diverse needs of all children.

The risks of hallucinations and the increased need for critical thinking skills are also significant concerns. With multiplied capacities compared to search engines and social media, GenAI has even more potential to shape users' perception of the world, social interactions and experiences. As GenAI becomes more prevalent, it is essential to develop and implement effective educational programmes that promote critical thinking and media and GenAI literacy skills (see Section 6.2), enabling children to analyse and evaluate the information they encounter online. Moreover, the potential risks of social discrimination, and negative impacts on children's mental and physical health must be carefully considered and addressed. Starting in this direction, an umbrella review and an expert report on the impacts of social media use (some embedding GenAI tools such as Chatbots) on adolescents' mental health and well-being flag that platforms' adoption of responsible design principles could improve risk mitigation.[228] [229]

WAYS FORWARD

To ensure that GenAI is developed and used in a way that respects children's rights and promotes their well-being, it is essential to design trustworthy GenAI systems. An example of this is the AI system for collaborative storytelling,[230] which is in line with the EU ethics guidelines for trustworthy AI.[231] The adaptation of ethical guidelines has identified relevant aspects for this population, such as stakeholder involvement, risk management, diversity and inclusion, children's rights and capacities, role of parents or carers, and the implementation of age-appropriate behaviours.

Transparency is a core transversal requirement for GenAI systems, particularly when it comes to children. Children-centric GenAI transparency requires informing children about the system's nature using age-appropriate language during child-AI interaction. These measures need to be adapted to age, linguistic, and cultural context, ensuring that children from diverse backgrounds can understand and engage with GenAI systems.

Evaluating the impact of GenAI on children's mental development is essential for better anticipatory governance response, and longitudinal studies can provide valuable insights into the effects of GenAI on children's cognitive processes and brain development.

4.7 Generative AI and Mental Health

KEY MESSAGES

- The use of AI chatbots and companion apps can lead to various mental health issues. These include addiction-like behaviours, validation-seeking tendencies, and in some cases,

228. Arianna Sala, Lorenzo Porcaro, Emilia Gómez, Social Media Use and adolescents' mental health and well-being: An umbrella review, Computers in Human Behavior Reports, Volume 14, 2024, 100404, ISSN 2451-9588, https://doi.org/10.1016/j.chbr.2024.100404.

229. Beullens, K., Bozzola, E., Cataldo, I., Hale, L., Kent, M., Montag, C., Nivins, S., O'reilly, M., Rubæk, L., Schiøtz Thorud, H.-M., Sterpenich, V. And Vandenbosch, L., Minors' health and social media: an interdisciplinary scientific perspective, Manolios, S., Sala, A., Sundorph, E., Chaudron, S. And Gomez, E. editor(s), Publications Office of the European Union, Luxembourg, 2025, https://data.europa.eu/doi/10.2760/3795891, JRC141090.

230. Escobar-Planas, M. et al. (2025). Implementing and Evaluating Trustworthy Conversational Agents for Children. In: Plácido da Silva, H., Cipresso, P. (eds) Computer-Human Interaction Research and Applications. CHIRA 2024. Communications in Computer and Information Science, vol 2370. Springer, Cham. https://doi.org/10.1007/978-3-031-82633-7_29

231. European Commission: Directorate-General for Communications Networks, Content and Technology and Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, Ethics guidelines for trustworthy AI, Publications Office, 2019, https://data.europa.eu/doi/10.2759/346720

encouragement of harmful actions such as self-harm and disordered eating.

- Additionally, the potential for digital impersonation adds another layer of emotional risk, potentially causing distress to individuals affected by such practices.
- The rise of deep fakes and manipulated media can pose significant risks, also in the form of non-consensual explicit content and emerging forms of online harassment.

GenAI is reshaping how individuals interact with technology, introducing new risks concerning mental health. In the coming years, we can expect GenAI to become even more integrated into digital services, and GenAI content to become more pervasive online. This may be accompanied by the emergence of new types of risks.

Recent research has highlighted the potential risks that AI chatbots pose to a person's mental and physical well-being.[232] Specifically, some GenAI characteristics such as the chatbot's perceived sentience, its anthropomorphism or human-likeness, or its surprising ability to tell users what they want to hear (i.e. "sycophancy") may contribute to problematic or addiction-like use of GenAI systems,[233] with emerging research aiming to understand the complex phenomena at play.

Recent reports have highlighted cases where AI companion apps encouraged users to engage in self-harm, eating disorder behaviour and violence.[234] Since then, new incidents of self-harm involving teenage users of GenAI have been widely reported, as well as cases involving the digital impersonation of real people, including deceased minors.[235]

Moreover, the rise of deep fakes enabled by GenAI, where it becomes difficult to distinguish between real and fake content, poses significant risks to mental health, increasing the risks of cyberbullying. This can take the form of creation and diffusion of non-consensual explicit content, especially among adolescents who are more prone to be vulnerable during this developmental stage. Deep fakes introduce a new dimension of harm through the creation of convincing falsified videos that not only damage reputation but also cause psychological trauma.[236] [237] The risk is not new, considering that the first photorealistic videos appeared on online platforms as early as 2017,[238] but the capacities of GenAI have exacerbated it tremendously. A recent study carried out by Oxford Research Institute shows that deep fake generators have been downloaded almost 15 million times since late 2022, and 96% of the deep fake models primarily targeted women.[239] A case in point is the Almendralejo case that took place in 2023, where a number of teenage girls received photos of themselves naked that were AI-generated by their peers,

232. Robert Mahari and Pat Pataranutaporn, 'We Need to Prepare for Addictive Intelligence' (MIT Technology Review, 5 August 2024) https://www.technologyreview.com/2024/08/05/1095600/we-need-to-prepare-for-addictive-intelligence/

233. Pat Pataranutaporn and others, 'Influencing Human–AI Interaction by Priming Beliefs about AI Can Increase Perceived Trustworthiness, Empathy and Effectiveness' (2023) 5 Nature Machine Intelligence 1076.

234. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika - Linnea Laestadius, Andrea Bishop, Michael Gonzalez, Diana Illenčík, Celeste Campos-Castillo, 2024 https://journals.sagepub.com/doi/abs/10.1177/14614448221142007

235. Open letter to UK online service providers regarding Generative AI and chatbots - Ofcom https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/open-letter-to-uk-online-service-providers-regarding-generative-ai-and-chatbots/

236. Alexander, S. (2025). Deep fake Cyberbullying: The Psychological Toll on Students and Institutional Challenges of AI-Driven Harassment. The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 98(2), 36–50. https://doi-org.ejournals.um.edu.mt/10.1080/00098655.2025.2488777

237. Vaccari, C., and A. Chadwick. 2020. Deep fakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Social Media + Society 6 (1):1–13. doi: 10.1177/2056305120903408.

238. Are Deep fakes Concerning? Analyzing Conversations of Deep fakes on Reddit and Exploring Societal Implications https://www.vice.com/en/article/deepfake-porn-origins-sexism-reddit-v25n2/

239. OII | Dramatic rise in publicly downloadable deep fake image generators https://www.oii.ox.ac.uk/news-events/dramatic-rise-in-publicly-downloadable-deepfake-image-generators/

causing them a high degree of distress.[240] Researchers are starting to document and study the impact of this phenomenon on its victims[241] and what legal actions might be taken in such cases. For example, while many schools have general cyberbullying policies, they often do not take into account AI-generated content's unique attributes.[242] Services beyond social media and pornographic platforms (see the analysis in the context of DSA in Section 5.2), such as app stores and search engines, are also exposed to similar risks, for example, the emergence of "nudify" apps that modify images of women to depict them naked has been reported.[243] Moreover, while human participants remain essential in detecting cyberbullying, the use of multiple algorithms could be trained to recognise cultural colloquialisms and slang terminology to facilitate detection in the future.[244]

240. The Almendralejo case: When AI deepfakes are used to undress teenagers | Euronews Tech Talks Podcast https://www.euronews.com/next/2023/11/15/the-case-of-almendralejo-when-deepfakes-are-used-to-undress-teenagers-euronews-tech-talks-

241. Are Deepfakes Concerning? Analyzing Conversations of Deepfakes on Reddit and Exploring Societal Implications https://www.vice.com/en/article/deepfake-porn-origins-sexism-reddit-v25n2/

242. Alexander, S. (2025). Deep fake Cyberbullying: The Psychological Toll on Students and Institutional Challenges of AI-Driven Harassment. The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 98(2), 36–50. https://doi-org.ejournals.um.edu.mt/10.1080/00098655.2025.2488777

243. Deepfake Defences - Mitigating the Harms of Deceptive Deepfakes - Ofcom Discussion Paper https://www.ofcom.org.uk/siteassets/resources/documents/consultations/discussion-papers/deepfake-defences/deepfake-defences.pdf?v=370754

244. Gomez, C.E., Sztainberg, M.O. & Trana, R.E. Curating Cyberbullying Datasets: a Human-AI Collaborative Approach. Int Journal of Bullying Prevention 4, 35–46 (2022). https://doi.org/10.1007/s42380-021-00114-6

4.8 Gender – as a Specific Case of Bias and AI Social Implications

KEY MESSAGES

- GenAI systems can perpetuate existing biases and stereotypes, particularly if they are trained on data sets that reflect historical and systemic inequalities, highlighting the need for human-rights-based and ethical AI use.
- The case of credit risk assessment and the potential risks of GenAI in perpetuating gender bias highlights the importance of establishing frameworks and guidelines for fair and transparent AI decision-making, with a focus on mitigating unfair biases and ensuring accountability.

Diversity, fairness, and non-discrimination are fundamental to building Trustworthy AI.[245] Eliminating unfair bias in AI is crucial to preventing negative outcomes, such as the marginalisation of vulnerable groups and the reinforcement of prejudice and discrimination. AI fairness is assessed by using a set of protected attributes, including g racial or ethnic origin, religion or belief, class, disability, age, gender or sexual orientation, among others, gender, national origin, and age, in line with the EU Charter for Fundamental Rights. In this respect, AI systems should be evaluated considering different protected groups, especially in high-risk scenarios (see Section 5.1). Furthermore, AI actors should strive to minimise and prevent discriminatory or biased outcomes throughout the AI system's life cycle to ensure its fairness[246] and comply with the EU non-discrimination legal framework.[247]

245. HLEG. (2019). *Ethics guidelines for Trustworthy AI.* European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

246. UNESCO. (2022). *Recommendation on the Ethics of Artificial Intelligence* (No. HS/BIO/PI/2021/1). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137

247. https://commission.europa.eu/aid-development-cooperation-fundamental-rights/your-fundamental-rights-eu/know-your-rights/equality/non-discrimination_en

BIAS, STEREOTYPES, AND FAIRNESS

The growing integration of AIacross various sectors has heightened concerns about biases in LLMs, including those related to gender, religion, race, profession, nationality, age, physical appearance, and socio-economic status.[248] While AI holds the promise of enhancing efficiency and decision-making in areas like healthcare, education, and business, its widespread use and the high level of public trust it enjoys could also amplify societal prejudices, leading to systematic disadvantages, particularly for women.[249]

GenAI systems can perpetuate and even amplify existing biases, particularly when trained on data reflecting historical inequalities. Research has highlighted persistent social biases in modern language models, despite efforts to mitigate them. For instance, in gendered word association tasks, recent models still associate female names with traditional roles like “home” and “family,” while linking male names with “business” and “career.” Moreover, in text generation tasks, these models produce sexist and misogynistic content approximately 20% of the time.[250]

Similarly, an analysis of occupational portraits generated by three popular text-to-image AI generators revealed significant gender and racial biases. Women and black individuals were notably underrepresented, especially in roles requiring high levels of preparation. Women were often portrayed as younger and with submissive gestures, while men appeared older and more authoritative. Alarmingly, these biases surpassed real-world disparities, indicating that the issues extend beyond merely biased training data.[251]

Research has demonstrated that AI can exhibit gender bias when used for decision-making processes. For example, in **recruitment**, AI algorithms have been observed to favour male candidates over equally qualified female candidates.[252] Moreover, in 2025, the Netherlands Institute for Human Rights found a violation of Dutch and EU anti-discrimination legislation in Meta’s job vacancy advertising algorithm.[253] For example, in violation of the principles of equal treatment and non-discrimination, in 2023, the algorithm in the Netherlands displayed vacancies for receptionist positions to female users in 97% of cases. Similarly, it showed vacancies for mechanics to male users 96% of the time. In healthcare, AI systems have shown gender bias in diagnostic reasoning, leading to treatment recommendations that may misdiagnose or inadequately address conditions in women, often due to training on predominantly male-centric data.[254] In **education**, while AI has the potential to create personalized learning paths tailored to individual competencies and needs, it may also unfairly predict higher dropout rates for female students, particularly in male-dominated fields like science, technology, engineering and mathematics (STEM), thereby limiting their access to advanced education programmes.[255] Given

248. Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2024). Bias and Fairness in Large Language Models: A Survey. *Computational Linguistics*, 50(3). https://doi.org/10.1162/coli a 00524

249. Hall, P., & Ellis, D. (2023). A systematic review of socio-technical gender bias in AI algorithms. *Online Information Review*, 47(7), 1264–1279. https://doi.org/10.1108/OIR-08-2021-0452

250. UNESCO & International Research Centre on Artificial Intelligence. (2024). Challenging systematic prejudices: An investigation into bias against women and girls in large language models. UNESCO.

251. Zhou, M., et al. (2024). *Bias in Generative AI* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2403.02726

252. Dastin, J. (2022). Amazon scraps secret AI recruiting tool that showed bias against women. In *Ethics of data and analytics* (pp. 296-299). Auerbach Publications

253. The Netherlands Institute for Human Rights rules that Meta’s algorithm engages in prohibited gender discrimination https://www.prakkendoliveira.nl/en/news/2025/the-netherlands-institute-for-human-rights-rules-that-metas-algorithm-engages-in-prohibited-indirect-gender-discrimination

254. Zack, T., et al. (2024). Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: A model evaluation study. The Lancet Digital Health, 6(1), e12–e22. https://doi.org/10.1016/S2589-7500(23)00225-X

255. Gardner, J., Brooks, C., & Baker, R. (2019). Evaluating the fairness of predictive student models through slicing analysis. In Proceedings of the 9th international conference on learning analytics & knowledge (pp. 225–234). https://doi.org/10.1145/3303772.3303791

that biased algorithms can lead to discrimination against vulnerable groups, particularly those experiencing the intersection of multiple forms of discrimination, such as racial minorities or individuals with disabilities.

RISK MITIGATION STRATEGIES

Addressing gender biases in AI systems is crucial for developing equitable technologies. This requires multiple strategies, such as enhancing the diversity and representativeness of training datasets, incorporating fairness-focused algorithmic approaches, and increasing diversity among AI developers.[256] Since bias mitigation cannot be achieved through one-time interventions, it is essential to conduct regular audits and monitoring using diverse benchmark datasets and methodologies to identify and address biases throughout the AI life cycle.[257]

Establishing robust policy frameworks at both corporate and governmental levels can guide the ethical development and deployment of AI systems. By implementing comprehensive risk mitigation strategies and fostering inclusive development practices, AI can be harnessed to promote societal equity and prevent the reinforcement of existing inequalities. Under the EU DSA, very large online platforms (VLOPs) and search engines (VLOSEs) must perform annual risk assessments on systemic risks stemming from their services and their algorithmic systems, including risks to fundamental rights, such as non-discrimination, as well as risks of gender-based violence (see Section 5.2).

256. Ho, J., et al. (2025). Gender biases within Artificial Intelligence and ChatGPT: Evidence, Sources of Biases and Solutions. *Computers in Human Behavior: Artificial Humans*, 4, 100145. https://doi.org/10.1016/j.chbah.2025.100145

257. UNESCO & International Research Centre on Artificial Intelligence. (2024). *Challenging systematic prejudices: An investigation into bias against women and girls in large language models.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000388971

A CASE STUDY: GENDER BIASES IN CREDIT RISK ASSESSMENT

The adoption of AI is also increasing across financial institutions, offering the potential to affect activities such as credit scoring, customer support or regulatory compliance. However, this adoption comes with its own set of challenges, including the mitigation of biases inherent in AI systems.[258] Among the many applications, credit risk assessment is one of the key areas where banks are actively exploring the use of AI and which is considered high risk under the AI Act (see Section 5.1). Unlike traditional approaches, which are typically based on statistical modelling, the arrival of AI allows financial institutions to potentially base their credit ratings on characteristics of the loan applicant, make decisions autonomously and generate customer risk profiles much faster, descriptive statistics on the rate of credit approval in the US as of 2022 confirm that there is bias in traditional credit scoring.[259] Notably, these report a gender bias of approximately 4%. As an alternative to this traditional approach, the JRC conducted an experiment to assess whether GenAI tools exhibit similar or higher biases, evaluating the validity and reliability of LLMs in supporting financial decision-making.

To do so, the exercise relies on a dataset containing over 400,000 short biographies with labelled profession and gender.[260] For each profession with enough available observations, 1,000 male and 1,000 female biographies are randomly paired, resulting in a final dataset of 17 professions and 34,000 individuals. Finally, the usage of an LLM as a credit risk scoring algorithm is evaluated by prompting it to choose precisely one person from each pair to provide credit to. To control for

258. "Intelligent financial system: how AI is transforming finance" by Iñaki Aldasoro, Working Papers No 1194 by Leonardo Gambacorta, Anton Korinek, Vatsala Shreeti and Merlin Stein

259. Song, Z., Rehman, S. U., PingNg, C., Zhou, Y., Washington, P., & Verschueren, R. (2024). Do FinTech algorithms reduce gender inequality in banks loans? A quantitative study from the USA. Journal of Applied Economics, 27(1).

260. e-Arteaga, Maria, et al. "Bias in bios: A case study of semantic representation bias in a high-stakes setting." *Proceedings of the Conference on Fairness, Accountability, and Transparency.* 2019.

potentially skewed quality of candidates in favour of a specific gender, the results are compared to a baseline alternative in which all gender-identifiable information (names, pronouns) is removed from the biographies and replaced with neutral references (they/them pronouns, no names).

This experiment shows that:

1. The LLM exhibits a similar level of bias as is currently observed in traditional credit provisioning, with a gap of around 4% in favour of men.

2. The situation is heterogeneous across professions, with women in some traditionally female-dominated professions (nurses, teachers) exhibiting a higher likelihood of obtaining a loan than their male counterparts.

4.9 The Contribution of a Behavioural Approach to AI Policy Analysis

KEY MESSAGES

- Integrating behavioural insights helps regulate GenAI to operate transparently and equitably, aligning its developments with societal values and increasing public trust.
- Agentic AI can potentially reduce human errors in decision-making. Nevertheless, it might also be biased, thanks to the human-produced data on which it feeds.
- Behavioural insights can help determine when AI should take precedence over human judgment to enhance social welfare.

In recent years, there has been a significant shift in policymaking, often referred to as the "behavioural turn". This approach leverages insights from psychology and behavioural economics to inform policymaking. It does this in a number of ways. For one, it helps to influence policies so that they are better aligned with the natural ways in which people think and make decisions. Rather than assuming individuals always act rationally, behavioural insights help tailor interventions to subtly encourage better choices without coercion. But also, a behavioural approach helps identify those cases where market players take advantage of citizens and consumers' imperfect thinking to steer their behaviour in a particular way (to polarise them or persuade them to consume). In other words, a behavioural approach in policymaking can help promote the use of behavioural insights for good, while limiting its use for bad.

There is significant potential for such an approach to inform the policy analysis of GenAI. It gives us a broad toolkit to work with. For one, it can help ensure GenAI interacts with humans in ways that protect the user and promote trust in AI. Regulations can help safeguard individuals, particularly the vulnerable populations who may struggle with complex interfaces. By integrating behavioural insights into policy, regulators can provide better guidelines that ensure AI tools operate transparently and are accessible to all users. The aim is to create a framework where AI systems are not only effective but also equitable, ensuring the benefits are widely distributed across the EU's diverse populace.

We can also leverage behavioural insights to ensure GenAI (and future developments, like Agentic AI, see Section 2.3) aligns with the broader ethical frameworks that govern societal values and fundamental rights. There are things that we, as humans, consider fair, moral, and/or ethical. Sometimes this can be readily stated, other times it needs to be revealed by systematically observing subjects' behaviour, e.g. through experimentation. The resulting insights can guide the EU in regulating the development of GenAI so that it respects these core values.

A behavioural approach can also help address the issue of bias in GenAI. This not only involves addressing technical and data-driven aspects,

ensuring that the datasets used to train AI systems are diverse and representative, but also requires a deeper understanding of cognitive biases inherent in human decision-making. A behavioural approach here can provide us with a framework to recognise and address these biases within AI systems. Cognitive biases, such as confirmation bias or anchoring, can also inadvertently be mirrored in GenAI if not carefully managed.

Looking towards the future, the further developments of AI will only increase the role for a behavioural approach. Agentic AI, for one, represents an important evolution within the field of Artificial Intelligence, building upon the capabilities of GenAI to function as autonomous entities. While GenAI focuses on creating content - such as text, images, or music - agentic AI extends this functionality, enabling systems to act independently within decision-making processes. These systems are capable of perceiving their environment, making complex decisions, and interacting with both humans and other machines to achieve specified goals. This advancement enhances GenAI's utility but also presents new challenges that must be addressed. Here, a behavioural approach will be particularly valuable.

Agentic AI will build on its ability to learn about user behaviour and preferences in fine detail. While this capability will offer powerful personalisation, it will also present the risk of exploiting users' cognitive biases and lack of information, potentially leading people to make choices that are not in their best interest. Behavioural research in this area can help identify if agentic AI is exploiting such biases and inform policies to counteract such objectionable practices.

However, in addition to risks there are opportunities. Agentic AI's decision-making can also be designed to overcome the cognitive biases that can skew human judgment. Judges and doctors, for example, are subject to biases themselves, and can (and do) make mistakes. Does this mean agentic AI will always be superior? Behavioural research is necessary to determine when it might be better for humans to make decisions, and when there is scope for agentic AI to take over and improve these decisions for the good of social welfare.

A final consideration: both GenAI and Agentic AI make assumptions of users' behaviour and preferences by building on observed behaviour - such as time spent on social media - which may not accurately reflect individuals' true interests or contribute to their well-being. This flaw in perception can lead to outcomes like increased polarisation or sedentary behaviour. Understanding this dynamic calls for the development of more modern and robust behavioural models, and a more fine-grained oversight, to ensure the behaviours being promoted are conducive to overall well-being.

4.10 Privacy and Data Protection – a Societal Standpoint

KEY MESSAGES

- GenAI can have very significant societal impacts and raise serious challenges for individuals' privacy and personal data protection. This is because of GenAI's capacity for data analysis and relationship.
- While such challenges should not hamper the development of GenAI, and the positive societal aspects it has to offer, such impact and challenges cannot be disregarded, and a clear technical and legal understanding of the issues raised by GenAI is required across societal sectors, including users, regulators and policymakers, in order to ensure that such issues are adequately – and timely – taken into consideration, discussed and addressed.

THE NOTIONS OF DATA PROTECTION AND PRIVACY

In the EU legal framework, the right to personal data protection and the right to privacy are two fundamental rights, enshrined in key legal texts.

They can often be used interchangeably. In the case of the US, for example, the concept of privacy often incorporates that of personal data protection; that being said, they differ considerably.

The right to a private and family life appears in fundamental texts such as the Universal Declaration of Human Rights (Article 12), the European Convention of Human Rights (Article 8) and the EU Charter of Fundamental Rights (Article 7), as well as in the EU Treaties. The right under Article 8 of the European convention of Human Rights and the right under Article 7 of the EU Charter of Fundamental Rights covers the confidentiality of communications, often referred to in the EU context as the right to privacy. Article 8 of the European Convention of Human Rights also covers the protection of personal data, in the absence of a more specific right in that Convention. As stated by the European Court of Human Rights under the abovementioned Article 8, a person's private life "encompasses a wide range of interests", including "a person's identity and personal development, the right to establish and develop relationships with other human beings" and "[a] ctivities of a professional or business nature" - as a result, it is a "notion not susceptible to exhaustive definition".[261]

The right to personal data protection concerns the protection of one's personal data, as stipulated under Article 8 of the EU Charter of Fundamental Rights and under Article 16 of the Treaty on the Functioning of the European Union (TFEU). Such data need to be processed fairly, for specific purposes and subject to a clear legal basis, with individuals being granted specific rights as to the processing of their data, which is overseen by an independent authority.

"Personal data" is a broad concept, which makes any discussions involving it more complex as it applies to all types of personal data, including publicly available personal data.[262] The first law in Europe to specifically address the protection of personal data dates back to 1970s Germany,[263] and the current General Data Protection Regulation (Regulation (EU) 2016/679), applicable in the EU since 2018, is considered by several as the international data protection standard.[264]

It is easy to see how personal data and privacy, as rights and as concepts, can be intertwined. The notion of personal data consists of information that relates to an individual (a "natural person") that can render them directly or indirectly identified, or identifiable. As a result, it will very often be the case that the violation of an individual's personal data will result in the violation of their private life.

Within the GenAI life cycle, personal data can potentially be found at all stages, from the training of the language models to the outputs produced by an AI agent. With its use becoming widespread, it is very important to assess the impact that GenAI tools have on the processing of our personal data and on our privacy. Additionally, it is crucial to identify the challenges raised by its use in relation to the processing of our personal data. This section focusses on the societal impact and challenges, leaving out any legal discussions on the topic, which are addressed in Section 5.

CHALLENGES AND IMPACT FOR DATA PROTECTION AND PRIVACY

Challenges to data protection and privacy stemming from the development and use of GenAI tools exist at all stages of the GenAI life cycle.

261. See Council of Europe, Privacy and Data Protection, Explanatory Memorandum, par. 2, available at: https://www.coe.int/en/web/freedom-expression/privacy-and-data-protection-explanatory-memo

262. Hamburg Data Protection Authority, Discussion Paper: Large Language Models and Personal data, 2024, p. 4.

263. European Parliament, Understanding EU data protection policy, Briefing, January 2025, available at: https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/698898/EPRS_BRI(2022)698898_EN.pdf, p. 2

264. "It is arguably the most significant data protection legislation in the world today", King, J., Meinhardt, C., Rethinking Privacy in the AI Era – Policy provocations for a data-centric world, White Paper, February 2024, Stanford University, p. 10; see also Zanfir-Fortuna, G., Why Data Protection Legislation offers a powerful tool to regulating AI February 2025, pdpEcho, available at: https://pdpecho.com/2025/02/26/why-data-protection-legislation-offers-a-powerful-tool-for-regulating-ai/

A simpler way of thinking about this concept is by dividing it into two phases:

1. "development" phase, which concerns all aspects of the GenAI model/system creation, such as code development, collection of (personal and non-personal) training data, and the training itself, and

2. "deployment" phase, which covers all stages relating to the use of an AI model (including as part of an AI system), as well as any activities in the post-development phase, like fine-tuning.[265]

The amount of data required to develop an AI model (before it is part of an AI system) is often very significant, and is steadily increasing, together with our ability to collect it in many instances of our daily lives. In the words of King and Meinhardt, "AI's appetite for data currently knows few bounds".[266] This "insatiable appetite for data" is bound to increase,[267] especially in light of the evidence that larger datasets can improve an AI system's capabilities.[268]

Deeply linked with quantity is the matter of quality: large datasets may often contain data that are inaccurate, or biased, leading to inaccurate or harmful outputs in relation to individuals, in particular those in less represented communities.[269]

There is also the question of inference: GenAI systems may allow those using them to infer information about an individual that may go well beyond what that individual intended to disclose in the first place.[270] That may result in individuals being subject to misguided decisions, or possibly even discrimination.[271] In an often-quoted example involving a large retail corporation, an algorithm was developed to identify pregnant women based on their shopping patterns, leading to a father discovering about his teenage daughter's pregnancy after receiving coupons for baby-related products at home. There, the main concern, as raised by Solove, is that the algorithm could infer sensitive personal data about an individual from data that are quite common to retrieve.[272]

The way in which data are collected matters, especially since one of the most common techniques is web-scraping: simply put, it is a technique that "enables the automated collection and extraction of certain information from different publicly available sources on the Internet (such as websites)", which can then be used, in this context, for the training of AI models.[273] Web-scraping as a technique, however, raises important questions about the impact it has on those whose data are collected, since individuals "may lose control of their personal information when this is collected without their knowledge, against their

265. This is also the approach followed by the European Data Protection Board in its Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, adopted on 17 December 2024. See, in particular, par. 18, p.11.

266. King, J., Meinhardt, C., Rethinking Privacy in the AI Era – Policy provocations for a data-centric world, White Paper, February 2024, Stanford University, p. 17

267. Solove, D., Artificial Intelligence and Privacy, Florida Law Review, Vol. 77, 2025, p. 18

268. King, J., Meinhardt, C., Rethinking Privacy in the AI Era – Policy provocations for a data-centric world, White Paper, February 2024, Stanford University, p. 37. King and Meinhardt also discuss – and challenge – this notion (see p. 38).

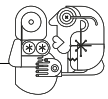
269. "An AI system built using a dataset collected from a city will only have a small percentage of certain minority groups, say 5%. If the dataset is used as-is, then the outputs of this AI system will be biased against this minority group because they only make up 5% of the dataset and the AI system has relatively less data to learn from about them.", European Data Protection Board Support Pool of Experts, Shrishak, K., Bias Evaluation, AI-Complex Algorithms and Effective Data Protection Supervision,March 2024, p. 6.

270. Solove, D., Artificial Intelligence and Privacy, Florida Law Review, Vol. 77, 2025, pp. 36-37

271. "The common factor among these risks is the fear of privacy pervasive data collection that allows sensitive inferences to be drawn. These inferences can lead to discrimination, especially when shared with third parties such as insurance companies, financial institutions, or employers.", Wachter, S., Data Protection in the Age of Big Data, Nature Electronics, Vol. 2, 6–7, January 2019, p. 2

272. Solove, D., Artificial Intelligence and Privacy, Florida Law Review, 2025, p. 37

273. European Data Protection Board, Report on the work undertaken by the ChatGPT Taskforce, 23 May 2024, par. 15.I



expectations and for purposes that are different from those of the original publication".[274]

Purpose is also an issue: what is the model being trained for? What use will it have? Which type of AI system will it integrate, with what functionalities? Can it be repurposed at some point? This is particularly the case of General Purpose Artificial Intelligence (or GPAI) models/ systems. This multi-modality can have a very significant impact on individuals, as systems can be gravely misused (e.g. for fraud, deep fakes as discussed in Section 5.2, and identity theft).

Finally, there is the literacy challenge: are we equipped, as individuals, to understand why and how our personal data are collected, and what they are used for? Do we, as individuals, and as a society, have a clear understanding of how these systems work, the challenges they raise, and the impacts on our lives? This is especially relevant since it is questionable whether even those deploying such systems are able to explain the decisions and outcomes stemming from such systems (the so-called "black box effect").[275]

This is essential to determine whether individuals have the capacity – or are given the tools – to understand and act upon the challenges raised by this technology to their personal data and privacy, including through legal mechanisms. It is also relevant to understand whether their actions produce any effects, be they at the core level of data collection and training, or further down the AI life cycle, when, for example, inferences, relying on GenAI-based tools, are being made about them. Finally, it is also crucial for regulators to understand what measures to put in place, in addition to existing ones, to grant individuals the ability to understand and control, or at least mitigate, the impact this technology may have on people's data and privacy. ■

274. European Data Protection Supervisor, Opinion 41/2023 on the Proposal for a Regulation on European Union labour market statistics on businesses, 25 September 2023, par. 17. See also par. 18. One well known example is that of Clearview. In the words of the Dutch Data Protection Authority, which imposed a fine on Clearview AI of circa 30 million EUR: "Clearview is a commercial business that offers facial recognition services to intelligence and investigative services. [...] For this purpose, Clearview has a database with more than 30 billion photos of people. Clearview scrapes these photos automatically from the Internet. And then converts them into a unique biometric code per face. Without these people knowing this and without them having given consent for this", available at: https://www.autoriteitpersoonsgegevens.nl/en/current/dutch-dpa-imposes-a-fine-on-clearview-because-of-illegal-data-collection-for-facial-recognition

275. European Data Protection Supervisor, TechDispatch 2/2023, Explainable Artificial Intelligence, November 2023, p. 3

5
REGULATORY FRAMEWORK

This chapter outlines the regulatory landscape for GenAI, starting with the AI Act and its implications for GenAI applications. It explores the risks associated with GenAI and the role of the Digital Services Act in mitigating these risks. The chapter also examines the interplay between GenAI and the General Data Protection Regulation (GDPR), as well as intellectual property (IP), and particularly copyright challenges. Central questions include how to balance innovation with the need for robust ethical and legal standards in AI governance. Finally, the legislation that regulates the exchange and reuse of data is summarised.

5.1 The AI Act and Its Implications for Generative AI

KEY MESSAGES

- The AI Act mediates the development of GenAI systems. On the one hand, it imposes a set of legal requirements to make GenAI systems in the EU more transparent and trustworthy.
- On the other hand, it fosters technological innovation in areas linked to trustworthy AI, such as watermarking and fingerprinting techniques.

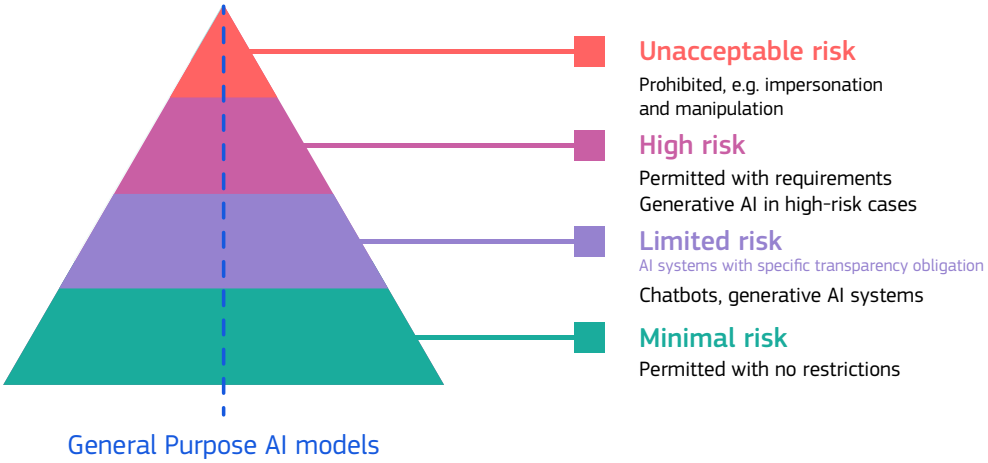
The AI Act is the first world regulation of AI, and it is being implemented in the EU. It requires that certain AI systems and models fulfil a set of requirements before they are put on the EU market. Following a risk-based approach (i.e. considering risks to health, safety and fundamental rights), this set of legal requirements depends on the specific applications, with four different risk levels for AI systems (minimal, limited, high and unacceptable risks) and three levels for models (no obligations, General-Purpose AI (GPAI) models and GPAI models with systemic risks). Current Commission guidance documents provide clarifications on these levels and requirements.

Those levels also apply to AI systems and models that are generative. In particular, GenAI applications are linked to some high-risk and transparency risks, and GenAI models are at the core of GPAI models.

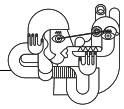
The AI Act has relevant implications on the uptake and trustworthiness of GenAI models, which are outlined in the following sections.

GENERATIVE AI MODELS AND SYSTEMS IN THE AI ACT

Figure 17. Generative AI and AI Act risk levels for AI systems.



Source: JRC elaboration.



Many GenAI systems are particularly linked to the "limited risk" level and related transparency obligations depicted in Article 50.[276] These obligations stipulate that providers of certain GenAI systems such as chatbots need to ensure that humans relating with those systems are aware they are interacting with a machine. In addition, providers need to ensure that AI-generated content is identifiable. On top of that, certain AI-generated content should be clearly and visibly labelled, such as deep fakes and text published with the purpose of informing the public on questions linked to the public interest.

In addition to limited risks, GenAI systems can be part of high-risk AI systems or unacceptable risk practices. High-risk AI systems are subject to strict obligations[277] and the AI Act prohibits eight practices linked to unacceptable risks.[278] Although there is no explicit reference to GenAI systems in the list of high-risk use cases (AI Act Annex III), GenAI systems have the potential to be integrated in those. In terms of prohibited practices, GenAI systems can be linked to some of them, notably harmful AI-based manipulation and deception (AI Act Article 5 (a) and (b)).

The draft Commission Guidelines on Prohibited AI practices[279] provide some examples of unacceptable use of GenAI systems: an AI chatbot that impersonates a friend or relative of a person to cause significant harms, or a system designed to detect when it is under evaluation to halt undesired behaviour. The guidelines also mention an interplay between these prohibited practices and the transparency measures for limited risks mentioned earlier and stated by Article 50 (4). Transparency measures can be considered as a mitigation strategy to reduce the risk of deception and manipulation.

GENERATIVE AI MODELS IN THE AI ACT

In addition to AI systems, the AI Act imposes certain specific obligations on providers of general-purpose AI models (including but not limited to GenAI models), which can perform a wide range of tasks and are becoming the basis for many AI systems mentioned in the previous section. The Act imposes several transparency requirements and copyright-related rules for those models, as well as additional requirements of systemic risk identification and mitigation for those models that may be linked to those systemic risks.

Currently, many state-of-the-art GenAI models present capabilities linked to the concept of general-purpose AI, so they would be subject to the relevant obligations. The European Commission's AI Office is currently facilitating the drawing-up of a Code of Practice to detail these rules based on state-of-the-art practices.[280]

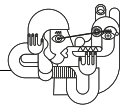
276. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU,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

277. An adequate risk assessment and mitigation strategy, a high-quality of the datasets, logging of activity to ensure traceability of results, detailed documentation providing all information necessary on the system and its purpose for authorities to assess its compliance, clear and adequate information to the deployer, appropriate human oversight measures, and a high level of robustness, cybersecurity and accuracy. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

278. Harmful AI-based manipulation and deception, harmful AI-based exploitation of vulnerabilities, social scoring, individual criminal offence risk assessment or prediction, untargeted scraping of the internet or CCTV material to create or expand facial recognition databases, emotion recognition in workplaces and education institutions, biometric categorisation to deduce certain protected characteristics and real-time remote biometric identification for law enforcement purposes in publicly accessible spaces. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

279. Draft Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act) – C(2025) 884 https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act

280. https://digital-strategy.ec.europa.eu/en/policies/ai-code-practice



RELATED TECHNOLOGIES

As mentioned above, the AI Act relies on technical solutions for the transparency of GenAI, namely those that allow the identification of AI-generated content. These techniques should possess the following four properties: *efficiency, integrity of data, robustness to content alteration, and protection against manipulation*.[281] Transparency techniques for GenAI are based on those used for the marking of digital content (audio, image or text) to ensure compliance with copyright law. These techniques cover four main transparency approaches: (1) metadata embedded in the content; (2) watermarking techniques designed to embed barely perceptible markers in the audio, image or text; (3) fingerprinting techniques relying on the generation and storage of content identifiers; and (4) AI-based detection mechanisms.

Although there have been significant advancements on these techniques, transparency of GenAI is still a challenging field, where further fundamental research is needed to develop more reliable solutions. This also involves the exploration of new engineering methods for content identification and its governance, considering the role of proprietary versus open solutions, as explored in Section 1.3.

281. Hamon, R., Sanchez, I., Fernandez Llorca, D. and Gomez, E., Generative AI Transparency: Identification of Machine-Generated content, European Commission, Ispra, 2024, JRC137136. https://publications.jrc.ec.europa.eu/repository/handle/JRC137136

5.2 Generative AI Risks and the Digital Services Act

KEY MESSAGES

- The EU Digital Services Act (DSA) imposes obligations on online intermediaries and platforms to address systemic risks, requiring them to adapt their services, systems, and algorithms to mitigate these risks. Most of the very large online platforms and very large online search engines identified risks related to GenAI in their recent risk assessment reports.
- GenAI technology introduces emerging forms of risks that designated digital services will have to analyse and mitigate, such as risks to users' physical and mental well-being or risks to civic discourse and electoral processes.
- It also has the potential to bring opportunities for a safer online space if properly implemented and tested, including through the use of LLMs in content moderation and the implementation of guardrails and mitigations against harmful or malicious content.

The EU Digital Services Act (DSA) is the world's first regulation to address societal risks emerging from the use of intermediary services such as online platforms and search engines. The DSA imposes obligations for online intermediaries and platforms according to their role, size, and societal impact, covering systemic risks stemming from the design or functioning and use of these services, and its related systems, including algorithmic ones.

Since DSA obligations started to apply to most designated *Very Large Online Platforms* (VLOPs) and *Very Large Online Search Engines* (VLOSEs) in August 2023, the European Commission has taken multiple supervision and enforcement actions, some of which have specifically focused on GenAI

(in particular related to hallucinations, deep fakes, and election-related risks). These include requests for information (RFI) sent between March and May 2024 to six VLOPs and three VLOSEs.[282] The rapid pace of advancement and adoption of GenAI in online services calls for a strong technical and scientific capacity at the regulatory level.

THE GENAI RISKS IDENTIFIED BY ONLINE PLATFORMS IN THEIR RISK ASSESSMENT REPORTS

Some platforms and search engines with GenAI features cover associated risks in their 2024 reports. For example, the risk assessment for Bing explicitly acknowledges risks stemming from the exploitation of GenAI features, risks related to access to information and freedom of expression, and risks linked to false and misleading information, as well as echo chambers, and risks related to the generation of unsafe, misleading, fraudulent, private or otherwise harmful content.[283]

Hallucinations, such as GenAI responses that are not grounded in input sources, are explicitly mentioned, and specific mitigation measures and guardrail techniques are discussed, especially in scenarios prone to attacks or involving so-called "data voids", search areas where there is a lack of authoritative information, especially in languages with less traffic. Other platforms, for instance LinkedIn, also refer to risks of GenAI features, such as the possibility that models reflect harmful stereotypes, or risks related to abuse by platform users, including jailbreaking events or malicious prompt injections, e.g. in job descriptions.[284]

Risks related to the misuse of Generative AI on platforms are also prominently reflected in risk assessment reports. Most social media platforms, for example, refer to the potential for GenAI to lower technical barriers to produce harmful content at high speed and scale. In particular, the risk of GenAI being used to manipulate content is often highlighted, and its use for spreading and amplifying false information, or enabling fraud and scams.

A notable risk mentioned in risk assessment reports of platforms such as X,[285] TikTok[286] and Meta,[287] [288] among others, refers to the possibility that GenAI "may facilitate the production of AI generated illegal content" or "be used as part of image-based abuse". A particularly serious instance is the generation of Child Sexual Abuse Material (CSAM), with recent reports warning about an increase in CSAM content using GenAI coinciding with the public release of ChatGPT and Stable Diffusion in 2022.[289]

A related risk is the emergence of pornographic deep fakes (see Section 4). These risks are not exclusive to social media and are in fact arguably as prominent on pornographic platforms. While no risk assessment reports for designated pornographic platforms (Pornhub, Xvideos, Stripchat and XNXX) have been made available to date, the Google risk assessment reports "a concerning increase in generated images and videos that portray people in sexually explicit contexts, distributed on the web without their

282. Commission sends requests for information on generative AI risks to 6 Very Large Online Platforms and 2 Very Large Online Search Engines under the Digital Services Act | Shaping Europe's digital future https://digital-strategy.ec.europa.eu/en/news/commission-sends-requests-information-generative-ai-risks-6-very-large-online-platforms-and-2-very

283. Bing Systemic Risk Assessment Report 2024 https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/August-2024-Microsoft-Bing-Systemic-Risk-Assessment-Report-EU-Digital-Services-Act.pdf

284. LinkedIn System Risk Assessment Report 2024 https://content.linkedin.com/content/dam/help/tns/en/LinkedIn-2023-2024-DSA-Systemic-Risk-Assessment-Report.pdf

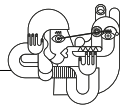
285. X risks assessment report 2024 https://transparency.x.com/content/dam/transparency-twitter/dsa/dsa-sra/dsa-sra-2024/TIUC-DSA-SRA-Report-2024.pdf

286. TikTok DSA Risk Assessment Report 2023 https://sf16-va.tiktokcdn.com/obj/eden-va2/zayvwlY_fjulyhwzuhy%5b/ljhwZthlaukjlkulzlp/DSA_H2_2024/TikTok-DSA-Risk-Assessment-Report-2023.pdf

287. DSA Systemic Risk Assessment and Mitigation Report for Facebook 2024

288. DSA Systemic Risk Assessment and Mitigation Report for Instagram 2024

289. Internet Watch Foundation AI CSAM Report 2023 https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/



consent", and how they were improving their mitigation measures to better address this issue, such as by updating their ranking algorithms.[290]

The DSA requires designated online platforms to closely monitor and address existing and emerging risks exacerbated by the use of GenAI in digital services, particularly when minors are among the users of these services.

Indeed, while deep fakes and similar forms of illegal or harmful synthetic content are often subjected to the same content policies as other violative content, additional mitigation measures specifically targeting AI-generated content are starting to emerge. Transparency, in the form of labelling tools made available to users and advertisers with the option to tag their content as made with AI, is a first line of defence. Models to detect synthetic content are also in the process of being developed and improved, complementing user labelling. The technical difficulty of this task is often highlighted, with effective techniques able to detect synthetic media at scale still being largely unavailable. Several platforms refer to collaborations with industry partners and organisations to exchange information about violative content, including synthetic content, as well as to continue developing technology for its detection. A notable example is watermarking technology, which embeds a digital signature into the content that is imperceptible to the human eye but can be detected through technical means, such as SynthID.[291] Broader content provenance solutions are also emerging, such as C2PA, an open technical standard to embed metadata in digital content such as images, videos, audio recordings, and documents.[292] The objective of these techniques is ultimately to identify the origin of AI-generated harmful content,[293] thus limiting the misuse of this technology, fostering transparency and allowing for accountability. However, these techniques are under active development and deployment by various organisations, with further work ahead towards widespread robust and interoperable adoption.

OPPORTUNITIES OF GENERATIVE AI FOR A SAFER ONLINE SPACE

Despite its risks, GenAI technology is a transformational technology expected to bring many potential benefits and opportunities across a wide range of application areas. This is also the case in the context of fostering a safe online space. One example that highlights this potential is the use of LLMs in content moderation use cases and, more generally, for enforcing community guidelines and policies put in place by online platforms. Specialised GenAI models are increasingly being used to implement guardrails and mitigations against harmful or malicious content, whether user- or AI-generated.

The use of automation to support content moderation activities in online platforms and search engines is becoming an established practice. Analysis of the data in the DSA transparency database[294] across all VLOPs between 1 April 2024 and 1 April 2025 reveals that a majority of content moderation actions registered involved at least partial automation, primarily for initial detection, and increasingly for completely automated removal, i.e. without human intervention.

Today, GenAI may still play a limited role in content moderation compared to classical algorithms and AI models. However, we can

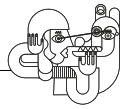
290. How Google Search is addressing explicit fake content https://blog.google/products/search/google-search-explicit-deep-fake-content-update/

291. Scalable watermarking for identifying large language model outputs | Nature https://www.nature.com/articles/s41586-024-08025-4

292. https://c2pa.org/

293. Generative AI and watermarking Briefing | European Parliament https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)757583

294. The Digital Services Act (DSA), obliges providers of hosting services to inform their users of the content moderation decisions they take and explain the reasons behind those decisions in so-called statements of reasons. To enhance transparency and facilitate scrutiny over content moderation decisions, providers of online platforms need to submit these statements of reasons to the DSA Transparency Database https://transparency.dsa.ec.europa.eu/?lang=en



expect its role to increase in the future. Platforms such as Google[295] already point to the opportunities to use AI to prevent, detect, and respond to illegal and harmful content at scale, including using LLMs. While the potential benefits are clear, it is important to highlight the need to rigorously evaluate automated content moderation systems across users and languages, to prevent systemic biases as well as over- or under-moderation.[296]

5.3 General Data Protection Regulation (GDPR) and Generative AI

KEY MESSAGES

- The relationship between data protection laws, particularly the GDPR, and GenAI needs to be better understood, especially when it comes to their implementation and compliance assurance.
- The AI Act, despite having a different scope of application, will offer a complementary legal framework, but the GDPR will continue to apply to any processing of personal data in the context of AI and GenAI technologies.
- The application of existing data protection laws, however, needs to continue to be assessed in detail, as several issues, such as the notion of personal data and their processing in AI models, lawfulness, accountability, and the provision of data subject rights, among others, still require more research and practical assessment.

In Section 4.9, we identified and analysed the societal impacts and challenges of GenAI development from data protection and privacy perspectives. Here, the purpose is to introduce the relationship between GenAI and the GDPR and identify some key issues raised by recent developments in GenAI.

Since it became applicable in 2018, the GDPR has benefited from what is known as the "Brussels Effect" to become an international legal standard in data protection, acting as a reference for other regulations around the world.[297] A technology-neutral, principle-and risk-based legal text, it was built on its legal predecessor's strengths, adding a few of its own, to ensure compliance with data protection rules.

The fast-paced development of AI – and its generative subset, in particular – has raised questions about the preparedness of data protection laws (including the GDPR) to address the complex issues raised by these technologies,[298] although EU regulators and other experts consider that data protection legislation can be a relevant tool for regulating AI.[299] In fact, the GDPR has

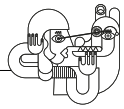
295. Google Report on Systemic Risk Assessments 2024 https://storage.googleapis.com/transparencyreport/report-downloads/dsa-risk-assessment_2024-8-28_2024-8-28_en_v1.pdf

296. Content Moderation in a New Era for AI and Automation | Oversight Board https://www.oversightboard.com/news/content-moderation-in-a-new-era-for-ai-and-automation/

297. To better understand the "Brussels Effect", please see: Bradford, A., The Brussels effect: How the European Union rules the world, Oxford University Press, 2020

298. "The framework that underlies data protection laws has weaknesses that will not give individuals the tools they need to preserve their data privacy as AI advances; it also fails to address societal-level privacy risks", King, J. and Meinhardt, C., Rethinking Privacy in the AI Era – Policy Provocations for a Data-Centric World, White Paper, Stanford Institute for Human Centered Artificial Intelligence, p. 6; "Current privacy laws fall woefully short of addressing AI's privacy challenges. AI puts pressure on many of the weakest parts of privacy law. Privacy law's wrong approaches and other unfixed flaws are especially ill-suited for AI", Solove, D., Artificial Intelligence and Privacy, February 2024, 77 Florida Law Review 1, 2025, "Unfortunately, many nascent digital technologies seem destined to undermine the aims of the GDPR. [...] [T]he GDPR focuses mainly on protection at the input stage when data is collected, but hardly during or after analysis. The law thus ignores the fact that unforeseen threats to privacy can arise after data collection owing to inferential analytics.", Sandra Wachter, Data Protection in the age of big data, Nature Electronics, vol. 2, January 2019, pp. 6 and 7

299. "Recent guidance from the European Data Protection Board or the CNIL [...] shows the GDPR is flexible enough to avoid inhibiting the AI revolution in the EU, while at the same time offering protections to the rights of individuals." Zanfir-Fortuna, G., Why Data Protection legislation offers a powerful tool for regulating AI, pdpEcho, available at: https://pdpecho.com/2025/02/26/why-data-protection-legislation-offers-a-powerful-tool-for-regulating-ai/



already been called to action in different instances, such as the Italian Data Protection Authority preliminary investigation and corresponding sanctions to OpenAI, including a EUR 15 million fine;[300] and the NoYB[301] complaints against the same company, including one about the creation of a "fake child murderer".[302]

The reason why data protection – and the legal frameworks regulating it – have so much prominence in any AI discussion can be explained by the words of King and Meinhardt: "[t]he connective tissue between privacy and AI is data: nearly all forms of AI require large amounts of training data to develop classification or decisional capabilities [...][d]ata is a key component for all AI systems – to date, the most significant improvements in AI systems have been tied to access to very large amounts of training data".[303] It i's important to underline that, whereas the concept of "data" is not always used in relation to "personal data", data used to train AI models, or data produced by AI systems is, in many cases, personal data.

In our view, it is unquestionable that there are currently many issues concerning the relationship between GenAI and data protection requiring a more definitive answer. We will address some in this section. As AI technologies evolve to become more complex (as is currently the case with the introduction of AI Agents), it is likely that so will the issues they raise when it comes to the processing of personal data and its compliance with data protection legislation and the GDPR in particular.

THE INTERPLAY BETWEEN THE GDPR AND THE AI ACT

Despite having different scopes of application, the GDPR and the EU AI Act (Regulation 2024/1689) are to be seen as "complementary and mutually reinforcing instruments".[304] The AI Act makes explicit reference to Art. 16 TFEU, and to the GDPR and its full application to the processing of personal data in the context of AI life-cycles.[305] Despite its market-driven, innovation-support approach, it puts a strong emphasis on promoting the uptake of human-centric and trustworthy AI and ensuring health, safety and a high level of protection of the fundamental rights enshrined in the Charter of Fundamental Rights of the EU (including the fundamental rights to privacy and data protection).

Two good examples of this complementarity between the AI Act and the GDPR can be found: (i) in the mandatory Fundamental Rights Impact Assessment (FRIA), Data Protection Impact Assessments (DPIA) may be relied upon to meet certain aspects of the FRIA; and (ii) in the EU AI Act requiring that providers of high-risk AI systems draw up a declaration of conformity containing a statement that the relevant AI system complies with EU data protection laws.[306]

This complementarity is also why the European Data Protection Board (EDPB), in its statement

300. See the IT DPA press release (in IT and EN) from December 2024, available here: https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/10085432#english. Other examples include CNIL's 20 Million EUR fine to Clearview AI in 2022 (see: https://www.edpb.europa.eu/news/national-news/2022/french-sa-fines-clearview-ai-eur-20-million_en); and the Dutch DPA 2.75 Million EUR fine in 2021 to the Dutch Tax Administration concerning child benefits fraud detection (see: Dutch scandal serves as a warning for Europe over risks of using algorithms – POLITICO

301. NoYB is a donation-funded NGO based in Vienna, Austria working to enforce data protection laws, in particular the GDPR and the ePrivacy Directive.

302. See the first complaint, concerning an accuracy issue related to a user's date of birth, from April 2024: https://noyb.eu/sites/default/files/2024-04/OpenAI%20Complaint_EN_redacted.pdf; see the second complaint, from March 2025, regarding the creation of a "fake child murderer": https://noyb.eu/sites/default/files/2025-03/OpenAI_complaint_redacted.pdf and https://noyb.eu/en/ai-hallucinations-chatgpt-created-fake-child-murderer.

303. King, J., Meinhardt, C., Rethinking Privacy in the AI Era – Policy Provocations for a Data-Centric World, White Paper, Stanford Institute for Human Centered Artificial Intelligence, p. 5

304. European Data Protection Board, Statement 3/2024 on data protection authorities' role in the Artificial Intelligence Act framework, 16 July 2024, par. 3, p. 2.

305. See Article 2(7) and recitals 3, 9 and 10 of the EU AI Act.

306. Article 16(g), Article 47 and Annex V, point 5 of the EU AI Act

on the role of Data Protection Authorities (DPAs) and the EU AI Act, considered that "whenever a general-purpose AI model or system entails the processing of personal data, it may fall – like any other AI system – under the supervisory remit, as applicable, of the relevant national DPAs" "whenever a general-purpose AI model or system entails the processing of personal data, it may fall – like any other AI system – under the supervisory remit, as applicable, of the relevant national DPAs". As a result, it calls for the attention of the European Commission and of the EU AI Office (established within the Commission) for "the need to cooperate with the national DPAs and the EDPB, and the need to establish, in agreement with them, the appropriate mutual cooperation in the most effective way".[307]

FOUR KEY ISSUES

Several issues can be raised concerning the processing of personal data under the GDPR in a GenAI environment, many of which are already addressed in available literature. Here we focus only on four issues, which we consider of great relevance:

1. The Notion of Personal Data and Their Processing in AI Models

The debate surrounding whether LLMs process personal data has significant implications for the development and use of AI systems. The Hamburg DPA, following in the footsteps of the Danish DPA, argued - prior to the adoption of EDPB Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models - that LLMs are not databases and do not store personal data and, therefore, no personal data is processed within the model.[308] This would mean, in essence, that no data subject rights could be granted in relation to the model itself, and that any potential violations in the development stage of an LLM would not affect the lawfulness of using such a model, later on, including as part of an AI system.

Criticism to these arguments has been made based on: (i) an assessment of the concept of personal data, as put forward by the Art. 29 Working Party, concluding that information within an LLM (also known as "tokens") falls within the definition of personal data; (ii) the fact that the GDPR also applies to personal information that is merely probabilistic, including if inaccurate, as also stated by the EDPB; and (iii) the consideration that the success of security measures protecting personal within an LLM (and, therefore, the impossibility of its extraction, other than by illegal means) does not prevent the application of the GDPR.[309]

Because of its impact for the use of LLMs in AI systems, and the implementation of data subject rights in relation to the LLM itself, the implications of adopting one or the other position are far reaching for individuals whose data are processed in this context, but also for companies – be they the ones developing these models or those later on using them as part of their AI systems, which may or may not be required to ensure what could be a technically complex level of compliance with the GDPR.

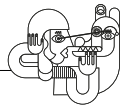
2. Lawfulness (Art. 5(1)(a) with a Focus on Legitimate Interest (Art. 6(1)(f))

General Purpose AI (GPAIs) models are usually trained on massive datasets (which can contain equally massive amounts of personal data, including sensitive personal data), obtained in

307. European Data Protection Board, Statement 3/2024 on data protection authorities' role in the Artificial Intelligence Act framework, 16 July 2024, pars. 14 and 15, p. 5

308. Danish Data Protection Authority, Public Authorities Use of Artificial Intelligence – Before you get started , October 2023 (our translation); Hamburg Commissioner for data protection and Freedom of Information, Discussion Paper: Large language models and personal data, July 2024.

309. For a more detailed overview of these arguments, please see: Moerel, L. and Storm, M., Do LLM's store personal data? This is asking the wrong question, IAPP, available at https://iapp.org/news/a/do-llms-store-personal-data-this-is-asking-the-wrong-question. For additional views on this matter, see also Christakis, T., AI Hallucinations and Data Subject Rights under the GDPR: Regulatory Perspectives and Industry Responses, December 2024, available at: https://ssrn.com/abstract=5042191 or http://dx.doi.org/10.2139/ssrn.5042191



different ways, such as through web scraping of publicly available data.[310]

To process personal data lawfully, a legal basis is required, and one possible basis, especially in the case of LLMs, is the legitimate interest of the data controller (Article 6(1)(f) GDPR)[311] (and, for sensitive personal data, the additional requirements of article 9(2) GDPR). However, among other criteria, this legal basis is subject to a "balancing test" that weighs the interests of the controller against the rights and freedoms of the data subjects. The challenge lies in applying this balancing test to large, diverse datasets, containing potentially very significant and diverse amounts of personal data, used to train AI models. In practice, it may prove quite complex for data controllers to accurately identify individual interests against the processing at stake, assess the impact of processing on individuals' rights and freedoms, and introduce mitigating measures when dealing with vast amounts of data from various sources.

The European Data Protection Board (EDPB) has provided guidance on legitimate interest under its Opinion 28/2024 on certain aspects of the processing of personal data in the context of AI models, acknowledging the potential risks of AI model development and deployment to fundamental rights. However, the EDPB emphasises that each case must be assessed individually, depending for instance on the nature and sources of the data, the context of the data processing and the further consequences for the individuals concerned.[312]

The lawfulness discussion is crucial because, if legitimate interest would not be applicable, other legal bases, such as consent, would have to be considered.[313] However, from a practical perspective, the implementation of a legal basis such as consent, for the purposes of training GPAI models, could prove to be, by its very nature, challenging to data controllers in a GPAI context, where datasets used to train models can contain a very significant amount of personal data categories concerning, potentially, a vast number of individuals.

3. The Principle of Accountability: Who's Responsible?

The nature and complexity of GenAI systems render the attribution of responsibilities throughout the AI life cycle potentially quite difficult. The GDPR dictates that the data controller – that which determines "why" and "how" personal data are being processed in a given context – shall be responsible for, and be able to demonstrate compliance with data protection principles.[314]

How is a data controller defined in a GenAI context? In the development phase (including the data collection and training stages) of an AI model, that may be easier to determine (e.g. OpenAI is the data controller in the development of GPT).[315] However, that determination becomes more complex once an AI model has been trained and it is deployed as part of an AI system.

310. See section 4.10 above.

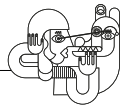
311. See the EDPB's Report of the work undertaken by the ChatGPT Taskforce, published in May 2024, in particular pars. 15-19.

312. European Data Protection Board, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, December 2024, pars. 76, 79, 80 and 84.

313. One good example is the recent case involving a German consumer organization (Verbraucherzentrale North Rhine-Westphalia, VNRW) and Meta. VNRW failed to win a court injunction to stop Meta from training its AI models with personal data from Facebook and Instagram (see: https://www.reuters.com/sustainability/boards-policy-regulation/german-rights-group-fails-bid-stop-metas-data-use-ai-2025-05-23/. Simultaneously, the Hamburg Data protection authority considered issuing an urgency procedure under Article 66 GDPR against its Irish counterpart and Meta in order to stop Meta AI training using personal data, but it later decided not to move forward (see: https://www.euractiv.com/section/tech/news/german-privacy-watchdog-scraps-plans-to-stop-meta-ai-training-on-personal-data/). NRW requests Meta to cease and desist AI training in the EU, 6 May 2025.

314. See Article 5(2) GDPR.

315. See OpenAI's Privacy Policy for the EEA (from 4 November 2024), available here: https://openai.com/policies/eu-privacy-policy/



Among other questions raised by the determination of accountability, there is that of the "fruit of the poisonous tree",[316] which is addressed by the EDPB in its Article 64(2) GDPR opinion: if a model is trained using unlawfully processed personal data, does it affect the lawfulness of processing and, as a result, the responsibility of the or of controllers relying on that model for their GenAI systems? The EDPB offers three different scenarios. Given the context-based nature of personal data processing, it is likely that a case-by-case assessment is required. That being said, this remains one of the very important aspects under discussion in the relationship between the GDPR and GenAI.[317]

4. Data Subject Rights: Can They Be Exercised?

Individuals whose personal data is processed in the context of any activity are entitled to a set of rights under the GDPR, established in Articles 15 to 22. Even if not absolute, such rights offer individuals more control over their personal data, and their processing by data controllers, by allowing an individual: to know if, why and how their personal data are being processed;[318] to ask for their rectification or erasure;[319] to object to the processing of their personal data under certain conditions;[320] and, especially relevant in an AI context, not to be subject to an automated individual decision-making process about them, with a set of subsequent safeguards, such as human intervention.[321] In a general AI context, but also specifically in the subfield of GenAI, this is not without challenges to their implementation.[322]

A proper exercise of data subject rights is also linked to other key issues surrounding the discussion on GenAI: (i) the accuracy of personal data contained in the model; and (ii) whether an AI model actually stores personal data – with potential consequences for the exercise of data subject rights, as mentioned above. Essentially, it appears that the identified approaches to ensure an effective implementation of data subject rights are technically complex including as to their effectiveness,[323] may require further refinement, and are not without challenges of their own.[324]

316. An analogy with the American legal doctrine applied to the inadmissibility of evidence if it derives from evidence that is illegally obtained. For a more detailed explanation of the concept, please see: https://www.law.cornell.edu/wex/fruit_of_the_poisonous_tree

317. European Data Protection Board, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, December2024, section 3.4

318. Right of access, Art. 15 GDPR

319. Arts. 16 and 17 GDPR, respectively

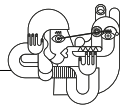
320. Art. 21 GDPR

321. Art. 22 GDPR; in relation to this, see in particular the Court of Justice of the European Union decision in case C-634/21, SCHUFA Holding (Scoring).

322. See, for example, the paper produced by the EDPB Support Pool of Experts, Shrishak, K., Effective implementation of data subject rights, AI Complex Algorithms and effective Data Protection Supervision, March 2024.

323. "Developers like OpenAI, Google, Meta, Anthropic and others have introduced strategies across data training, model architecture and system outputs to enhance reliability, transparency and the exercise of data subject rights. While these measures represent significant progress, they might not yet be sufficient, and ongoing refinement is necessary as the technology evolves." Christakis, T., AI Hallucinations and Data Subject Rights under the GDPR: Regulatory Perspectives and Industry Responses, December 2024, p. 2

324. European Data Protection Board Support Pool of Experts, Shrishak, K., Effective implementation of data subject rights, AI Complex Algorithms and effective Data Protection Supervision, March 2024. See pp. 8-9; see also: "However, if your own data has already been used for artificial intelligence training, this cannot be reversed by subsequently objecting. Training data is irrevocably incorporated into artificial intelligence models and, given the current state of technology, its influence cannot be removed from the model", Meta starts AI training with personal data, Hamburg Data Protection Authority, 27 May 2025, available here: https://datenschutz-hamburg.de/news/meta-starts-ai-training-with-personal-data.



5.4 Copyright Challenges

KEY MESSAGES

- The AI revolution has impacted the copyright landscape, and several key issues require clarification in the near future. These range from the proper application of the text and data mining (TDM) exception to the training process of GenAI models, to liability concerns in case AI-generated outputs might infringe on third-party copyrights linked to content used in the training process, and if so, who would be liable for such infringements.
- There is a need to find harmonised approaches to reserving rights and to avoiding the current lack of consistency being exacerbated by differing opinions of courts in EU Member States. Intensified efforts towards standardisation are essential for creating a unified and predictable copyright framework, benefiting all stakeholders. This need is further underscored by the fact that existing technologies, such as robots.txt, which is widely used in web servers, may not be well-suited for GenAI applications.
- Solutions must be sought to fairly compensate creators whose works are used in the AI training process.

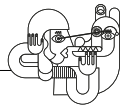
Artificial intelligence has introduced considerable challenges to the area of intellectual property. These challenges are pressing as it is crucial to find a balance between two important goals: on the one hand, protecting the IP rights of creators, concerned about the unauthorised use of their work or the lack of compensation thereof, and on the other hand, facilitating rapid innovation by ensuring AI developers have access to the content necessary for training their models in a timely, yet complaint way. The challenges, however, extend beyond this conflict, arising at various stages throughout the life cycle of an AI model, as explored in the following paragraphs.

1. From the input perspective, the most important question is that of the training of AI on materials that are protected by copyright. To access content, AI solution providers need to resort to activities such as data scraping, web scraping, web crawling and further to data mining. If there has ever been a doubt, the AI Act clarifies that "Text and data mining techniques may be used extensively in this context for the retrieval and analysis of such content"[325] – i.e. development and training of AI models.

Under the Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market ("DSM Directive"), text and data mining ("TDM") is defined 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". All these activities inherently involve the reproduction of text and data, which is an exclusive right of the rightholders.

The DSM Directive provides for two exceptions that permit TDM activities on protected works. Article 3 provides for an exception that allows TDM for scientific research purposes on lawfully accessed works, including the verification of research results, without restrictions. The other exception detailed in Article 4 allows for the extraction and reproduction of lawfully accessed works for TDM purposes if rightholders have not expressly reserved such activities in an appropriate manner. However, the exception outlined in Article 4 introduces complexity, as it is subject to rightsholders' ability to opt out of having their copyright-protected content used, provided they do so in an appropriate manner, such as via machine-readable means for content made publicly available online. The definition of "machine-readable means" appears challenging, particularly as Recital 18 of the DSM Directive seems to leave room for interpretation on whether rights can be reserved through terms

325. Recital 105 in the AI Act.



and conditions alone or if both metadata and terms need to be machine-readable.

Currently, the most recognised machine-readable method is the Robot Exclusion Protocol (robots.txt). However, robots.txt presents several drawbacks as it does not allow "rightholders to indicate that their opt-out applies to a particular class of uses such as no-tdm or no-generative-ai". Efforts are underway to address this limitation, including the work of the IETF's AI Preference (AIPREF) working group,[326] which is developing a common vocabulary to enable more fine-grained expressions of opt-out preferences in protocols like robots.txt. In light of these challenges, it has been suggested that compliance policies should differentiate between at least two distinct opt-out mechanisms: a comprehensive TDM opt-out ("no-tdm") and a specific opt-out from generative AI training ("no-generative-ai"), which would apply exclusively to the use of works for training a subset of AI models described in Recital 105 of the AI Act.

Currently, a large number of data providers prohibit the use of their content, whether specifically for AI training or more generally "for any purposes"[327] through the terms of service on their websites. The extent to which such means is acceptable as a valid opt out under the DSM Directive is not yet supported by consolidated jurisprudence. A German decision[328] stated, though without taking a binding stance, that "a reservation of use drawn up solely in 'natural language' (unlike the presumably predominant view in legal literature" can be regarded as 'machine-readable; however such reservation must be assessed in the light of "the technical development existing at the relevant time of use of the work."

In a ruling from October 2024, the District Court of Amsterdam[329] also had to consider whether the rightholders (publishers) had opted-out of TDM in a machine-readable form. The defendant (a commercial news aggregator) argued that the prohibition on automatic searching was in this case limited to specifically designated AI bots such as GPTBot, ChatGPT-User, CCBOT, and anthropic-ai, an allegation that the claimants failed to refute. The wording of the judgment is somewhat ambiguous, leaving it unclear whether the ruling merely concerned the evidence presented, or whether the court was deciding on the substance what constitutes a proper reservation of rights.

In a recent US court ruling,[330] it was determined that the fair use defence, which may be considered the common law equivalent of the exceptions and limitations in EU copyright law, cannot be used for training an AI model. Although the relevance of this case to the current discussion is debatable as it was held that the AI in question "is not generative AI (AI that writes new content itself)", it will nevertheless be interesting to monitor how this ruling could influence AI providers in the US, especially those engaged in training generative AI models. In an ongoing case[331] started by the New York Times ("NYT") against OpenAI for the unauthorised use of NYT content during model training, OpenAI also relies on the fair use defence.

326. IETF AI Preferences Working group charter: https://datatracker.ietf.org/wg/aipref/about/

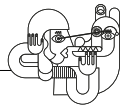
327. For instance, the terms of EBSCO provide that "You agree not to use (or attempt to use) any robot, spider or other automatic device, process or means to access the Website for any purpose, including monitoring or copying any of the material on the Website and not to conduct any systematic or automated data collection activities (including without limitation scraping, data mining, data extraction and data harvesting) on or in relation to the Website without EBSCO Information Services' express written consent", available at https://more.ebsco.com/website-terms-of-use.html, accessed at 21 April 2025,

328. The decision suffers from several shortcomings. For a more complete analysis please refer to this article available here: https://ipkitten.blogspot.com/2024/10/the-german-laion-decision-problematic.html.

329. Decision of the District Court of Amsterdam in case number: C/13/737170 / HA ZA 23-690, available at: ECLI:NL:RBAMS:2024:6563, District Court of Amsterdam, C/13/737170 / HA ZA 23-690,

330. Decision of the District Court of Delaware in Case 1:20-cv-00613-SB available at: https://www.ded.uscourts.gov/sites/ded/files/opinions/20-613_5.pdf.

331. Case no. Case 1:23-cv-11195 before the Sothern District of New York. The complaint is available at: https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf.



2. From the perspective of AI-generated outputs, the primary issue is whether these outputs qualify for copyright protection and if they do, who owns the copyright to the generated work. Another concern is whether AI-generated outputs might infringe on third-party copyrights related to the content used in the training process, and if so, who is liable for such infringement.

Regarding copyright protection, it is reasonable to argue that if an AI-generated output is influenced by the user's interaction with the AI tool through precise instructions and creative choices that reflect a personal touch, it can be considered original and thus eligible for copyright protection. However, the creator must demonstrate the prompting process. In a ruling[332] of the Prague City Court, the judge denied copyright protection for an image because the creator failed to prove authorship with concrete evidence, relying solely on a personal statement.

In the US, decisions in cases such as Théâtre D'opéra Spatial[333] and A Single Piece of American Cheese[334] highlight initial doubts in the Copyright Office's handling of AI-generated content. However, the recent report on Copyright and Artificial Intelligence[335] issued in January 2025 by the US Copyright Office concludes that "the use of AI tools to assist rather than replace human creativity does not impact the availability of copyright protection for the output".

In China, an early decision[336] on the copyrightability of AI outputs ruled that the plaintiff's work qualified for copyright protection due to the "intellectual input" of the creator. The "adjustment and modification process reflects the plaintiff's aesthetic choice and personalized judgment," rendering the work original.

As concerns authorship, traditional IP laws are based on human authorship, which creates ambiguity when an AI system generates content. In line with the case-law of the CJEU, when a work created with an AI tool is deemed 'original' and thus protected by copyright, the copyright will belong to the author who prompted the AI tool;[337] otherwise, there cannot be any copyright protection.

When AI outputs infringe on copyrighted content, this typically results from a process known as "memorization" during model training, where the model learns the data too precisely, reproducing them verbatim. This reproduction becomes an issue if the AI provider's rights to use the materials were limited. For instance, if the AI provider relied on a copyright exception such as TDM, which allows data use solely for text and data mining, subsequent reproduction or communication to the public through output generation are not covered by the TDM exception.

The subsequent question is who is liable in cases of copyright infringement. Users and AI providers can both be held liable depending on the concrete rights that have been infringed and the context of the use of the infringing content.

Thus far, most cases have targeted AI providers. In the US,[338] a series of cases were started from the end of 2023 and early 2024 by publishers such as the New York Times or representatives of

332. Decision of the Prague City Court in case 10 C 13/2023-16 available at: 108cad3e-d9e8-454f-bfac-d58e1253c83a.

333. Decision of the Copyright Review Board in case SR # 1-11743923581 available at https://www.copyright.gov/rulings-filings/review-board/docs/Theatre-Dopera-Spatial.pdf.

334. The work is registered as shown here: https://publicrecords.copyright.gov/detailed-record/37990563.

335. Copyright and Artificial Intelligence Part 2: Copyrightability available at: https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf.

336. Decision of the Beijing Internet Cour in case (2023) Jing 0491 Min Chu No 11279 available at BeijingInternetCourtCivilJudgment112792023.pdf.

337. For an analysis of the notion of 'authorship' under EU law, you may refer to the following publication: Johannes Fritz, The notion of 'authorship' under EU law—who can be an author and what makes one an author? An analysis of the legislative framework and case law, *Journal of Intellectual Property Law & Practice*, Volume 19, Issue 7, July 2024, Pages 552–556, https://doi.org/10.1093/jiplp/jpae022.

338. For an overview of US cases, you may refer to the following article: https://www.bakerlaw.com/services/artificial-intelligence-ai/case-tracker-artificial-intelligence-copyrights-and-class-actions/.

creators such as Getty Images or Concord Music against AI providers such as OpenAI, Stability AI or Anthropic.[339] The end of 2024 has seen similar cases in Europe. In Germany, GEMA, a collective rights management organisation, announced that it was suing OpenAI and SunoAI for copyright infringement.[340] In France, the Syndicat national de l'édition (SNE), the Société des Gens de Lettres (SGDL) and the Syndicat national des auteurs et des compositeurs (SNAC) announced that they started proceedings against META for the unlawful training of its AI models.[341]

In the UK, Getty Images and others have brought an action for copyright infringement, among other charges, before the High Court of Justice against Stability AI. The case highlights a representation matter as claimants filed the lawsuit for over 50,000 copyright holders with exclusive licences to Getty Images Group, prompting the judge to ask the parties to resolve "serious case management issues" before the trial in June 2025.[342]

In response to the growing body of case law, particularly in the US, several agreements have been initiated by providers like OpenAI with numerous publishers. These agreements are meant not only to tackle the use of such content for training purposes, but also the reproduction of excerpts through outputs, "with clear citations and direct links to original sources".[343] In a reply[344] provided by the Centre for Regulation of the Creative Economy based at the University of Glasgow ("CREATe") in February 2025 to the UK Government's Consultation on Copyright and AI,[345] CREATe identified 83 commercial agreements between content providers and AI developers, 68% of which were concluded in respect of news and news media content. Licensing, and particularly collective licensing and bargaining[346] have been identified as potential, albeit challenging, solutions to address the issue of fair remuneration for rightholders.

As it seems, GenAI has altered the copyright landscape, and several key issues require clarification in the near future, among these, the applicability of TDM and proper reservation of rights, compensation of rightholders and copyright liability. To this end, effective solutions will most likely arise from collaboration among various stakeholders.

5.5 Horizontal Data Legislation

KEY MESSAGES

- The European Strategy for Data aims to enhance data availability and cloud infrastructure to support AI and GenAI applications, establishing Common European Data Spaces for secure and trustworthy data sharing.
- Key legislative measures, including the Data Governance Act and the Data Act, focus on improving data accessibility and reuse, especially for AI and GenAI, with the upcoming Data Union Strategy and Data Labs initiatives aiming to improve data quality and organisation to maximise AI potential.

339. You may check details of the cases at the following links: Getty Images v. Stability AI, New York Times v. OpenAI, Concord Music Group, Inc. v. Anthropic https://www.courtlistener.com/docket/68117049/the-new-york-times-company-v-microsoft-corporation/

340. https://www.gema.de/en/news/ai-and-music/ai-lawsuit.

341. https://www.sne.fr/actu/unis-auteurs-et-editeurs-assignent-meta-pour-imposer-le-respect-du-droit-dauteur-aux-developpeurs-doutils-dintelligence-artificielle-generative/.

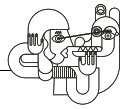
342. Decision of the High Court of Justice in case IL-2023-000007 available at: https://www.judiciary.uk/judgments/getty-images-and-others-v-stability-ai/.

343. https://openai.com/index/partnering-with-axios-expands-openai-work-with-the-news-industry/.

344. Kretschmer, M., Meletti, B., Bently, L., Cifrodelli, G., Eben, M., Erickson, K., Iramina, A., Li, Z., McDonagh, L., Perot, E., Porangaba, L., & Thomas, A. (2025). Copyright and AI: Response by the CREATe Centre to the UK Government's Consultation. CREATe. https://doi.org/10.5281/zenodo.14931964.

345. The Consultations is available here: https://www.gov.uk/government/consultations/copyright-and-artificial-intelligence/copyright-and-artificial-intelligence.

346. Quintais, João Pedro, Generative AI, Copyright and the AI Act (January 30, 2025). Computer Law & Security Review, Volume 56, 2025, 106107, https://doi.org/10.1016/j.clsr.2025.106107. Available at SSRN: https://ssrn.com/abstract=4912701.



As discussed in Section 1.3, no GenAI application would be possible without data, which should be of suitable quality and quantity. This Section provides an overview of the broad EU policy context regulating data sharing, focusing on the main EU policy initiatives and related legislation and their effects on, and implications for, GenAI.

Released in 2020 together with the White Paper on AI,[347] the European Strategy for Data[348] set an ambitious policy agenda aimed to achieve the full potential of data-driven innovation in the EU, recognising, among others, the needs to improve data availability and the uptake of cloud computing infrastructures to fuel AI applications. The strategy also envisioned the establishment of Common European data spaces as decentralised, federated and sovereign environments for sharing data across actors and sectors, in a secure, reliable and trustworthy manner in full alignment with European rules and values. Funded by the Digital Europe Programme with additional contributions from Horizon Europe, data spaces in 14 thematic domains have been conceptualised and are currently (Summer 2025) undergoing their real-world deployment.[349] Data spaces hold strong synergies with GenAI applications as they act as mutual enablers. This interplay is explored in a recent white paper[350] from the Data Spaces Support Centre, the project tasked to coordinate the development of Common European Data Spaces.

IMPLEMENTING THE VISION

To help achieve the ambition of the European Strategy for Data, a set of horizontal legal provisions on data sharing matters was put forward. First, this includes the Data Governance Act,[351] applicable from September 2023 and implementing cross-cutting measures to increase data availability and overcome technical obstacles to the reuse of data, including for GenAI applications. To this end, the DGA has created a specific legal regime applicable to a new class of players - data intermediaries, that provide a framework of governance and trust between data providers and data seekers. The Data Act,[352] applicable from September 2025, aims to make more data accessible for reuse by setting measures on, among others, the reuse of data generated from Internet of Things (IoT) devices. Moreover, the Implementing Act on High-Value Datasets,[353] applicable from June 2024, implements the Open Data Directive[354] by defining a list of datasets from public sector organisations (see also Section 6.5), whose reuse - especially by SMEs - holds the potential to generate high economic benefits to the EU economy and society, including through AI and GenAI applications. Such datasets shall be made available free of charge, under open access licences and accessible through Application Programming Interfaces (APIs).

THE WAY FORWARD

Given the amount of new data-related legislation described above, under the European Commission 2024–2029 the focus is placed on the implementation of such legal instruments together with their potential simplification. A

347. White Paper on Artificial Intelligence – A European approach to excellence and trust. COM(2020) 65 final

348. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – A European strategy for data. COM(2020) 66 final

349. https://digital-strategy.ec.europa.eu/en/policies/data-spaces

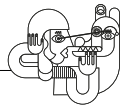
350. The new "Generative AI and Data Spaces" white paper of the Strategic Stakeholder Forum is now available

351. Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act): https://eur-lex.europa.eu/eli/reg/2022/868/oj

352. 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): https://eur-lex.europa.eu/eli/reg/2023/2854/oj

353. Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use: http://data.europa.eu/eli/reg_impl/2023/138/oj

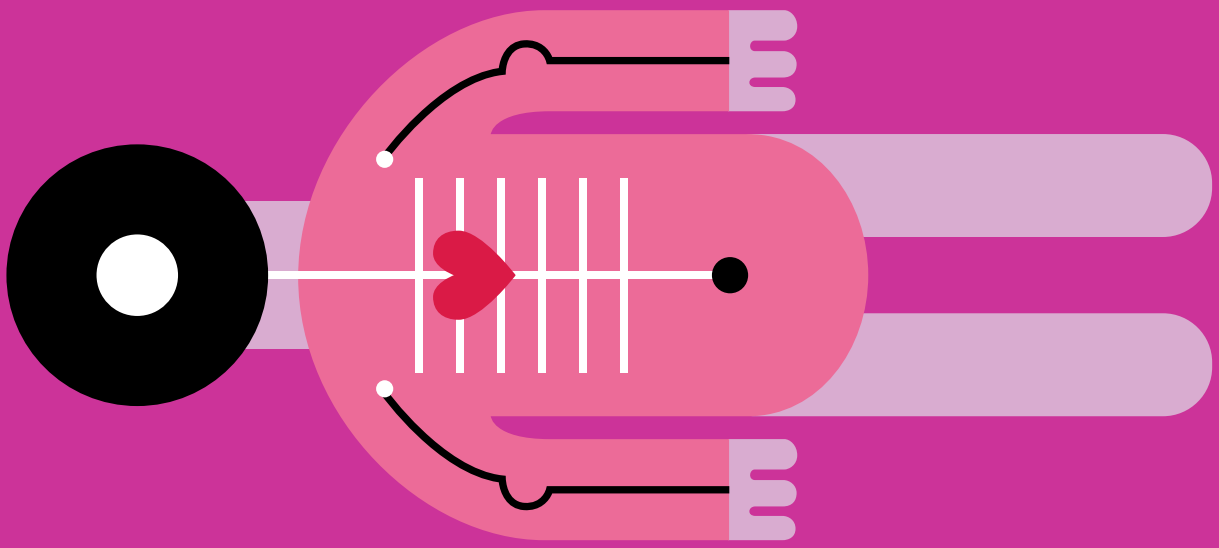
354. Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast): http://data.europa.eu/eli/dir/2019/1024/oj



Data Union Strategy, to be published in the second half of 2025, will chart the path towards improving access to reliable, high-quality and well-organised data to unlock the full power of AI applications and make the most out of Europe's data ecosystem. This will be reinforced by Data Labs - a new tool to enable the provision, pooling, and secure sharing of high-quality data, acting as central hubs that bring together and organise data in relation to AI Factories operating within the same sector. Additionally, Data Labs will align with and connect to Common European Data Spaces, which, in turn, transform fragmented data sources into federated, high-quality datasets to support the development of GenAI models.[355] ■

355. AI Continent Action Plan

6
DEEP DIVES



This chapter presents specific deep dives that illustrate the transformative impact of GenAI across various sectors. It begins with healthcare, examining how GenAI can provide many positive socioeconomic effects on the whole field if AI developers, clinicians, researchers and regulators were able to balance GenAI's many possibilities versus its pitfalls and risks. The educational sector is explored, together with the impact of GenAI in science, cybersecurity, and the public sector, showcasing its diverse applications and the unique challenges each domain faces. Key considerations include the responsible and effective integration of GenAI technologies to maximise their societal benefits.

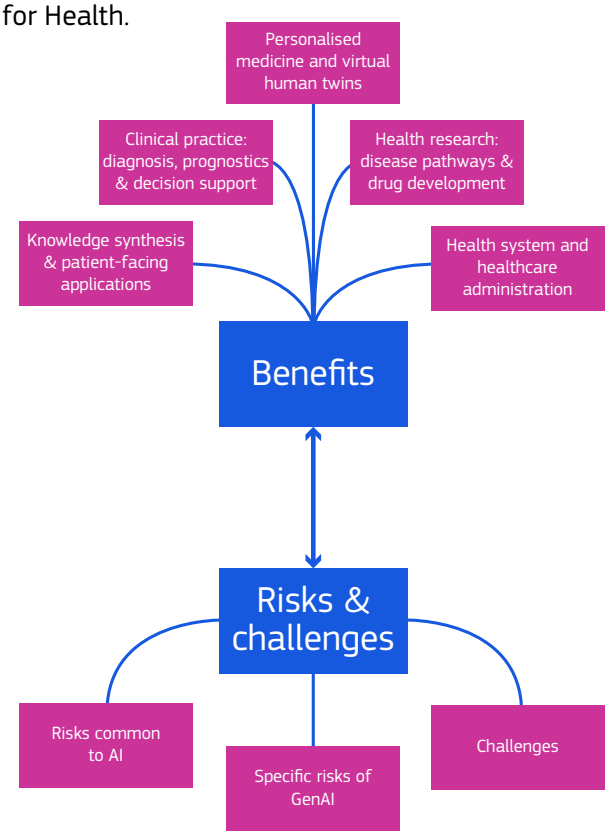
6.1 Healthcare

KEY MESSAGES

- GenAI could provide many positive socioeconomic impacts in health, e.g. by creating and exploiting electronic health records (EHRs), accelerating drug development, enhancing personalised medicine, fostering prevention and early diagnosis of diseases, enhancing healthcare efficiency, addressing health inequities and empowering patients.
- The extent to which the impact will materialise will depend on whether AI developers, clinicians, researchers and regulators will be able to balance GenAI's many possibilities versus risks. These include data biases, propagation of health inequities, deskilling of clinicians and the deterioration of the human dimension of care through automation bias and overreliance.
- Addressing this tension is not only a regulatory question[356] but also a question of responsible use within healthcare workflows.[357] Further, a broad rollout in health systems would require significant investments into decentralised IT infrastructure, colliding with a chronic state of underfunding.

Responsible development and use of GenAI in health and care will require targeted gradual uptake, open debate,[358] multidisciplinary collaboration, and pragmatic laws alongside ethical guidelines.[359] Building trust among clinicians, patients, and healthcare organisations is paramount for widespread adoption.

Figure 18. Benefits, risks and challenges of GenAI for Health.



Source: JRC elaboration.

356. AI and GenAI in healthcare fall under an ecosystem of EU legislations and policies, inter alia: the Medical devices and In vitro medical devices Regulations concerning software with a medical purpose. The AI Act concerning general requirements based on risks. The General data protection Regulation concerning personal including sensitive health data. The Data Act concerning data sharing and the European Health Data Space, making provisions for primary and secondary use of health data, notably also for training AI and GenAI systems.

357. OECD (2024) collective action for responsible AI in health. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/01/collective-action-for-responsible-ai-in-health_9a65136f/f2050177-en.pdf

358. Reddy S (2024) Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. Implementation Sci 19, 27. Online: https://doi.org/10.1186/s13012-024-01357-9

359. Upcoming JRC publication: Griesinger CG, Reina V, Panidis D, Chassaigne H (2025) Towards an evidence pathway for operationalizing trustworthy AI in health: an ontology to bridge the gap between ethical principles and fundamental concepts. Submitted.

GENERATIVE AI: A POSSIBILITY TO RESPOND TO GLOBAL HEALTHCARE CHALLENGES

In healthcare, traditional discriminative AI focuses on learning a function to map numerical inputs, such as patient data or medical image features, to specific outputs for tasks like predicting diseases (e.g., cancer from a CT scan) or classifying medical images (e.g. identifying fractures in X-rays). It excels at analysing existing data to make predictions or classifications, for instance, by identifying patterns in electronic health records to forecast patient risk for a particular condition.

In contrast, GenAI in healthcare learns the underlying probability distribution and structure within health data itself. This enables it to generate novel, statistically similar data points. Examples include creating synthetic medical images (like MRIs or X-rays) for training purposes, especially for rare diseases, generating realistic medical reports or literature summaries, designing new drug candidates by producing novel chemical structures, or synthesising omics data for research. Multi-modal GenAI (see Section 2.3) expands on this by processing diverse inputs, such as a patient's medical history (text), lab results (numerical), and imaging scans (images), to generate varied outputs, like a comprehensive diagnostic report or a personalised treatment plan, and can even translate between these data types, for example, by generating a textual radiology report directly from an X-ray image.

GenAI arrives at a critical juncture of healthcare: health systems globally face unprecedented demographic pressures from aging populations[360] and a rising burden of non-communicable diseases.[361] Workforce shortages endanger existing care models.[362] Estimates predict shortages of approximately 4.1 million healthcare professionals within EU countries by 2030.[363] This coincides with healthcare inequities, especially in rural versus urban areas.[364] In the following, we will detail potential benefits, risks and challenges in the use of GenAI for health and care, as presented in Figure 18.

APPLICATIONS OF GENERATIVE AI IN HEALTHCARE

The integration of GenAI into healthcare concerns multiple applications. We structure these in four areas: knowledge synthesis and patient-facing applications, clinical practice, personalised medicine and health research.

→ **Knowledge synthesis and patient-facing applications:** GenAI systems can act to reduce the cognitive overload[365] of healthcare professionals, who – often operating under pressure - frequently resort to heuristic decision-making.[366] GenAI may support differential diagnoses, suggest diagnostic tests or treatment protocols.[367] Multimodal GenAI can simultaneously process and synthesise vast quantities of medical information at high speed,

360. Iuga, I. C., Nerişanu, R. A., & Iuga, H. (2024). The impact of healthcare system quality and economic factors on the older adult population: A health economics perspective. Frontiers in Public Health, 12, 1454699.

361. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)00685-8/fulltext

362. Liu, J. X. & others. (2017). Global health workforce labor market projections for 2030. Human Resources for Health, 15, 1–12.

363. https://www.europarl.europa.eu/news/en/agenda/briefing/2025-02-10/14/healthcare-sector-addressing-labour-shortages-and-working-conditions

364. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/11/health-at-a-glance-europe-2024_bb301b77/b3704e14-en.pdf

365. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.

366. Whelehan DF et al. (2020) Medicine and heuristics: cognitive biases and medical decision-making. Ir J Med Sci. 2020 Nov;189(4):1477-1484. Online: doi: 10.1007/s11845-020-02235-1

367. Li, Y.-H. & others. (2024). Innovation and challenges of artificial intelligence technology in personalized healthcare. Scientific Reports, 14(1), 18994.

generating novel medical insights.[368] [369] [370] [371] [372] [373] Thus GenAI can reduce cognitive overload of healthcare professionals, augmenting human judgement by compiling insights based on millions of patient records and billions of literature data points.[374] GenAI systems can process and synthesize vast quantities of medical information at high speed, potentially reducing diagnostic delays and treatment errors.[375] LLMs can also empower patients by creating coherent narratives from fragmented medical information, improving treatment adherence and facilitating informed consent.[376] GenAI applications include conversational agents ("chatbots") that aid in preliminary assessments, health education, preparedness and fast response to public health threats,[377] and patient support. GenAI may facilitate healthcare delivery by aiding with the interpretation of electronic health records (EHRs).[378] GenAI could improve healthcare accessibility by providing medical summaries to enable first-line interventions by local healthcare professionals in rural areas of the global south.[379] Clinicians spend a significant amount of time on routine tasks, e.g. administrative documents, clinical documentation, medical coding, billing, patient scheduling and communication or workflow management. Automating these would reduce operational costs and free clinicians' time to focus on patient interactions. When paired with retrieval techniques like Retrieval Augmented Generation (RAG), GenAI can achieve good accuracy and reproducibility on most knowledge synthesis health tasks.[380]

→ **Clinical practice - diagnostics, prognostics and decision support:** There are numerous GenAI applications in clinical practice: medical imaging analysis[381] for diagnostics and prognostics,[382] clinical decision support systems and robotic surgical systems. GenAI can create synthetic but anatomically plausible medical images

368. Esteva, A. & others. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118.

369. Nazi, Z. A., & Peng, W. (2024). Large language models in healthcare and medical domain: A review. Informatics, 11(3).

370. Rajpurkar, P. & others. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Medicine, 15(11), e1002686.

371. Thirunavukarasu, A. J. & others. (2023). Large language models in medicine. Nature Medicine, 29(8), 1930–1940.

372. Tu, T. & others. (2024). Towards generalist biomedical AI. Nejm Ai, 1(3), AIoa2300138.

373. Tomašev, N. & others. (2019). A clinically applicable approach to continuous prediction of future acute kidney injury. Nature, 572(7767), 116–119.

374. Rajkomar & others, 2018

375. Rajpurkar, P. & others. (2022). AI in health and medicine. Nature Medicine, 28(1), 31–38.

376. Stanceski, Kristian, et al. (2024). "The quality and safety of using generative AI to produce patient-centred discharge instructions." npj Digital Medicine 7.1: 329.

377. S. Consoli, P. Markov, N. I. Stilianakis, L. Bertolini, A. Puertas Gallardo, and M. Ceresa (2024). Epidemic Information Extraction for Event-Based Surveillance using Large Language Models. In X.-S. Yang et al. (Eds.), Proceedings of Ninth International Congress on Information and Communication Technology (ICICT 2024), volume 1011 of Lecture Notes in Networks and Systems, pages 241-252, Springer Nature, Switzerland, doi:10.1007/978-981-97-4581-4_17

378. Yang X et al. (2022) A large language model for electronic health records. npj Digit. Med. 5, 194 (2022). https://doi.org/10.1038/s41746-022-00742-2

379. NITI Aayog (2018) National strategy for artificial intelligence. https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf

380. See JRC report: Ceresa, M; Bertolini, L., Comte, V.; Spadaro N.; Raffael, B.; Toussaint, B.; Consoli, S.; Muñoz Piñeiro A.; Patak, A.; Querci M.; Wiesenthal T. Retrieval Augmented Generation Evaluation for Health documents, Publications Office of the European Union,Luxembourg, JRC138904.

381. These advancements are particularly relevant for complex diseases like cancer. Initiatives such as the European Cancer Imaging Initiative, a flagship initiative under Europe's Beating Cancer Plan (EBCP), aim to foster and exploit innovative imaging, AI solutions and deployment of digital tools to improve cancer diagnosis and treatment across Europe.

382. Jiang, LY et al. (2023) Health system-scale language models are all-purpose prediction engines. Nature 619, 357–362 (2023). https://doi.org/10.1038/s41586-023-06160-y

of various typologies,[383] [384] which might help in tackling the limited access to large, expertly annotated medical image datasets and the high associated costs.[385] Synthetic images cannot replace real-world data for clinical validation but might support both development and testing of new models, e.g. by rebalancing datasets or supplementing scarce sets in case of rare diseases.[386] GenAI can enhance image quality and analysis to facilitate diagnosis[387] and produce "counterfactual scans" that illustrate outcomes under different hypothetical circumstances[388] including alternative treatments. Finally, robotic surgical systems enhanced by GenAI[389] might act as members of the surgery team,[390] augmenting human agency,[391] especially under conditions of fatigue. As an example, the "AI for Public Good" initiative is driving an innovative cancer imaging project for breast and prostate diagnosis, showcasing how AI models can be adapted across diverse settings without sharing sensitive data—safeguarding privacy while improving access, especially in low-income countries.

→ **Personalised medicine (PM):** PM seeks to overcome the traditional "one-size-fits-all" approach by advancing individualised healthcare through tailored risk assessment, prevention and treatment strategies for specific groups of individuals or single patients. PM draws on the integration of diverse data including multi-omics, clinical history, and lifestyle factors and compiles these in patient profiles. Such profiles allow predicting individual treatment responses and adverse events.[392] Virtual Human Twins (VHTs)[393] can support PM by replicating the complex physiology and pathology of individual patients based on the integration of diverse data, including from genomics, imaging, clinical records, and wearable sensors.[394] GenAI with its multimodal generative capabilities will accelerate the development and use of VHTs.[395] Finally, synthetic populations of VHTs can be utilised for health research in the context of in silico trials.

→ **Health research:** GenAI can support various aspects of health research. We distinguish two major strands:

383. Dar, S. U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., & Cukur, T. (2019). Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE Transactions on Medical Imaging, 38(10), 2375–2388.

384. Sandfort, V., Yan, K., Pickhardt, P. J., & Summers, R. M. (2019). Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Scientific Reports, 9(1), 16884.

385. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139–144.

386. Yang, Y., Zhang, H., Gichoya, J. W., Katabi, D., & Ghassemi, M. (2024). The limits of fair medical imaging AI in real-world generalization. Nature Medicine, 30(10), 2838–2848.

387. Pan, S., Wang, T., Qiu, R. L., Axente, M., Chang, C. W., Peng, J., Fei, B., & Yang, X. (2023). 2D medical image synthesis using transformer-based denoising diffusion probabilistic model. Physics in Medicine & Biology, 68(10), 105004.

388. Koohi-Moghadam, M., & Bae, K. T. (2023). Generative AI in medical imaging: Applications, challenges, and ethics. Journal of Medical Systems, 47(1), 94.

389. Schmidgall, Samuel, et al. (2024). "General-purpose foundation models for increased autonomy in robot-assisted surgery." *Nature Machine Intelligence*: 1-9.

390. Marcus, Hani J., et al. (2024). "The IDEAL framework for surgical robotics: development, comparative evaluation and long-term monitoring." *Nature medicine* 30.1: 61-75.

391. Goldberg, Ken, and Gary Guthart. (2024). "Augmented dexterity: How robots can enhance human surgical skills." *Science Robotics* 9.95: eadr5247.

392. In essence, while classical AI helps us understand "what is" or "what will likely be" based on existing data, GenAI empowers us to explore "what could be" by generating new possibilities and simulating complex interactions.

393. The European Virtual Human Twins initiative supports the emergence and adoption of the next generation of VHT solutions in health and care. https://digital-strategy.ec.europa.eu/en/policies/virtual-human-twins

394. See JRC report: Enhancing Digital Health Innovation in the EU with Effective Industrial Strategy Policies - A Focus on Wearable Medical Devices. JRC138798

395. Chiaro, Diletta, et al. (2025). "Generative AI-Empowered Digital Twin: A Comprehensive Survey With Taxonomy." *IEEE Transactions on Industrial Informatics*.

a. Drug discovery, repurposing, development: The pharmaceutical industry has embraced GenAI to tackle lengthy and costly drug development processes.[396] GenAI is used to design novel molecular structures optimised for specific therapeutic targets,[397] thus enhancing efficacy and reducing side effects. It can generate and screen millions of candidate molecules *in silico*, accelerating early-stage drug discovery. GenAI can predict drug-drug interactions and identify potential repurposing opportunities.[398] By generating synthetic clinical trial data that supplement real-world evidence, GenAI may facilitate clinical testing while reducing risks for patients.

b. In silico trials: *In silico* trials use synthetic population models to simulate clinical trials. GenAI can generate virtual patient cohorts with specific demographic and clinical characteristics, enabling the testing of drug efficacy, toxicity, and optimal dosage under diverse scenarios before initiating human trials,[399] addressing ethical challenges of clinical testing in humans. GenAI can simulate treatment effects, predict adverse events, optimise trial designs by identifying ideal patient subgroups, and even generate counterfactual scenarios to aid understanding of treatment mechanisms. The interplay between GenAI and VHTs promises more personalised, efficient pathways for therapeutic development and precision medicine.[400] Challenges such as model validation, predictive accuracy and computational requirements remain to be addressed.

CHALLENGES AND RISKS OF GEN AI IN HEALTHCARE

Risks Specific to Generative AI in Health and Care

Clearly, the ethical, scientific and regulatory challenges and risks applicable to AI systems[401] in general also apply to GenAI. These include patient safety, accountability, transparency and intelligibility of models to ensure failure transparency, traceability and informed patient consent.[402] In addition, GenAI poses specific challenges that still require considerable research. We structure these along three lines:[403] bias and equity, incorrect content (hallucinations), and stochastic echo chamber.

→ **Bias and equity:** GenAI may propagate inadequacies hidden in the training data. These include historical biases (race, gender, social status) or outdated medical concepts, encapsulated in scientific and clinical publications. This might lead to diagnostic errors, inequitable treatment recommendations, and further health

396. Doron, G., Genway, S., Roberts, M., & Jasti, S. (2025). Generative AI: Driving productivity and scientific breakthroughs in pharmaceutical R&D. Drug Discovery Today, 30(1), 104272.

397. Cheng, Y., Gong, Y., Liu, Y., Song, B., & Zou, Q. (2021). Molecular design in drug discovery: A comprehensive review of deep generative models. Briefings in Bioinformatics, 22(6), bbab344.

398. Drug repurposing or repositioning refers to the use of medicinal products for medical indications other than the one(s) for which the product was originally developed and/or marketed.

399. Hamed, Ahmed Abdeen, Tamer E. Fandy, and Xindong Wu. (2024). "Accelerating Complex Disease Treatment Through Network Medicine and GenAI: A Case Study on Drug Repurposing for Breast Cancer." 2024 IEEE International Conference on Medical Artificial Intelligence (MedAI). IEEE.

400. Su, Chengxun, et al. "Optimizing metabolic health with digital twins." npj Aging 11.1 (2025): 20.

401. S. Consoli, D. Reforgiato Recupero, and M. Petkovic, 2019. Data Science for Healthcare: Methodologies and Applications, Springer Nature, Switzerland, ISBN: 978-3-030-05248-5, doi:10.1007/978-3-030-05249-2

402. Howell MD (2024) Generative artificial intelligence, patient safety and healthcare quality: a review. BMJ Qual Saf. Oct 18;33(11):748-754. Online: doi: 10.1136/bmjqs-2023-016690

403. Griesinger CG, Reina V, Panidis D, Chassaigne H (2025) Towards an evidence pathway for operationalizing trustworthy AI in health: an ontology to bridge the gap between ethical principles and fundamental concepts. Submitted to arXiv

disparities. Addressing bias requires conscious efforts in data collection, algorithmic design, model auditing, and deploying fairness-aware machine learning techniques throughout the GenAI life cycle. In these scenarios, the distinction between Open Weights and Open Source models discussed in Section 1.3 might have an active and relevant role. Carefully measuring the extent to which a model has inherent biases from the training data requires the data to be publicly available. Although this risk is not unique to GenAI, the potentially future role of GenAI cutting through many health aspects poses a particular risk. Notably, most models are currently either trained on too narrow datasets or are evaluated on tasks that now allow gauging their real-world usefulness for health systems.[404] In this context, initiatives like the European Health Data Space (EHDS) Regulation[405] are highly relevant. The EHDS aims to integrate health data of EU citizens from various sources, such as hospitals, into a distributed infrastructure. This system is intended to facilitate data sharing for primary use (between healthcare points) and for secondary use, including research and the training of AI models. By tackling issues of data fragmentation and interoperability, the EHDS could make broader and potentially more representative datasets available, which is a necessary step - though not sufficient on its own - for training less biased GenAI models. Furthermore, this framework may also help address GDPR-related complexities concerning the secondary use of health data (see Section 5.5).

→ **Incorrect content:** GenAI may generate contents that seem, prima facie, plausible but are, on closer inspection, nonsensical or not rooted in true epistemic data[406] – so-called "hallucinations" or "confabulations".[407] Factual relevance of predictions may be improved through various approaches including "Retrieval Augmented Generation"[408] and reinforcement learning.[409] Taken together, caution is required when employing FMs in medicine and healthcare, and appropriate training of healthcare professionals including the limitations and risks of GenAI will be critical.

→ **Echo chamber of probabilistic processes:** contents created by GenAI are ultimately rooted in probabilistic processes of data representations and learned "semantic contexts" embedded in training data and incorporated in billions of parameters within artificial neural networks. There is the risk that the output is to some extent a merely sophisticated "echo chamber" of the data used to train the model and their stochastic connections shaped during machine learning. Paired with automation bias and complacency, this could devalue human medical expertise and creativity, leading to the propagation of care models rooted in specific data and algorithms. Methods that allow quantification of the uncertainty

404. Wornow M et al. (2023) The shaky foundations of large language models and foundation models for electronic health records. npj Digit. Med. 6, 135 (2023). https://doi.org/10.1038/s41746-023-00879-8

405. https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en

406. Rawte V et al. (2023) The Troubling Emergence of Hallucination in Large Language Models - An Extensive Definition, Quantification, and Prescriptive Remediations. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, pages 2541–2573, Singapore. Association for Computational Linguistics. Online: https://aclanthology.org/2023.emnlp-main.155/

407. Sun Y et al. (2024) AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. Humanit Soc Sci Commun 11, 1278.

408. See JRC report: Ceresa, M; Bertolini, L., Comte, V.; Spadaro N.; Raffael, B.; Toussaint, B.; Consoli, S.; Muñoz Piñeiro A.; Patak, A.; Querci M.; Wiesenthal T. Retrieval Augmented Generation Evaluation for Health documents, Publications Office of the European Union, Luxembourg, JRC138904.

409. Roit et al. (2023) Factually consistent summarization via reinforcement learning with textual entailment feedback. Online: http://arxiv.org/abs/2306.00186

outputs should be routinely used in health applications.[410]

Challenges for the development and deployment of GenAI include data and privacy, infrastructure, interoperability and cybersecurity aspects:

→ **Data and privacy:** Both performance and reliability of GenAI models depend on large-scale, high-quality and diverse multi-modal datasets – their **availability** however poses a major challenge as medical data are often fragmented and lack standardisation.[411] Accessing sufficient data, particularly for rare diseases or underrepresented populations, remains a bottleneck. While synthetic data (see Section 1.3) generated by GenAI can augment datasets, questions remain about fidelity, bias amplification and overfitting. Re-identification of data remains a persistent problem. Thus, more research into privacy-preserving techniques is required.[412]

→ **Infrastructure, interoperability and cybersecurity:** Many healthcare facilities currently lack the requisite IT **infrastructure** for GenAI implementation. GenAI models demand significant computational resources, data storage and high-bandwidth networking, necessitating local investments in hardware and cloud computing platforms. Some infrastructures are under construction at the European level, e.g. EuroHPC and the AI Factories initiatives. There also remains a tension between the momentum towards centralisation of infrastructure for a variety of reasons and the need to use federated/distributed learning to protect data privacy. **Interoperability** poses a substantial challenge: health data reside in multiple disconnected systems such as electronic health records (EHRs), picture archiving and communication systems (PACS) or laboratory information systems (LIS). These often use proprietary formats or outdated standards, hindering data aggregation for model training and complicating GenAI development across different settings and/or clinical workflows. Like other digital technologies, GenAI is not immune to **cybersecurity** issues. Moreover, its application in the medical field introduces additional challenges, linked to sensitive patient health information, exacerbating the complexity of security concerns. GenAI may introduce new vulnerabilities that attackers may exploit to compromise systems, spread inaccurate information, influence user behaviour or hold extracted personal data for ransom.[413] [414] Threat actors may employ a diverse range of cybersecurity attacks against GenAI, including model inversion attacks, compromising user privacy, as well as data poisoning attacks.[415] Prompt injection attacks can disrupt the model's normal functioning[416] and potentially lead to safety risks. Ultimately, the implementation of advanced detection

410. Hulsman, Roel, et al. (2024). "Conformal Risk Control for Pulmonary Nodule Detection." *arXiv preprint arXiv:2412.20167*.

411. M. van Hartskamp, S. Consoli, W. Verhaegh, M. Petkovic, and A. van de Stolpe (2019). Artificial Intelligence in Clinical Health Care Applications: Viewpoint. Journal of Medical Internet Research, 21(4):e12100, doi:10.2196/12100

412. The European Health Data Space (EHDS) aims to address these problems.

413. Teo, Z. L., Quek, C. W. N., Wong, J. L. Y., & Ting, D. S. W. (2024). Cybersecurity in the generative artificial intelligence era. Asia-Pacific Journal of Ophthalmology, 13(4). https://doi.org/10.1016/j.apjo.2024.100091

414. Reina V & Griesinger CB (2024) Cyber security in the health and medicine sector: a study on available evidence of patient health consequences resulting from cyber incidents in healthcare settings. European Commission: Joint Research Centre. https://publications.jrc.ec.europa.eu/repository/handle/JRC138692

415. Das, A., Tariq, A., Batalini, F., & others. (2024). Exposing Vulnerabilities in Clinical LLMs Through Data Poisoning Attacks: Case Study in Breast Cancer.

416. Clusmann, Jan, et al. (2025). "Prompt injection attacks on vision language models in oncology." *Nature Communications* 16.1: 1239.

systems and proactive mitigation measures is crucial to detect and contain cyber threats.[417]

Building trust among clinicians, patients, and healthcare organisations is essential to address such challenges and foster uptake. This involves addressing clinician concerns about accuracy, reliability, workflow integration, deskilling, and liability, as well as patient worries regarding data privacy, biased recommendations, and dehumanised care. Overcoming these challenges necessitates robust technical validation, strong ethical principles, regulatory clarity, user-centric design, transparent communication, and demonstrating GenAI's value to all stakeholders.

6.2 Impact on learning and teaching

KEY MESSAGES

- The integration of GenAI into educational systems is transforming the landscape of learning, teaching and assessment. This technological innovation has the potential to shape, or even disrupt education and training, and its impact is being felt by various stakeholders.
- To ensure that GenAI is used effectively and responsibly, policymakers, educators, and students must work together to develop the competences and policies required to support its integration into educational systems.

THE IMPACT OF GENAI IN EDUCATION AND TRAINING

GenAI is increasingly being used in educational settings, resulting in substantial changes in teaching and learning with far-reaching implications for stakeholders, including policymakers, educational institutions leadership, educators, and students. Initially, concerns surrounding potential misuse led to restrictions on GenAI in various institutions; however, the discourse quickly shifted towards exploring its potential to enhance learning and teaching outcomes. As GenAI systems become more capable, it also became evident that education systems would need to reassess the competences that would be required in the coming years.[418] Despite its growing influence, there is a pressing need for rigorous and empirical evidence to further understand the impact of GenAI on educational practices, particularly with regard to whether it can effectively improve teaching and learning[419] and also its implications for assessment In this section we examine results from a number of JRC studies looking at emerging trends and consequences of GenAI's integration in education, as well as the use and perception of GenAI by different educational stakeholders and the different requisites for educators to effectively leverage this technology in their pedagogical practices.[420] The different studies found that GenAI is seen as an opportunity for teaching and learning enhancement, but it requires careful implementation, ongoing professional development, and the development of AI literacy to ensure it is used effectively and responsibly.

The ethical guidelines on the use of AI and data in teaching and learning for educators[421] (2022) provide guidance to integrate ethical considerations and requirements based on examples and practical questions. Targeting

417. Vassilev, A., Oprea, A., Fordyce, A., Anderson, H., Davies, X., & Hamin, M. (2025). Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations (NIST Trustworthy and Responsible AI No. NIST AI 100-2e2025). National Institute of Standards and Technology.

418. OECD (2025), "What should teachers teach and students learn in a future of powerful AI?", *OECD Education Spotlights*, No. 20, OECD Publishing, Paris, https://doi.org/10.1787/ca56c7d6-en.

419. Sallai, D., Cardoso-Silva, J., Barreto, M., Panero, F., Berrada, G. and Luxmoore, S. (2024) 'Approach Generative AI Tools Proactively or Risk Bypassing the Learning Process in Higher Education', LSE Public Policy Review, 3(3), p. 7. Available at: https://doi.org/10.31389/lseppr.108.

420. Forthcoming JRC Policy Brief: Uses and perceptions of Generative AI in secondary education across five Member States

421. https://data.europa.eu/doi/10.2766/153756

teachers and educational staff, mainly at primary and secondary level, they are useful for the wider educational community and stakeholders involved in digital education. Their revision is ongoing to consider new technological developments, such as Gen AI for pedagogical practices and ensure an enhanced practical approach. Also, the 2025 Erasmus+ Forward looking projects call[422] looks for large-scale projects promoting the ethical and effective use of GenAI systems in education and training.

The 2030 Roadmap on the future of digital education and skills will foster the strategic and ethical uptake of AI in education, including through support and capacity building for teachers and education institutions and will promote the development of AI literacy from primary and secondary education.

AI Literacy Framework

The European Commission and the OECD, with the support of Code.org and a pool of leading international experts, are currently developing an AI Literacy Framework. The framework will outline the knowledge, skills and attitudes that will adequately prepare students in primary and secondary education. The framework will delineate how to deepen learning on the use of AI tools, how to co-create with them, as well as how to reflect on responsible and ethical use in subjects that are essential for AI Literacy, such as statistics, social science and computer science. The framework will be finalised in early 2026 after extensive stakeholder consultations. Based on this framework, the first assessment of AI literacy in the OECD Programme for International Student Assessment (PISA) will be developed. This will support the EU's goals to promote quality and inclusive digital education and skills.

422. https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/ERASMUS-EDU-2025-PI-FORWARD-DIGITAL-AI

EMERGING TRENDS AND TECHNOLOGIES IN EDUCATION

The JRC has conducted foresight research on emerging technologies, including GenAI, that have significant potential to shape the future of education by redefining educational practices, processes and organisations.[423] GenAI applications, such as video captioning, translation, speech-to-text, and text-to-animation, offer numerous opportunities for pedagogic purposes. These applications can generate video or text learning materials from existing content, enabling teachers to create personalised learning experiences for their students. Also, AI systems can act as partners, co-designers, and Socratic opponents or motivators, aiding thought development. However, pure language models are known for generating convincing but incorrect text. Despite scalability uncertainties, linking language models to existing knowledge sources could enhance trustworthiness, vital for education and learning.

Integrating AI with human learning processes makes agency central in education. AI could enable new agentic distribution forms in education, involving students, teachers, and parents with AI. This requires expanding individualistic views on agency and competence to include social and technical resources underpinning agentic action.[424]

The use of GenAI in education is not limited to teaching and learning, it also has the potential to transform educational administration and policy. The growing role of technology in education and the extensive use of digital platforms are increasing dependence on global players, raising concerns about data privacy and digital

423. Tuomi, I., Cachia, R. and Villar Onrubia, D., On the Futures of Technology in Education: Emerging Trends and Policy Implications, Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/079734, JRC134308.

424. Tuomi, I. (2022). Artificial intelligence, 21st century competences, and socio-emotional learning in education: More than high-risk? European Journal of Education, 57(4), 601–619. https://doi.org/10.1111/ejed.12531

sovereignty. On a systemic level, discussions continue on the potentially conflicting interests of commercial stakeholders and educators, and understanding educators and learners interests and needs remains crucial.[425] [426] [427] [428] [429]

In conclusion, the integration of emerging technologies, particularly generative AI, into educational practices holds transformative potential for redefining the landscape of learning and teaching. This evolution demands a nuanced understanding of the interplay between technological advancements and educational policies, ensuring that innovations are harnessed to enhance learning outcomes sustainably and equitably. As AI systems become integral to educational environments, they offer opportunities for personalised learning, efficient administration, and the democratisation of educational resources. However, this shift also necessitates addressing ethical considerations, governance issues, and the alignment of AI with educational values. By fostering collaboration among policymakers, educators, and technologists, the education sector can navigate these challenges, leveraging AI to support a future where education is more accessible, inclusive, and effective for all learners.

425. Blikstein, P., Zheng, Y., & Zhou, K. Z. (2022). Ceci n'est pas une école: The discourses of artificial intelligence in education through the lens of semiotic analytics. European Journal of Education, 57(4), 571–583.

426. Selwyn, N. (2022). The future of AI and education: Some cautionary notes. European Journal of Education, 57(4), 620–631. https://doi.org/10.1111/ejed.12532

427. Selwyn, N. (2023). Constructive Criticism? Working with (Rather than Against) the AIED Backlash. International Journal of Artificial Intelligence in Education. https://doi.org/10.1007/s40593-023-00344-3

428. Williamson, B. (2021, May 28). Google's plans to bring AI to education make its dominance in classrooms more alarming. Fast Company. https://www.fastcompany.com/90641049/google-education-classroom-ai

429. Williamson, B., Gulson, K. N., Perrotta, C., & Witzenberger, K. (2022). Amazon and the new global connective architectures of education governance. Harvard Educational Review, 92(2), 231–256.

JRC RESEARCH ON THE IMPACT OF GENAI ON EDUCATION AND TRAINING

A scoping literature review was conducted to examine the emerging body of research on the impact of GenAI on education.[430] The review analysed 283 publications and identified key features and gaps in the current research. The results show that:

- → Research on GenAI in education is predominantly focused on Western, Educated, Industrialised, Rich, and Democratic (WEIRD) countries, with a lack of representation from Latin America and Africa.
- → Most studies focus on higher education, with a need for more research on other education sectors.
- → The evaluation of GenAI in tests and tasks, and its general application in education, are the most prominent areas of research.
- → There is a need for more confirmatory studies to consolidate existing knowledge and assumptions.
- → Ethical issues, such as the responsible use of GenAI and its potential impact on academic integrity, and technological issues, while present, receive relatively less attention.

Based on these findings, the authors recommend that not only future research should broaden its scope to include more diverse population samples, education sectors, and controversial topics, such as ethics, but also educational practice should focus on equipping teachers and educators with the necessary skills to navigate GenAI, and consider curricula that integrate education about and with AI.

430. Forthcoming JRC Report: Scoping review of GenAI research in education studies.

→ Educational policy should support research on GenAI in education, ensure individuals' rights and sovereignty, and navigate tensions between technological determinism and ethical standards.

→ Policies should be devised to ensure academic and educational integrity, and curricula should be revised to systematically integrate education about and with AI.

EMERGING USES AND PERCEPTIONS OF GENAI IN EDUCATION

The JRC conducted a study across five Member States (Ireland, Finland, Germany, Luxembourg, and Spain) to explore the emerging uses and perceptions of GenAI in education.[431] The study involved interviews and focus groups with policymakers, teacher educators, school leaders, teachers, and students. The findings suggest that GenAI is seen as a technological innovation with high potential to shape education, but its integration requires careful implementation, ongoing professional development, and the development of AI literacy.

Key findings and implications for each stakeholder group include:

1. **Teacher educators** need comprehensive training and guidance on GenAI, and initial teacher education programmes should include both technical skills and the ethical implications of GenAI.

2. **Educators** see GenAI as a tool to enhance teaching but have concerns about its potential to hinder learning. They need comprehensive training and guidelines for using GenAI effectively.

3. **Students** are more focused on the benefits of GenAI, such as personalised learning. They are already using tools for tasks like brainstorming, language skills development, and generating learning resources, but require guidance to understand potential risks. They need equitable access to GenAI tools to prevent disparities in learning opportunities.

4. **School leaders** face challenges in integrating GenAI, including lack of time, resources and guidelines. They need to develop internal policies and standards for the ethical and effective use of GenAI.

5. **Policymakers** recognise the need for clear policies and standards to guide the use of GenAI in schools. They should encourage the development of AI literacy among educators and students to foster informed and responsible use of GenAI technologies.

Overall, the study highlights the need for careful consideration and planning to ensure the effective and responsible integration of GenAI in education, including the development of AI literacy, comprehensive training, and guidelines for educators, as well as equitable access to GenAI tools for all students.

PROFESSIONAL SKILLS FOR EDUCATORS

Educators play a crucial role in facilitating learners' digital competence, including AI literacy and support them to benefit from digital technologies opportunities while harnessing challenges and risks. The continuous changing technological landscape, and particularly GenAI, require even more complex competencies and skills for citizens to develop. This may be achieved through:

1. **Teacher training:** To ensure that they are trained effectively to integrate AI into their teaching practices.

2. **Curriculum updates:** AI technology, including its ethical, social, and societal dimensions, should be incorporated into educational curricula.

431. Forthcoming JRC Policy Brief: Uses and perceptions of Generative AI in secondary education across five Member States.

3. **Education-specific AI models:** Systems like EdGPT should be developed to support learning and teaching.

4. **Address AI-specific challenges:** Educators need to be able to address issues like data usage, data privacy, information bias, and equal access.

AI IN VOCATIONAL EDUCATION AND TRAINING

Vocational Education and Training (VET) systems also play a crucial role in promoting AI literacy. VET systems are uniquely positioned to adapt to technological changes and provide practical, hands-on learning experiences that align with industry need. The integration of AI in VET is driven by motivations to enhance educational quality, personalise learning, and prepare students for job markets. The study "Emerging Technologies and Trends in VET"[432] underscores the necessity for strategic policy interventions (e.g. investments in teacher education and equipment/infrastructure; and further research to better understand the impact on technology in learning processes in VET) and to address the challenges and opportunities these technologies present to VET systems.

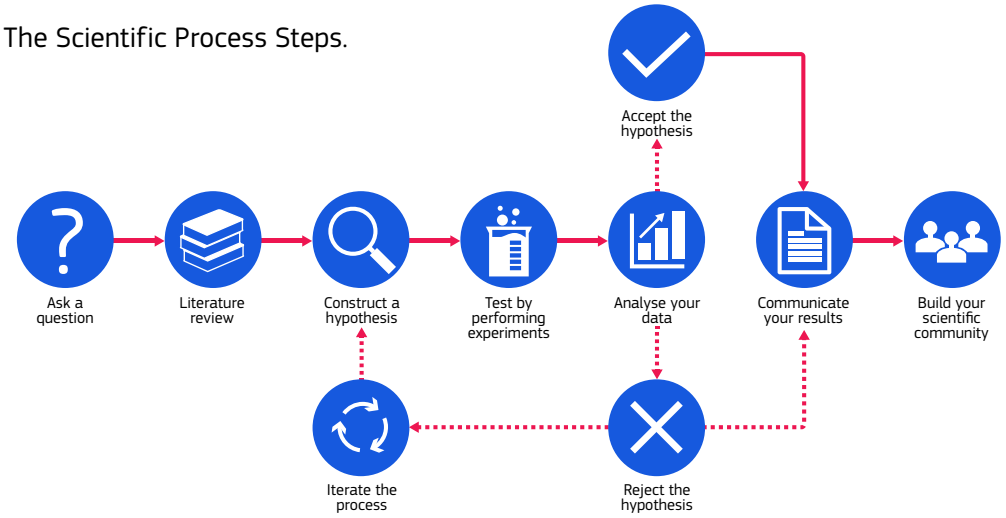
6.3 Impact of Generative AI in Science

KEY MESSAGES

- GenAI is reshaping the scientific process. It offers unprecedented efficiency and creativity but requires careful oversight to maintain scientific integrity.
- While GenAI facilitates advancements in science by democratising access and fostering collaboration, it also poses challenges such as potential biases and the risk of reinforcing dominant narratives, necessitating a balanced integration of AI with human expertise.

This Section provides an overview of the impact of GenAI on science. Specifically, it discusses how it influences all steps of the scientific process and effectively modifies research methodologies. The scientific process (Figure 19), commonly known as the scientific method, is a systematic approach that forms the basis of scientific understanding. It assists researchers in developing questions, planning experiments, and making conclusions about their observations of the world.

Figure 19. The Scientific Process Steps.



Source: Own elaboration, adapted from the literature.[433] [434]

432. The study is published as a chapter of the monography European Commission, Joint Research Centre, Herrero, C. and López Cobo, M. (editors), Supporting the digital transformation of Vocational Education and Training. Publications Office of the European Union, Luxembourg, 2025, JRC141881.

433. Wright, G. (2023). Scientific method definition. Technical report. https://www.techtarget.com/whatis/definition/scientific-method

434. Mitchell, P. I. (2024). Steps of the scientific method. Technical report. https://www.dbu.edu/mitchell/medieval-resources/sciencemethodoverview.html

→ **Ask a question (or make an observation).** The process begins with asking questions or making observations, guiding empirical exploration. GenAI can revolutionise this step by helping researchers, especially novices, refine and contextualise inquiries.[435] [436] [437] These models synthesise literature to highlight patterns and gaps, aiding in formulating creative questions.[438] [439] [440] AI can autonomously generate research goals or hypotheses, effectively "observing" phenomena,[441] [442] accelerating question formulation, and expanding scientific curiosity.[443] [444] [445] However, it may inadvertently reinforce dominant narratives, limiting exploration of new ideas.[446]

→ **Literature review.** GenAI tools and systems such as Elicit, Scite, and Scopus AI transform this step by integrating literature search, retrieval, and summarisation.[447] [448] [449] [450] [451] While concerns remain,[452] GenAI can facilitate broaden access to

435. França, C. (2023). AI empowering research: 10 ways how science can benefit from AI (arXiv:2307.10265). arXiv. https://doi.org/10.48550/ARXIV.2307.10265

436. Khlaif, Z. N., Mousa, A., Hattab, M. K., Itmazi, J., Hassan, A. A., Sanmugam, M., & Ayyoub, A. (2023). The Potential and Concerns of Using AI in Scientific Research: ChatGPT Performance Evaluation. JMIR Medical Education, 9, e47049. https://doi.org/10.2196/47049

437. Burger, B., Kanbach, D. K., Kraus, S., Breier, M., & Corvello, V. (2023). On the use of AI-based tools like ChatGPT to support management research. European Journal of Innovation Management, 26(7), 233–241. https://doi.org/10.1108/EJIM-02-2023-0156

438. Bail, C. A. (2024). Can Generative AI improve social science? Proceedings of the National Academy of Sciences, 121(21), e2314021121. https://doi.org/10.1073/pnas.2314021121

439. Xu, R., Sun, Y., Ren, M., Guo, S., Pan, R., Lin, H., Sun, L., & Han, X. (2024). AI for social science and social science of AI: A survey. Information Processing & Management, 61(3), 103665. https://doi.org/10.1016/j.ipm.2024.103665

440. Erduran, S. (2023). AI is transforming how science is done. Science education must reflect this change. Science, 382(6677), eadm9788. https://doi.org/10.1126/science.adm9788

441. Ifargan, T., Hafner, L., Kern, M., Alcalay, O., & Kishony, R. (2024). Autonomous LLM-driven research from data to human-verifiable research papers (arXiv:2404.17605). arXiv. https://doi.org/10.48550/arXiv.2404.17605

442. Zenil, H., et al. (2023). The Future of Fundamental Science Led by Generative Closed-Loop Artificial Intelligence (arXiv:2307.07522). arXiv. https://doi.org/10.48550/arXiv.2307.07522

443. Gao, J., & Wang, D. (2024). Quantifying the Benefit of Artificial Intelligence for Scientific Research (arXiv:2304.10578). arXiv. https://doi.org/10.48550/arXiv.2304.10578

444. Cai, Y., Deng, Q., Lv, T., Zhang, W., & Zhou, Y. (2025). Impact of GPT on the Academic Ecosystem. Science & Education, 34(2), 913–931. https://doi.org/10.1007/s11191-024-00561-9

445. Van Noorden, R., & Perkel, J. M. (2023). AI and science: What 1,600 researchers think. Nature, 621.

446. Conroy, G. (2023). How ChatGPT and other AI tools could disrupt scientific publishing. Nature, 622(7982), 234–236. https://doi.org/10.1038/d41586-023-03144-w

447. Alqahtani, T., et al. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. Research in Social and Administrative Pharmacy, 19(8), 1236–1242. https://doi.org/10.1016/j.sapharm.2023.05.016

448. Park, K.-S., & Choi, H. (2024). How to Harness the Power of GPT for Scientific Research: A Comprehensive Review of Methodologies, Applications, and Ethical Considerations. Nuclear Medicine and Molecular Imaging, 58(6), 323–331. https://doi.org/10.1007/s13139-024-00876-z

449. Chen, X. S., & Feng, Y. (2024). Exploring the use of generative artificial intelligence in systematic searching: A comparative case study of a human librarian, ChatGPT-4 and ChatGPT-4 Turbo. IFLA Journal, 03400352241263532. https://doi.org/10.1177/03400352241263532

450. Dashkevych, O., & Portnov, B. A. (2024). How can generative AI help in different parts of research? An experiment study on smart cities' definitions and characteristics. Technology in Society, 77, 102555. https://doi.org/10.1016/j.techsoc.2024.102555

451. Fernández-López, J., Borrás-Rocher, F., Viuda-Martos, M., & Pérez-Álvarez, J. Á. (2024). Using Artificial Intelligence-Based Tools to Improve the Literature Review Process: Pilot Test with the Topic "Hybrid Meat Products." Informatics, 11(4), 72. https://doi.org/10.3390/informatics11040072

452. Tomczyk, P., Brüggemann, P., & Vrontis, D. (2024). AI meets academia: Transforming systematic literature reviews. EuroMed Journal of Business. https://doi.org/10.1108/EMJB-03-2024-0055

theoretical foundations across disciplines.[453] [454] In addition, as explained in the communication point below, the possibility to quickly generate summaries of complex research papers or answer questions in a conversational manner can help researchers digest larger amounts of information than in the past. Researchers must, however, confront possible inaccuracy and hallucinations.

→ **Construct a hypothesis.** Formulating a hypothesis transforms a research question into a testable proposition. GenAI supports this by identifying patterns in data and literature, enabling plausible hypotheses grounded in evidence.[455] GenAI assists in reframing problems, identifying blind spots, and suggesting structured, testable statements aligned with research goals.[456] Advanced approaches explore open-ended hypothesis spaces beyond traditional methods. With capabilities like step-by-step logic, GenAI supports hypothesis construction but may miss insights requiring human expertise.

→ **Test your hypothesis by performing an experiment.** This step validates scientific claims. GenAI automates experiment design, code generation, and execution, ensuring seamless transitions from hypothesis to analysis.[457] [458] [459] AI aids in controlling variables and recognising patterns, improving the accuracy of experimental results. Experimental comparisons of AI systems reveal GenAI's.[460]

→ **Analyse your data.** Data analysis transforms results into insights, and GenAI can accelerate this by processing datasets, applying statistical methods, and generating visualisations. Tools like PROTEUS and Agent Laboratory automate workflows using LLM reasoning.[461] AI aids coding, modelling, and pattern discovery across domains. GenAI enhances clarity and speed in interpretation, but transparency and reliability are key.

→ **Draw conclusions based on acceptance or rejection of the hypothesis.** This involves determining whether results support or not the hypothesis and integrating findings with prior knowledge. GenAI tools assist in summarising results and aligning them with hypotheses. GenAI aids in interpreting outputs, drafting conclusions, and identifying inconsistencies. Caution is needed as risks like bias and overconfidence require human validation.

453. Borger, J. G., et al. (2023). Artificial intelligence takes center stage: Exploring the capabilities and implications of ChatGPT and other AI-assisted technologies in scientific research and education. Immunology & Cell Biology, 101(10), 923–935. https://doi.org/10.1111/imcb.12689

454. Microsoft Research AI4Science, & Microsoft Azure Quantum (2023). The Impact of Large Language Models on Scientific Discovery: A Preliminary Study using GPT-4 (arXiv:2311.07361). arXiv. https://doi.org/10.48550/arXiv.2311.07361

455. Liu, H., Zhou, Y., Li, M., Yuan, C., & Tan, C. (2025). Literature Meets Data: A Synergistic Approach to Hypothesis Generation (arXiv:2410.17309). arXiv. https://doi.org/10.48550/arXiv.2410.17309

456. Kabir, A., Shah, S., Haddad, A., & Raper, D. M. S. (2025). Introducing Our Custom GPT: An Example of the Potential Impact of Personalized GPT Builders on Scientific Writing. World Neurosurgery, 193, 461–468. https://doi.org/10.1016/j.wneu.2024.10.041

457. Bersenev, D., Yachie-Kinoshita, A., & Palaniappan, S. K. (2024). Replicating a High-Impact Scientific Publication Using Systems of Large Language Models. https://doi.org/10.1101/2024.04.08.588614

458. Schmidgall, S., Su, Y., Wang, Z., Sun, X., Wu, J., Yu, X., Liu, J., Liu, Z., & Barsoum, E. (2025). Agent Laboratory: Using LLM Agents as Research Assistants. https://doi.org/10.48550/ARXIV.2501.04227

459. Owoahene Acheampong, I., & Nyaaba, M. (2024). Review of Qualitative Research in the Era of Generative Artificial Intelligence. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4686920

460. Dashkevych, O., & Portnov, B. A. (2024). How can generative AI help in different parts of research? An experiment study on smart cities' definitions and characteristics. Technology in Society, 77, 102555. https://doi.org/10.1016/j.techsoc.2024.102555

461. Ding, N., et al. (2024). Automating Exploratory Proteomics Research via Language Models (arXiv:2411.03743). arXiv. https://doi.org/10.48550/arXiv.2411.03743

When integrated responsibly, GenAI supports reflection on findings and encourages hypothesis refinement.[462]

→ **Communicate your results.** This step ensures scientific knowledge is shared and evaluated. GenAI supports manuscript drafting, formatting, editing, and translation, enhancing clarity and reach. Tools like ChatGPT aid in structuring papers and producing summaries and multilingual reports.[463] GenAI assists with grant writing and presentations, reshaping documentation.[464] While efficient, transparency and oversight are essential to uphold integrity. Recent studies highlight GenAI's growing role in science communication – conveying scientific research and information to non-experts. These include helping create accessible summaries, interactive Q&As, automatic creation of data visualisations and insight generation (drafting short data-driven stories). However, risks of misinformation and public distrust demand human oversight to ensure accuracy and preserve credibility.[465] [466] [467]

→ **Build your scientific community.** Scientific communities shape research practices and foster collaboration. GenAI enables interdisciplinary engagement and inclusive collaboration.[468] GenAI tools bridge language and expertise gaps, democratising participation and fostering shared norms. While supporting capacity-building and research integrity,[469] overemphasis on GenAI could overshadow cultural perspectives.

6.4 GenAI in Cybersecurity

KEY MESSAGES

- GenAI is reshaping the scientific process. It offers unprecedented efficiency and creativity but requires careful oversight to maintain scientific integrity.
- While GenAI facilitates advancements in science by democratising access and fostering collaboration, it also poses challenges such as potential biases and the risk of reinforcing dominant narratives, necessitating a balanced integration of AI with human expertise.

The rapid advancements in AI technologies have significantly impacted the world of cybersecurity. After discussing cybersecurity issues related to GenAI in Section 2.2, here we delve into the most prominent research directions and potential real-life applications.

Defending Against Social Engineering. Humans, a frequent attack point in cybersecurity, can be empowered by GenAI systems to recognise, deflect, and report different attempts of attacks focusing on the human vector. Social

462. Bond, A., Cilliers, D., Retief, F., Alberts, R., Roos, C., & Moolman, J. (2024). Using an Artificial intelligence chatbot to critically review the scientific literature on the use of Artificial intelligence in Environmental Impact Assessment. Impact Assessment and Project Appraisal, 42(2), 189–199. https://doi.org/10.1080/14615517.2024.2320591

463. The Royal Society. (2024). Science in the age of AI - How artificial intelligence is changing the nature and method of scientific research.

464. Wiley. (2025). ExplanAItions—An AI study by Wiley.

465. Wu, Yang, et al. "Automated data visualization from natural language via large language models: An exploratory study." *Proceedings of the ACM on Management of Data* 2.3 (2024): 1-28.

466. Kessler, S. H., Mahl, D., Schäfer, M. S. and Volk, S. C. (2025). Science communication in the age of artificial intelligence JCOM 24(2), E. https://doi.org/10.22323/2.24020501

467. Schäfer, M. S., Kremer, B., Mede, N. G. and Fischer, L. (2024). Trust in science, trust in ChatGPT? How Germans think about generative AI as a source in science communication JCOM 23(09), A04. https://doi.org/10.22323/2.23090204

468. Bianchini, S., Müller, M., & Pelletier, P. (2023). Drivers and Barriers of AI Adoption and Use in Scientific Research. (arXiv:2312.09843). arxiv. https://doi.org/10.48550/arXiv.2312.09843

469. Sun, L., Chan, A., Chang, Y. S., & Dow, S. P. (2024). ReviewFlow: Intelligent Scaffolding to Support Academic Peer Reviewing. Proceedings of the 29th International Conference on Intelligent User Interfaces, 120–137. https://doi.org/10.1145/3640543.3645159

engineering attacks, such as spam and phishing, aim to influence and manipulate an individual to unwillingly or unknowingly yield system access to cybercriminals. To combat such malicious attempts, GenAI agents can be applied to scrutinise incoming textual, voice, and video content in email and chat messages and social media posts. Beyond attack management, GenAI can also explain why certain content is potentially predatory, strengthening the cybersecurity knowledge of individuals.[470] GenAI-powered monitoring of incoming content can also lead to recognition, notification, and treatment of targeted misinformation and deceitful narratives even at an organisational level.[471]

Automated Threat Detection and Response. Natural language is the prevailing input method when interacting with GenAI models. However, these systems can operate on other data formats, making them a compelling tool for analysis on formats such as network traffic and system logs. Such analysis might lead to the detection, correlation, understanding, and mitigation of cyber threats. While more traditional deep learning approaches are already used for such a scenario, GenAI relies on its wider knowledge gamut to perform such tasks while also providing insights and recommendations to cybersecurity specialists.[472]

Security Testing. Verification of security measures in place is a crucial step towards a secure system. Penetration testing is a widely adopted method to validate security controls and identify system vulnerabilities. Nowadays, penetration testing requires a high degree of manual expert work, making the process time-consuming and expensive. GenAI can contribute in multiple ways, making penetration testing faster, more automated, and leading to a more accurate and comprehensive verification process. Current trends foresee the usage of GenAI for threat planning, where an LLM is tasked to recommend an attack palette for a given organisation or infrastructure; agentic behaviour, where a GenAI system executes cyber attacks; and environment, data, and scenario generation used for penetration testing. While showing that potential, current GenAI approaches are capable of automated penetration on less complex targets, more sophisticated exploits still require human operation.[473]

Code Security. Beyond experts and end users, software developers playdnother important role in the cybersecurity lifecycle. The security of the code they write depends on their level of security knowledge and intent. That leads to accidental, or even malicious, introduction of code vulnerabilities. GenAI methods display potential to assist with recognising such vulnerabilities in previously produced code, or even as a programming companion, screening code as it is being written. State-of-the-art approaches based on current GenAI methods can correctly recognise and advise on more elementary vulnerabilities. However, similar to security testing, their accuracy drops when processing more complex problems that require a wider context and more complex inference to evaluate larger codebases.[474]

Education and Awareness. GenAI can support security training of individuals at any level of expertise. It can provide tailored interventions and more engaging material for organisational security awareness training,[475] as well as

470. Koide, Takashi, et al. "Chatspamdetector: Leveraging large language models for effective phishing email detection." *arXiv preprint arXiv:2402.18093* (2024).

471. Yu, Jingru, et al. "The Shadow of Fraud: The Emerging Danger of AI-powered Social Engineering and its Possible Cure." *arXiv preprint arXiv:2407.15912* (2024).

472. Ali, Tarek, and Panos Kostakos. "Huntgpt: Integrating machine learning-based anomaly detection and explainable ai with large language models (llms)." *arXiv preprint arXiv:2309.16021* (2023).

473. Deng, Gelei, et al. "{PentestGPT}: Evaluating and harnessing large language models for automated penetration testing." *33rd USENIX Security Symposium* (USENIX Security 24). 2024.

474. Khare, Avishree, et al. "Understanding the effectiveness of large language models in detecting security vulnerabilities." *arXiv preprint arXiv:2311.16169* (2023).

475. Greco, Francesco, Giuseppe Desolda, and Luca Viganò. "Supporting the Design of Phishing Education, Training and Awareness interventions: an LLM-based approach." *CEUR WORKSHOP PROCEEDINGS*. Vol. 3700. CEUR-WS, 2024.

automated generation of cybersecurity scenarios and exercises[476] as part of any formal educational programme. These capabilities can also be leveraged to improve public cybersecurity literacy, strengthening collective resilience. More generally, GenAI can facilitate the communication of security and privacy information. In this area, research has shown the applicability of artificial intelligence to simplify privacy policies of digital services, oftentimes verbose, confusing, and at a language complexity level that is not accessible to all citizens. This can be accomplished through summarisation, allowing interactive analysis of privacy guarantees, and providing intuitive visualisations about how personal data are used and shared.[477] GenAI brings the potential to move from this type of automated analysis to fully personalised privacy assistants.[478]

Challenges and Considerations. GenAI changes the cybersecurity playground by democratising access to artificial intelligence. This augments the capabilities of both malicious actors and security professionals to launch and defend against more sophisticated attacks at scale with lower technical friction. The main disruption factor comes from the natural language interface that allows humans to speak with AI systems, to instruct them, ask questions, reason together, and make or delegate decisions based on that interaction. While valuable work to secure GenAI is focusing on improving the machine learning pipeline and adding technical guardrails, research and policy interventions are needed on the human-AI interaction side. The anthropomorphising of AI-assistants by their users, combined with GenAI's capability to generate eloquent, convincing outputs and the increased agency of existing deployments, raises risks that target human cognitive and perception vulnerabilities. It is imperative to explore human-centred safeguards and interaction designs that foster critical thinking and adapt autonomy levels according to the context, especially when it comes to high stakes cybersecurity scenarios. This groundwork can support a technological future where both humans and agents collaborate efficiently, enhancing overall performance in cybersecurity practice. Moreover, not only experts will benefit. GenAI can serve as a catalyst for a broader cybersecurity literacy by supporting upskilling and reskilling initiatives across diverse user groups. Lowering barriers to access and understanding can help broaden participation in cybersecurity, strengthening our societal resilience in the face of evolving threats.

476. Yamin, Muhammad Mudassar, et al. "Applications of llms for generating cyber security exercise scenarios." *IEEE Access* (2024).

477. Woodring, Justin, Katherine Perez, and Aisha Ali-Gombe. (2024). "Enhancing privacy policy comprehension through privacy: A user-centric approach using advanced language models." *Computers & Security* 145: 103997.

478. Chen, Chaoran, et al. "CLEAR: Towards Contextual LLM-Empowered Privacy Policy Analysis and Risk Generation for Large Language Model Applications." *Proceedings of the 30th International Conference on Intelligent User Interfaces*. 2025.

6.5 Use of Generative AI in the Public Sector

KEY MESSAGES

- GenAI has the potential to transform public sector management and service delivery, but its adoption raises complex challenges and opportunities that require careful consideration and strategic attention.
- Effective governance and regulatory approaches are crucial in ensuring the safe, ethical, and lawful use of GenAI technologies in the public sector, and addressing the risks and benefits associated with its adoption.

As part of the AI-led transformation of our society, public sector organisations are increasingly using AI-based solutions to address internal operational needs and provide public services. The JRC has been closely following the uptake and use of AI in the public sector, including GenAI, to collect scientific evidence and provide policy advice. The JRC's work is crucial in ensuring that policies concerning the use of AI in the public sector are informed by factual and independent knowledge, recognising that the stakes in this

sector are high due to the potential impact on both citizens and businesses.

SETTING THE SCENE: AI IN THE PUBLIC SECTOR

The latest Public Sector Tech Watch report[479] provides a comprehensive picture of how public administrations across Europe use AI and other emerging technologies. Based on over 1,600 documented cases, the report shows widespread AI adoption for enhancing public services (53% of cases) and improving internal administrative efficiency (47%). National governments typically focus on streamlining internal processes, while local authorities prioritise citizen-focused applications. The report also highlights the growing interest in GenAI, with many pilot projects emerging across different public sector contexts. This increasing trend highlights new questions about governance, accountability, transparency, and public value, emphasising the need for ongoing research to guide future decision-making in this area.

A large-scale survey conducted by the JRC, involving public managers from seven Member States similarly reveals that AI is now widely implemented, especially in service delivery and internal operations, though its use in policymaking remains limited. The study highlights that adoption is driven by technical capabilities, leadership, innovation-friendly culture, and internal expertise, as well as the expectations of citizens. Ensuring a balanced and trustworthy integration of AI will depend on the continued strengthening of in-house capacities, ethical awareness, and citizen-oriented strategies.[480]

However, the successful implementation of AI (including GenAI) in public organisations remains a complex task and hinges on more than just the technology. Research has identified five interrelated sets of challenges: societal expectations, ethical concerns, legal and regulatory issues, technical implementation, and organisational change. This latter requires alignment with daily practices, capacity-building, and a clear understanding of how change is perceived and managed by staff, calling for a broader view of AI adoption that includes institutional learning and adaptive governance.[481]

Supporting this evolving landscape, the JRC has developed a detailed framework of the competences and governance practices needed to enable meaningful and responsible AI use in the public sector, highlighting that skill-building and good governance are mutually reinforcing. Strengthening both dimensions is essential not only for effective AI adoption but also for ensuring that its use aligns with public values and institutional goals.[482]

479. European Commission: Directorate-General for Digital Services, Brizuela, A., Montino, C., Galasso, G., Polli, G. et al., Adoption of AI, blockchain and other emerging technologies within the European public sector – A public sector Tech Watch report, Publications Office of the European Union, 2024, https://data.europa.eu/doi/10.2799/3438251

480. European Commission: Joint Research Centre, Grimmelikhuijsen, S. and Tangi, L., What factors influence perceived artificial intelligence adoption by public managers, Publications Office of the European Union, Luxembourg, 2024, https://data.europa.eu/doi/10.2760/0179285, JRC138684.

481. Tangi, L., van Noordt, C. and Rodriguez Müller, A. P. (2023), 'The challenges of AI implementation in the public sector: An in-depth case studies analysis', in Proceedings of the 24th Annual International Conference on Digital Government Research, Association for Computing Machinery, New York, pp. 414–422 https://doi.org/10.1145/3598469.3598516.

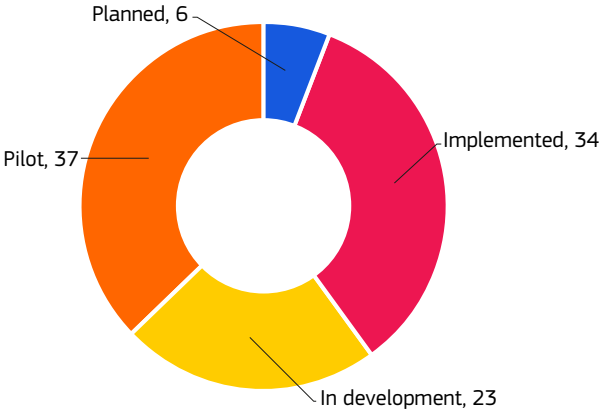
482. European Commission: Joint Research Centre, Medaglia, R., Mikalef, P. and Tangi, L., Competences and governance practices for artificial intelligence in the public sector, Publications Office of the European Union, Luxembourg, 2024, https://data.europa.eu/doi/10.2760/7895569, JRC138702.

TOWARDS THE USE OF GENAI IN THE PUBLIC SECTOR

So far, Public Sector Tech Watch[483] has identified around 100 cases of the use of GenAI, with public administrations exploring the adoption of a range of GenAI technologies in different application areas. This emerging trend of embracing GenAI demands dedicated attention to exploring the benefits and understanding the distinct challenges associated with its adoption.

Amongst the use cases identified, the primary applications are in public services and engagement (51) and improving internal administrative efficiency (31), followed by analysis, monitoring and regulatory research (16). The growing collection of cases reveals projects at different stages of maturity (planned, in development, piloted or fully implemented), and indicates a significant expansion ahead (see Figure 20).

Figure 20. Distribution of GenAI cases according to the state of development.

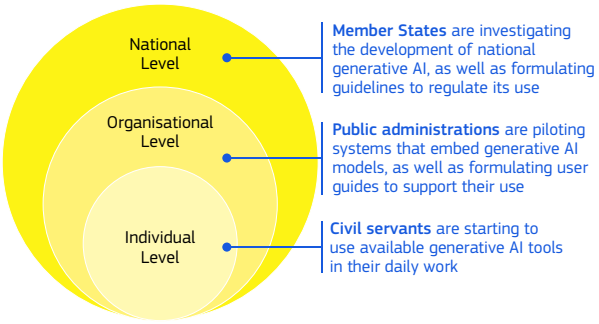


Source: Public Sector Tech Watch (PSTW) dataset.

A recent JRC study shows that GenAI is not only being piloted in formal public sector projects but is also increasingly used informally by public managers at an individual level across the EU. Around 30% of respondents already use GenAI in their daily work, and a further 44% intend to adopt it soon, while 26% report no current interest in using these tools. This uptake cuts across sectors and age groups, highlighting how GenAI tools - commonly and freely available online - are entering the workplace, often ahead of formal strategies or guidance.

However, as these trends increase, public administrations have begun establishing frameworks and guidelines to explicitly address the safe, ethical, and lawful use of GenAI technologies in public sector contexts.

Figure 21. GenAI current practices in the public sector.



Source: JRC elaboration.

GENAI GUIDELINES AND POLICIES ACROSS THE EU

Current evidence from across the EU shows public administrations actively developing policies, guidelines, and procedural frameworks to manage the use of GenAI, also taking into account the AI Act. A review of 33 such documents and guidelines, issued by national, regional, and local authorities, reveals most were published in 2023 and 2024 and focus specifically on GenAI. These documents respond to growing concerns about transparency, human oversight, data protection, and accountability.[484]

483. The Public Sector Tech Watch is an observatory dedicated to monitoring, analysing and disseminating the use of emerging technologies (e.g., Blockchain, Artificial Intelligence, etc.) within the public sector in Europe. It is managed by the Directorate-General for Informatics (DIGIT) and the Joint Research Centre (JRC) of the European Commission. This collection includes the data produced by the observatory. More details: https://joinup.ec.europa.eu/collection/public-sector-tech-watch

484. European Commission. Public Sector Tech Watch. Analysis of the generative AI landscape in the European

Common themes include the need for public employees to remain responsible for the content generated with GenAI tools, to avoid disclosing sensitive or non-public information, and to critically assess the accuracy and appropriateness of AI-generated outputs. Several guidelines also provide practical resources, such as technical annexes on prompt engineering, internal protocols for safe use, and templates for disclosing the involvement of GenAI in document creation.

The development of these frameworks reflects a broader shift toward operationalising trustworthy AI in public administration, with an emphasis on ensuring alignment with ethical principles, such as fairness, prevention of harm, and respect for autonomy.

The possibilities to use AI in the public sector – by public administrations, but also in many different public-private partnerships – have recently been boosted by the adoption and early implementation of the Interoperable Europe Act.[485] This piece of legislation empowers public administrations to provide public services seamlessly across territorial, sectoral, and organisational boundaries, while maintaining their sovereignty from local to EU level. It also sets the framework for interoperable regulatory sandboxes, supporting the digital commons (mainly open-source solutions), and stimulating small and medium-sized solution providers (GovTech) to innovate by fostering the update of emerging technologies. Complementing these efforts, the upcoming Apply AI Strategy[486] aims to accelerate the uptake of AI in key sectors, including public services, by improving access to trustworthy AI models, enhancing the public sector's capacity to experiment and procure AI, and supporting cross-border collaboration and infrastructure. Together, these initiatives reflect a growing momentum at EU level to strengthen the strategic and sovereign deployment of AI, including GenAI, in ways that align with public values and societal goals.

OUTLOOK/FUTURE PERSPECTIVES

As GenAI becomes increasingly embedded in public administrations, new questions are emerging that warrant further research and strategic attention. These questions span multiple levels (see Figure 21) – national, organisational, and individual – and reflect the evolving roles of governments, institutions, and civil servants in shaping how GenAI is developed, applied and governed. However, as these trends increase, public administrations have begun establishing frameworks and guidelines to explicitly address the safe, ethical, and lawful use of GenAI technologies in the public sector.

At the national level, some EU Member States are investing in open-source, language-specific GenAI models, which have the potential to help protect linguistic diversity, improve cultural alignment and contribute to national competitiveness, while also offering enhanced data protection and safeguarding national security interests. However, this trend also raises questions about sustainability, governance, and collaboration. The regulatory landscape is also in flux, with some countries moving quickly to set rules for potentially risky use cases, while others opt for a lighter-touch approach.

Public administrations are beginning to integrate GenAI into specific processes, often through pilot projects or by issuing internal guidance for staff. However, implementation remains uneven, and further inquiry is needed into what organisational conditions enable responsible and meaningful adoption.

At the individual level, the use of GenAI tools is growing rapidly, often outside formal strategies. Public servants increasingly rely on these widely available tools to explore new ideas or draft texts and summaries. These informal practices may raise questions about oversight, consistency,

public sector. Publications Office of the European Union, Luxemburg, 2025, https://data.europa.eu/doi/10.2799/0409819

485. https://interoperable-europe.ec.europa.eu/interoperable-europe/interoperable-europe-act

486. AI Continent Action Plan. COM(2025) 165 final

accuracy, data protection and equity, among others.

As GenAI supports more cognitive functions, it challenges the traditional boundaries between tool and collaborator. There is a need to explore how this affects policy development, problem-solving, and innovation in the public sector. The effects of GenAI in public administration are likely to be diffuse, context-dependent, and not easily captured by standard metrics. ■

CONCLUSIONS

As clarified in the beginning of this Outlook report, Generative AI refers to a subset of Artificial Intelligence technologies that enable machines to generate new content, such as images, videos, text, and music that is often indistinguishable from that created by humans. It is a revolutionary technology with tremendous disruptive potential, which needs to be better understood and that will require policy responses at EU level in many respects.

As highlighted, GenAI is an economic sector in its own right, and comes with its own set of challenges and opportunities. Although Europe is often seen as lagging behind other global leaders like the United States and China, our analysis of the global GenAI landscape reveals a more nuanced picture. In particular, we underscore the EU's strong position in research. To remain competitive, the EU must cultivate a vibrant and dynamic ecosystem of actors, with a strong presence of start-ups and established companies. This will require to address issues such as investment, talent, and innovation, and to create an environment that fosters the development and deployment of AI solutions, in line with the ambitious agenda put forward by the recently adopted AI Continent Action Plan[487], and EU's Competitiveness Compass.[488]

Beyond GenAI's role in competitiveness, we must acknowledge the profound societal impact that this technology is likely to have. GenAI is not merely a sectoral issue but a general-purpose technology that will permeate various aspects of our lives, from healthcare and education to transportation and employment. Its effects will reach across multiple industries and application domains, and it is crucial that we develop a holistic policy approach that takes into account its far-reaching consequences. Within this context, it becomes clear that while Generative AI presents a multitude of opportunities for innovation and economic growth within the European Union, it also poses significant challenges that deserve comprehensive responses to navigate its societal and competitive impacts effectively.

In closing, it is important to also emphasise that scientific evidence is essential in guiding the policies that relate to GenAI in ensuring that they are effective and fact-based. As new technologies quickly emerge, this report offers a broad overview of important aspects to take into account for a better understanding of the tecno-socio-economic implications of GenAI. Even as technology evolves, these dimensions, when taken into account would contribute to the emergence of sound GenAI initiatives that are aligned with our societal values and legal frameworks.

487. https://commission.europa.eu/topics/eu-competitiveness/ai-continent_en

488. https://commission.europa.eu/topics/eu-competitiveness/competitiveness-compass_en

REFERENCES

Abdelnabi, S., et al. (2023). Not what you've signed up for: Compromising real-world LLM-integrated applications with indirect prompt injection. In M. Pintor, X. Chen, & F. Tramèr (Eds.), *Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security (AISec 2023)* (pp. 79–90). ACM. https://doi.org/10.1145/3605764.3623985

Abendroth-Dias, K., et al. (2025). *DGTES handbook: A snapshot of EU digital competitiveness and dependencies*. Publications Office of the European Union.

Aldasoro, I., Gambacorta, L., Korinek, A., Shreeti, V., & Stein, M. (n.d.). *Intelligent financial system: How AI is transforming finance (Working Papers No. 1194)*.

Alexander, S. (2025). Deep fake cyberbullying: The psychological toll on students and institutional challenges of AI-driven harassment. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 98*(2), 36–50. https://doi.org/10.1080/00098655.2025.2488777

Ali, S., et al. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain trustworthy Artificial Intelligence. *Information Fusion, 99*, 101805. https://doi.org/10.1016/j.inffus.2023.101805

Ali, T., & Kostakos, P. (2023). HuntGPT: Integrating machine learning-based anomaly detection and explainable AI with large language models (LLMs). *arXiv preprint arXiv:2309.16021*.

Alqahtani, T., et al. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy, 19*(8), 1236–1242. https://doi.org/10.1016/j.sapharm.2023.05.016

Alvarado, R. (2023). AI as an epistemic technology. *Science and Engineering Ethics, 29*, Article 32. https://doi.org/10.1007/s11948-023-00451-3

Andriushchenko, M., Croce, F., & Flammarion, N. (2024). Jailbreaking leading safety-aligned LLMs with simple adaptive attacks. *arXiv:2404.02151 [cs.CR]*. https://arxiv.org/abs/2404.02151

Apollo Research. (2024). *We need a Science of Evals*. https://www.apolloresearch.ai/blog/we-need-a-science-of-evals

Apruzzese, G., et al. (2023). "Real Attackers Don't Compute Gradients": Bridging the gap between adversarial ML research and practice. In *2023 IEEE Conference on Secure and Trustworthy Machine Learning, SaTML 2023*.

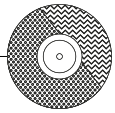
Bail, C. A. (2024). Can generative AI improve social science? *Proceedings of the National Academy of Sciences, 121*(21), e2314021121. https://doi.org/10.1073/pnas.2314021121

Barrett, C. W., et al. (2023). Identifying and mitigating the security risks of generative AI. *Foundations and Trends in Privacy and Security, 6*(1), 1–52. https://doi.org/10.1561/3300000041

Bashir, N., Donti, P., Cuff, J., Sroka, S., Ilic, M., Sze, V., Delimitrou, C., & Olivetti, E. (2024). The climate and sustainability implications of generative AI. *An MIT Exploration of Generative AI*. https://doi.org/10.21428/e4baedd9.9070dfe7

Bender, E. M., et al. (2021). On the dangers of stochastic parrots. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bersenev, D., Yachie-Kinoshita, A., & Palaniappan, S. K. (2024). Replicating a high-impact scientific publication using systems of large language models. *https://doi.org/10.1101/2024.04.08.588614*



Bertoldi, P. (2023). *Assessment framework for data centres in the context of Activity 8.1 in the Taxonomy Climate Delegated Act (JRC131733).* European Commission, Ispra.

Beullens, K., Bozzola, E., Cataldo, I., Hale, L., Kent, M., Montag, C., Nivins, S., O'Reilly, M., Rubæk, L., Schiøtz Thorud, H.-M., Sterpenich, V., & Vandenbosch, L. (2025). Minors' health and social media: An interdisciplinary scientific perspective. In S. Manolios, A. Sala, E. Sundorph, S. Chaudron, & E. Gomez (Eds.), *Publications Office of the European Union*. https://data.europa.eu/doi/10.2760/379589

Bianchini, S., Müller, M., & Pelletier, P. (2023). Drivers and barriers of AI adoption and use in scientific research. *https://doi.org/10.48550/arXiv.2312.09843*

Blikstein, P., Zheng, Y., & Zhou, K. Z. (2022). Ceci n'est pas une école: The discourses of artificial intelligence in education through the lens of semiotic analytics. *European Journal of Education, 57*(4), 571–583.

Bohr, A., & Memarzadeh, K. (2020). Chapter 2 - The rise of artificial intelligence in healthcare applications. In A. Bohr & K. Memarzadeh (Eds.), *Academic Press* (pp. 25–60). https://doi.org/10.1016/B978-0-12-818438-7.00002-2

Bond, A., Cilliers, D., Retief, F., Alberts, R., Roos, C., & Moolman, J. (2024). Using an artificial intelligence chatbot to critically review the scientific literature on the use of artificial intelligence in environmental impact assessment. *Impact Assessment and Project Appraisal, 42*(2), 189–199. https://doi.org/10.1080/14615517.2024.2320591

Borger, J. G., et al. (2023). Artificial intelligence takes center stage: Exploring the capabilities and implications of ChatGPT and other AI-assisted technologies in scientific research and education. *Immunology & Cell Biology, 101*(10), 923–935. https://doi.org/10.1111/imcb.12689

Bradford, A. (2020). *The Brussels effect: How the European Union rules the world.* Oxford University Press.

Brooks, C., Eggert, S., & Peskoff, D. *The Rise of AI-Generated Content in Wikipedia. arXiv.*

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics, 140*(2), 889–942.

Burden, J. (2024). Evaluating AI evaluation: Perils and prospects. *arXiv:2407.09221*

Burden, J., Tešić, M., Pacchiardi, L., & Hernández-Orallo, J. (2025). Paradigms of AI evaluation: Mapping goals, methodologies and culture. *arXiv:2502.15620*

Burger, B., Kanbach, D. K., Kraus, S., Breier, M., & Corvello, V. (2023). On the use of AI-based tools like ChatGPT to support management research. *European Journal of Innovation Management, 26*(7), 233–241. https://doi.org/10.1108/EJIM-02-2023-0156

Bush, S. (2025, April). Anime lessons in the limits of AI. *Financial Times.* https://on.ft.com/4iNW6Wl

Cai, Y., Deng, Q., Lv, T., Zhang, W., & Zhou, Y. (2025). Impact of GPT on the academic ecosystem. *Science & Education, 34*(2), 913–931. https://doi.org/10.1007/s11191-024-00561-9

Calza, E., et al. (2022). *A policy oriented analytical approach to map the digital ecosystem (DGTES).* Publications Office of the European Union.

Capone, C., & Paolucci, P. S. (2024). Towards biologically plausible model-based reinforcement learning in recurrent spiking networks by dreaming new experiences. *Nature News.* https://www.nature.com/articles/s41598-024-65631-y

Carlini, N., et al. (2019). The secret sharer: Evaluating and testing unintended memorization in neural networks. In N. Heninger & P. Traynor (Eds.), *Proceedings of the 28th USENIX*

Security Symposium (pp. 267–284). USENIX Association. https://www.usenix.org/conference/usenixsecurity19/presentation/carlini

Carlini, N., et al. (2024). Stealing part of a production language model. In *Forty-first International Conference on Machine Learning, ICML 2024,* Vienna, Austria, July 21–27, 2024. Retrieved from https://openreview.net/forum?id=VE3yWXt3KB

Carpentier, E., D'Adda, D., Nepelski, D., & Stake, J. (2025). *European Digital Innovation Hubs Network's activities and customers.* Publications Office of the European Union. https://data.europa.eu/doi/10.2760/7784020

Cedefop. (2025). *Skills empower workers in the AI revolution: First findings from Cedefop's AI skills survey (Policy brief).* Publications Office of the European Union. https://doi.org/10.2801/6372704

Ceresa, M., et al. (n.d.). *Retrieval augmented generation evaluation for health documents (JRC138904).* Publications Office of the European Union.

Chao, P., et al. (2023). Jailbreaking black box large language models in twenty queries. *arXiv preprint arXiv:2310.08419.* https://doi.org/10.48550/arXiv.2310.08419

Chaslot, G. (2017). How YouTube's A.I. boosts alternative facts. *Medium.*

Chen, C., et al. (2025). CLEAR: Towards contextual LLM-empowered privacy policy analysis and risk generation for large language model applications. *Proceedings of the 30th International Conference on Intelligent User Interfaces.*

Chen, X. S., & Feng, Y. (2024). Exploring the use of generative artificial intelligence in systematic searching: A comparative case study of a human librarian, ChatGPT-4 and ChatGPT-4 Turbo. *IFLA Journal, 50*(1), 03400352241263532. https://doi.org/10.1177/03400352241263532

Chen, Y. H., et al. (2017). Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks. *IEEE Journal of Solid-State Circuits, 52*(1), 127–138. https://ieeexplore.ieee.org/document/7738524

Cheng, Y., Gong, Y., Liu, Y., Song, B., & Zou, Q. (2021). Molecular design in drug discovery: A comprehensive review of deep generative models. *Briefings in Bioinformatics, 22*(6), bbab344.

Chiaro, D., et al. (2025). Generative AI-empowered digital twin: A comprehensive survey with taxonomy. *IEEE Transactions on Industrial Informatics.*

CHIH-HSUAN. (2024, December 27). SONAR: Sentence-Level Multimodal and Language-Agnostic Representations. *Medium.* https://medium.com/@chs.li.work/sonar-sentence-level-multimodal-and-language-agnostic-representations-73a81d3f5913

Choung, H., et al. (2022). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human-Computer Interaction, 39*(9), 1–13. https://doi.org/10.1080/10447318.2022.2050543

Christakis, T. (2024). *AI Hallucinations and Data Subject Rights under the GDPR: Regulatory Perspectives and Industry Responses.* https://ssrn.com/abstract=5042191 or http://dx.doi.org/10.2139/ssrn.5042191

Clusmann, J., et al. (2025). Prompt injection attacks on vision language models in oncology. *Nature Communications, 16*(1), 1239.

COM(2020) 65 final. *White Paper on Artificial Intelligence – A European approach to excellence and trust.*

Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use. http://data.europa.eu/eli/reg_impl/2023/138/oj

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – *A European strategy for data.*

Conroy, G. (2023). How ChatGPT and other AI tools could disrupt scientific publishing. *Nature, 622*(7982), 234–236. https://doi.org/10.1038/d41586-023-03144-w

Consoli, S., Markov, P., Stilianakis, N. I., Bertolini, L., Puertas Gallardo, A., & Ceresa, M. (2024). Epidemic information extraction for event-based surveillance using large language models. In X.-S. Yang et al. (Eds.), *Proceedings of the Ninth International Congress on Information and Communication Technology (ICICT 2024)* (pp. 241–252). Springer Nature. https://doi.org/10.1007/978-981-97-4581-4_17

Consoli, S., Reforgiato Recupero, D., & Petkovic, M. (2019). *Data science for healthcare: Methodologies and applications.* Springer Nature. https://doi.org/10.1007/978-3-030-05249-2

Copyright and Artificial Intelligence Part 2: *Copyrightability.* https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf

Crawford, K. (2024, February 20). Generative AI's environmental costs are soaring – and mostly secret. *Nature.*

Dar, S. U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., & Cukur, T. (2019). Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Transactions on Medical Imaging, 38*(10), 2375–2388.

Das, A., Tariq, A., Batalini, F., et al. (2024). Exposing vulnerabilities in clinical LLMs through data poisoning attacks: Case study in breast cancer.

Dash, S., et al. (2025, March 4). A Deepdive into Aya Vision: Advancing the Frontier of Multilingual Multimodality. *Hugging Face Blog.* https://huggingface.co/blog/aya-vision

Dashkevych, O., & Portnov, B. A. (2024). How can generative AI help in different parts of research? An experiment study on smart cities' definitions and characteristics. *Technology in Society, 77*, 102555. https://doi.org/10.1016/j.techsoc.2024.102555

Dastin, J. (2022). Amazon scraps secret AI recruiting tool that showed bias against women. In *Ethics of data and analytics* (pp. 296–299). Auerbach Publications.

Dauth, W., et al. (2021). The adjustment of labor markets to robots. *Journal of the European Economic Association, 19*(6), 3104–3153.

De-Arteaga, M., et al. (2019). Bias in bios: A case study of semantic representation bias in a high-stakes setting. *Proceedings of the Conference on Fairness, Accountability, and Transparency.*

Decision of the Beijing Internet Court in case. (2023). Jing 0491 Min Chu No. 11279. Retrieved from *BeijingInternetCourtCivilJudgment112792023.pdf*

Decision of the Copyright Review Board in case SR # 1-11743923581. Retrieved from https://www.copyright.gov/rulings-filings/review-board/docs/Theatre-Dopera-Spatial.pdf

Decision of the District Court of Delaware. Case 1:20-cv-00613-SB. Retrieved from https://www.ded.uscourts.gov/sites/ded/files/opinions/20-613_5.pdf

Decision of the High Court of Justice. (2023). Case IL-2023-000007, *Getty Images and others v. Stability AI*. Retrieved from https://www.judiciary.uk/judgments/getty-images-and-others-v-stability-ai/

Decision of the Prague City Court. (2023). Case 10 C 13/2023-16. Retrieved from *108cad3e-d9e8-454f-bfac-d58e1253c83a*

Del Rio-Chanona, R. M., Laurentsyeva, N., & Wachs, J. (2024). Large language models reduce public knowledge sharing on online Q&A platforms. *PNAS Nexus, 3*(9), 1–12. https://doi.org/10.1093/pnasnexus/pgae400

Deng, G., et al. (2024). PentestGPT: Evaluating and harnessing large language models for automated penetration testing. *Proceedings of the 33rd USENIX Security Symposium (USENIX Security 24).*

Deng, Y., et al. (2024). Multilingual Jailbreak Challenges in Large Language Models. In *The Twelfth International Conference on Learning Representations, ICLR 2024* (May 7–11, 2024, Vienna, Austria). OpenReview.net. https://openreview.net/forum?id=vESNKdEMGp

Dessart, F., Fernández Macías, E., & Gómez, E. (2025). Anticipating the impact of AI on occupations: A JRC methodology. *JRC Science for Policy Brief (JRC142580).*

Di Nuovo, E., Cartier, E., & De Longueville, B. (2024). Meet XLM-RLnews-8: Not just another sentiment analysis model. In *Natural Language Processing and Information Systems, 28th International Conference on Applications of Natural Language to Information Systems, NLDB 2024, Turin, Italy, June 25–27, 2024, Proceedings* (p. 1). Springer Science and Business Media Deutschland GmbH.

Ding, N., et al. (2024). Automating exploratory proteomics research via language models. *arXiv.* https://doi.org/10.48550/arXiv.2411.03743

District Court of Amsterdam. (2024). C/13/737170 / HA ZA 23-690. Retrieved from *ECLI:NL:RBAMS:2024:6563*

Doron, G., Genway, S., Roberts, M., & Jasti, S. (2025). Generative AI: Driving productivity and scientific breakthroughs in pharmaceutical R&D. *Drug Discovery Today, 30*(1), 104272.

Draft Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act) – C(2025) 884. https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act

Dulong De Rosnay, M., & Stalder, F. (2020). Digital Commons. *Internet Policy Review, 9*(4). https://doi.org/10.14763/2020.4.1530

EDPB Support Pool of Experts, Shrishak, K. (2024, March). Effective implementation of data subject rights, AI complex algorithms and effective data protection supervision.

Edwards, B. (2025, February 12). Sam Altman lays out roadmap for OpenAI's long-awaited GPT-5 model. *Ars Technica.* https://arstechnica.com/ai/2025/02/sam-altman-lays-out-roadmap-for-openais-long-awaited-gpt-5-model/

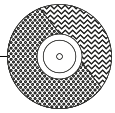
Enhancing Digital Health Innovation in the EU with Effective Industrial Strategy Policies – A Focus on Wearable Medical Devices (JRC138798).

Erduran, S. (2023). AI is transforming how science is done. Science education must reflect this change. *Science, 382*(6677), eadm9788. https://doi.org/10.1126/science.adm9788

Eriksson, M., Purificato, E., Noroozian, A., Vinagre, J., Chaslot, G., Gomez, E., & Fernandez-Llorca, D. (2025). Can we trust AI benchmarks? An interdisciplinary review of current issues in AI evaluation. *arXiv:2502.06559.*

Escobar-Planas, M., et al. (2025). Implementing and evaluating trustworthy conversational agents for children. In H. Plácido da Silva & P. Cipresso (Eds.), *Computer-Human Interaction Research and Applications: CHIRA 2024* (Vol. 2370, pp. unknown). Springer, Cham. https://doi.org/10.1007/978-3-031-82633-7_29

Esteva, A., & others. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118.



European Commission, Joint Research Centre. (2025). *Technology safeguards for the re-use of confidential data* (JRC141298). Ispra: European Commission.

European Commission. (2025). *Public sector tech watch: Analysis of the generative AI landscape in the European public sector*. Publications Office of the European Union. https://data.europa.eu/doi/10.2799/040981

European Commission. (n.d.). Data Centres Code of Conduct. https://e3p.jrc.ec.europa.eu/en/groups/data-centres-code-conduct

European Commission: Directorate-General for Communications Networks, Content and Technology, & Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji. (2019). *Ethics guidelines for trustworthy AI*. Publications Office. https://data.europa.eu/doi/10.2759/346720

European Commission: Directorate-General for Digital Services, Brizuela, A., Montino, C., Galasso, G., Polli, G., et al. (2024). *Adoption of AI, blockchain and other emerging technologies within the European public sector – A public sector Tech Watch report*. Publications Office of the European Union. https://data.europa.eu/doi/10.2799/3438251

European Commission: Directorate-General for Research and Innovation, Renda, A., Balland, P.-A., Soete, L., & Christophilopoulos, E. (2025). *A European model for artificial intelligence*. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/803464

European Data Protection Board Support Pool of Experts, Shrishak, K. (2024, March). Effective implementation of data subject rights, AI complex algorithms and effective data protection supervision.

European Data Protection Board. (2024). Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models.

European Data Protection Board. (2024, July 16). Statement 3/2024 on data protection authorities' role in the Artificial Intelligence Act framework.

European Data Protection Supervisor. (2023, September 25). Opinion 41/2023 on the proposal for a regulation on European Union labour market statistics on businesses.

European Parliament and the Council. (2019, June 20). Directive (EU) 2019/1024 on open data and the re-use of public sector information (recast). http://data.europa.eu/eli/dir/2019/1024/oj

European Parliament. (2022). *Understanding EU data protection policy*. Briefing. https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/698898/EPRS_BRI(2022)698898_EN.pdf

Fabiani et al. (2024). Strategic insights into the EU's advanced manufacturing industry: Trends and comparative analysis.

Falkenberg, M., Galeazzi, A., Torricelli, M., et al. (2022). Growing polarization around climate change on social media. *Nature Climate Change*, *12*, 1114–1121. https://doi.org/10.1038/s41558-022-01527-x

Farrell, E., Minghini, M., Kotsev, A., Soler Garrido, J., Tapsall, B., Micheli, M., Posada Sanchez, M., Signorelli, S., Tartaro, A., Bernal Cereceda, J., Vespe, M., Di Leo, M., Carballa Smichowski, B., Smith, R., Schade, S., Pogorzelska, K., Gabrielli, L., & De Marchi, D. (2023). *European Data Spaces - Scientific insights into data sharing and utilisation at scale* (EUR 31499 EN). Publications Office of the European Union. https://doi.org/10.2760/400188

Fatemi, M., et al. (2025). Concise reasoning via reinforcement learning. ArXiv.org. https://arxiv.org/abs/2504.05185

Fernández-López, J., Borrás-Rocher, F., Viuda-Martos, M., & Pérez-Álvarez, J. Á. (2024). Using artificial intelligence-based tools to improve the

literature review process: Pilot test with the topic "hybrid meat products." *Informatics*, *11*(4), 72. https://doi.org/10.3390/informatics11040072

Forthcoming JRC Policy Brief: Uses and perceptions of Generative AI in secondary education across five Member States.

Forthcoming JRC report: Scoping review of GenAI research in education studies.

Framework and Questionnaires for SMEs/PSOs: A guidance document for EDIHs. European Commission. (n.d.). JRC133234.

França, C. (2023). AI empowering research: 10 ways how science can benefit from AI. *arXiv*. https://doi.org/10.48550/arXiv.2307.10265

Fredrikson, M., et al. (2014). Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing. In K. Fu & J. Jung (Eds.), *Proceedings of the 23rd USENIX Security Symposium* (pp. 17–32). USENIX Association. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/fredrikson_matthew

Fredrikson, M., Jha, S., & Ristenpart, T. (2015). Model inversion attacks that exploit confidence information and basic countermeasures. In I. Ray, N. Li, & C. Kruegel (Eds.), *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (pp. 1322–1333). ACM. https://doi.org/10.1145/2810103.2813677

Fritz, J. (2024). The notion of 'authorship' under EU law—who can be an author and what makes one an author? An analysis of the legislative framework and case law. *Journal of Intellectual Property Law & Practice*, *19*(7), 552–556. https://doi.org/10.1093/jiplp/jpae022

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics*, *50*(3). https://doi.org/10.1162/coli_a_00524

Gao, J., & Wang, D. (2024). Quantifying the benefit of artificial intelligence for scientific research. *arXiv*. https://doi.org/10.48550/arXiv.2304.10578

Gardner, J., Brooks, C., & Baker, R. (2019). Evaluating the fairness of predictive student models through slicing analysis. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 225–234). https://doi.org/10.1145/3303772.3303791

Girdhar, R. (2025, January 30). LLMs can see and hear without any training. *arXiv*. https://arxiv.org/abs/2501.18096v1

Giudici, P., et al. (2024). Artificial intelligence risk measurement. *Expert Systems with Applications*, *235*, 121220. https://doi.org/10.1016/j.eswa.2023.121220

Goldberg, K., & Guthart, G. (2024). Augmented dexterity: How robots can enhance human surgical skills. *Science Robotics*, *9*(95), eadr5247.

Gomez, C. E., Sztainberg, M. O., & Trana, R. E. (2022). Curating cyberbullying datasets: A human-AI collaborative approach. *International Journal of Bullying Prevention*, *4*, 35–46. https://doi.org/10.1007/s42380-021-00114-6

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144.

Gottweis, J., & Natarajan, V. (2025, February 19). Accelerating scientific breakthroughs with an AI co-scientist. *Google Research Blog*. https://research.google/blog/accelerating-scientific-breakthroughs-with-an-ai-co-scientist/

Greco, F., Desolda, G., & Viganò, L. (2024). Supporting the design of phishing education, training and awareness interventions: An LLM-based approach. *CEUR Workshop Proceedings*, *3700*.

Griesinger, C. G., Reina, V., Panidis, D., & Chassaigne, H. (2025). Towards an evidence pathway for operationalizing trustworthy AI in health: An ontology to bridge the gap between ethical principles and fundamental concepts. (Submitted for publication).

Grimmelikhuijsen, S., & Tangi, L. (2024). What factors influence perceived artificial intelligence adoption by public managers. Publications Office of the European Union. https://data.europa.eu/doi/10.2760/0179285

Hall, P., & Ellis, D. (2023). A systematic review of socio-technical gender bias in AI algorithms. *Online Information Review*, *47*(7), 1264–1279. https://doi.org/10.1108/OIR-08-2021-0452

Hamed, A. A., Fandy, T. E., & Wu, X. (2024). Accelerating complex disease treatment through network medicine and GenAI: A case study on drug repurposing for breast cancer. In *2024 IEEE International Conference on Medical Artificial Intelligence* (MedAI). IEEE.

Hamon, R., Sanchez, I., Fernandez Llorca, D., & Gomez, E. (2024). Generative AI transparency: Identification of machine-generated content (JRC137136). European Commission. https://publications.jrc.ec.europa.eu/repository/handle/JRC137136

Han, Z., et al. (2024). Parameter-efficient fine-tuning for large models: A comprehensive survey. *arXiv:2403.14608* [*cs.LG*]. Retrieved from https://arxiv.org/abs/2403.14608

Harvard Business Review. (2024). *The uneven distribution of AI's environmental impacts*. https://hbr.org/2024/07/the-uneven-distribution-of-ais-environmental-impacts

Ho, J., et al. (2025). Gender biases within artificial intelligence and ChatGPT: Evidence, sources of biases and solutions. *Computers in Human Behavior: Artificial Humans, 4*, 100145. https://doi.org/10.1016/j.chbah.2025.100145

Holzinger, A., Saranti, A., Molnar, C., Biecek, P., & Samek, W. (2022). Explainable AI methods - A brief overview. In A. Holzinger, R. Goebel, R. Fong, T. Moon, K. R. Müller, & W. Samek (Eds.), *xxAI - Beyond explainable AI* (pp. unknown). Lecture Notes in Computer Science (Vol. 13200). Springer, Cham. https://doi.org/10.1007/978-3-031-04083-2_2

Howell, M. D. (2024). Generative artificial intelligence, patient safety and healthcare quality: A review. *BMJ Quality & Safety, 33*(11), 748–754. https://doi.org/10.1136/bmjqs-2023-016690

Hsu, J. (2023, October 12). Energy-efficient transistor could allow smartwatches to use AI. *New Scientist*. https://www.newscientist.com/article/2397235-energy-efficient-transistor-could-allow-smartwatches-to-use-ai/

Hu, E. J., et al. (2022). LoRA: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. Retrieved from https://openreview.net/forum?id=nZeVKeeFYf9

Hubinger, E., et al. (2024). Sleeper agents: Training deceptive LLMs that persist through safety training. *arXiv:2401.05566 [cs.CR]*. https://arxiv.org/abs/2401.05566

Huertas-García, A., et al. (2024). Camouflage is all you need: Evaluating and enhancing language model robustness against camouflage adversarial attacks. https://doi.org/10.48550/arXiv.2402.09874

Hulsman, R., et al. (2024). Conformal risk control for pulmonary nodule detection. *arXiv preprint arXiv:2412.20167.*

IEA. (2025). *Energy and AI*. https://www.iea.org/reports/energy-and-ai (Licence: CC BY 4.0)

Ifargan, T., Hafner, L., Kern, M., Alcalay, O., & Kishony, R. (2024). Autonomous LLM-driven research from data to human-verifiable

research papers. *arXiv.* https://doi.org/10.48550/arXiv.2404.17605

International Computer and Information Literacy Study. (n.d.). *ICILS*. https://op.europa.eu/en/publication-detail/-/publication/59721dc6-a0aa-11ef-85f0-01aa75ed71a1/language-en

Iuga, I. C., Nerişanu, R. A., & Iuga, H. (2024). The impact of healthcare system quality and economic factors on the older adult population: A health economics perspective. *Frontiers in Public Health, 12*, 1454699.

Jiang, L. Y., et al. (2023). Health system-scale language models are all-purpose prediction engines. *Nature, 619*, 357–362. https://doi.org/10.1038/s41586-023-06160-y

Kabir, A., Shah, S., Haddad, A., & Raper, D. M. S. (2025). Introducing our custom GPT: An example of the potential impact of personalized GPT builders on scientific writing. *World Neurosurgery, 193*, 461–468. https://doi.org/10.1016/j.wneu.2024.10.041

Kalpaka, A., Rissola, G., De Nigris, S., & Nepelski, D. (2023). Digital maturity assessment (DMA).

Kamiya, G., & Bertoldi, P. (2024). *Energy consumption in data centres and broadband communication networks in the EU*. Publications Office of the European Union. https://doi.org/10.2760/706491

Kaplan, J., et al. (2020). Scaling laws for neural language models. *arXiv.* https://arxiv.org/abs/2001.08361

Kessler, S. H., Mahl, D., Schäfer, M. S., & Volk, S. C. (2025). Science communication in the age of artificial intelligence. *JCOM, 24*(2), e. https://doi.org/10.22323/2.24020501

Khare, A., et al. (2023). Understanding the effectiveness of large language models in detecting security vulnerabilities. *arXiv preprint arXiv:2311.16169.*

Khlaif, Z. N., Mousa, A., Hattab, M. K., Itmazi, J., Hassan, A. A., Sanmugam, M., & Ayyoub, A. (2023). The potential and concerns of using AI in scientific research: ChatGPT performance evaluation. *JMIR Medical Education, 9*, e47049. https://doi.org/10.2196/47049

Kim, J. H., et al. (2022). Dataset condensation via efficient synthetic-data parameterization. *International Conference on Machine Learning*. PMLR.

King, J., & Meinhardt, C. (2024). Rethinking privacy in the AI era – Policy provocations for a data-centric world (White Paper, p. 17). Stanford University.

Koide, T., et al. (2024). Chatspamdetector: Leveraging large language models for effective phishing email detection. *arXiv preprint arXiv:2402.018093.*

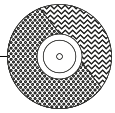
Koohi-Moghadam, M., & Bae, K. T. (2023). Generative AI in medical imaging: Applications, challenges, and ethics. *Journal of Medical Systems, 47*(1), 94.

Korinek, A. (2023). Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature, 61*(4), 1281–1317.

Kretschmer, M., Meletti, B., Bently, L., Cifrodelli, G., Eben, M., Erickson, K., Iramina, A., Li, Z., McDonagh, L., Perot, E., Porangaba, L., & Thomas, A. (2025). Copyright and AI: Response by the CREATe Centre to the UK Government's Consultation. *CREATe*. https://doi.org/10.5281/zenodo.14931964

Kurakin, A., Goodfellow, I., & Bengio, S. (2016). Adversarial machine learning at scale. https://doi.org/10.48550/arXiv.1611.01236

Kurita, K., Michel, P., & Neubig, G. (2020). Weight poisoning attacks on pre-trained models. https://arxiv.org/abs/2004.06660



Laestadius, L., Bishop, A., Gonzalez, M., Illenčík, D., & Campos-Castillo, C. (2024). Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika.

Laux, J., et al. (2023). Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance, 18*(1). https://doi.org/10.1111/rego.12512

Li, X., et al. (2024). DeepInception: Hypnotize large language model to be jailbreaker. *arXiv: 2311.03191 [cs.LG]*. https://arxiv.org/abs/2311.03191

Li, Y.-H., et al. (2024). Innovation and challenges of artificial intelligence technology in personalized healthcare. *Scientific Reports, 14*(1), 18994.

Liao, Z., & Sun, H. (2024). AmpleGCG: Learning a universal and transferable generative model of adversarial suffixes for jailbreaking both open and closed LLMs. *arXiv:2404.07921 [cs.CL]*. https://arxiv.org/abs/2404.07921

Lim, H., et al. (2025). Intelligent olfactory system utilizing in situ ceria nanoparticle-integrated laser-induced graphene. *ACS Nano*. https://doi.org/10.1021/acsnano.5c03601

Lindsey, J., et al. (2025, March 27). On the biology of a large language model. *Transformer Circuits*. https://transformer-circuits.pub/2025/attribution-graphs/biology.html

Liu, H., Zhou, Y., Li, M., Yuan, C., & Tan, C. (2025). Literature meets data: A synergistic approach to hypothesis generation. *arXiv*. https://doi.org/10.48550/arXiv.2410.17309

Liu, J. X., et al. (2017). Global health workforce labor market projections for 2030. *Human Resources for Health, 15*, 1–12.

Liu, X., et al. (2024). AutoDAN: Generating stealthy jailbreak prompts on aligned large language models. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. Retrieved from https://openreview.net/forum?id=7Jwpw4qKkb

Liu, X., et al. (2024). AutoDAN-Turbo: A lifelong agent for strategy self-exploration to jailbreak LLMs. https://doi.org/10.48550/arXiv.2410.05295

Longpre, S., et al. (2024). Consent in crisis: The rapid decline of the AI data commons. *arXiv*. https://doi.org/10.48550/arXiv.2407.14933

López Cobo, M., et al. (2019). Academic offer and demand for advanced profiles in the EU. *Artificial Intelligence, High Performance Computing and Cybersecurity (EUR 29629 EN)*. Publications Office of the European Union.

Luccioni, A. S., Strubell, E., & Crawford, K. (2025). From efficiency gains to rebound effects: The problem of Jevons' paradox in AI's polarized environmental debate. *arXiv preprint arXiv:2501.16548*.

Lv, H., et al. (2024). CodeChameleon: Personalized encryption framework for jailbreaking large language models. *arXiv:2402.16717 [cs.CL]*. https://arxiv.org/abs/2402.16717

Marcus, H. J., et al. (2024). The IDEAL framework for surgical robotics: Development, comparative evaluation and long-term monitoring. *Nature Medicine, 30*(1), 61-75.

Masanet, E., Shehabi, A., Lei, N., Smith, S., & Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science, 367*(6481), 984-986.

Medaglia, R., Mikalef, P., & Tangi, L. (2024). *Competences and governance practices for artificial intelligence in the public sector*. Publications Office of the European Union. https://data.europa.eu/doi/10.2760/7895569

Mehrotra, A., et al. (2024). Tree of attacks: Jailbreaking black-box LLMs automatically. In

A. Globersons et al. (Eds.), *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*. Retrieved from http://papers.nips.cc/paper%5C_files/paper/2024/hash/70702e8cbb4890b4a467b984ae59828a-Abstract-Conference.html

METR. (2024). *Evaluating AI models for critical harms*. Retrieved from https://metr.org/evaluating-ai-models-for-critical-harms.pdf

Meuser, T., et al. (2024). Revisiting edge AI: Opportunities and challenges. *IEEE Internet Computing, 28*(4), 49-59.

Meyerson, E. (2012, November 29). YouTube now: Why we focus on watch time. *YouTube Official Blog*.

Microsoft Research AI4Science, & Microsoft Azure Quantum. (2023). The impact of large language models on scientific discovery: A preliminary study using GPT-4 (arXiv:2311.07361). *arXiv*. https://doi.org/10.48550/arXiv.2311.07361

Microsoft. (2025, May 6). Boosting HR and IT services at Microsoft with our new employee self-service agent in Microsoft 365 Copilot. *Microsoft Inside Track*.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1-38. https://doi.org/10.1016/j.artint.2018.07.007

Mitchell, P. I. (2024). *Steps of the scientific method* (Technical report). Retrieved from https://www.dbu.edu/mitchell/medieval-resources/sciencemethodoverview.html

Moerel, L., & Storm, M. (n.d.). Do LLM's store personal data? This is asking the wrong question. *IAPP*. https://iapp.org/news/a/do-llms-store-personal-data-this-is-asking-the-wrong-question

Nazi, Z. A., & Peng, W. (2024). Large language models in healthcare and medical domain: A review. *Informatics, 11*(3).

NITI Aayog. (2018). *National strategy for artificial intelligence*. Retrieved from https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187-192.

NTT Data. (2025). *Global GenAI report: How organizations are mastering their GenAI destiny in 2025*. Retrieved from Global GenAI Report | NTT DATA

OECD. (2024). *Collective action for responsible AI in health*. Retrieved from https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/01/collective-action-for-responsible-ai-in-health_9a65136f/f2050177-en.pdf

OECD. (2025). What should teachers teach and students learn in a future of powerful AI? *OECD Education Spotlights*, No. 20. OECD Publishing. https://doi.org/10.1787/ca56c7d6-en

O'Sullivan, J. (2025, March 29). The case for AI illiteracy. *Substack*. https://substack.com/inbox/post/160133422

Owoahene Acheampong, I., & Nyaaba, M. (2024). Review of qualitative research in the era of generative artificial intelligence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4686920

Pan, S., Wang, T., Qiu, R. L., Axente, M., Chang, C. W., Peng, J., Fei, B., & Yang, X. (2023). 2D medical image synthesis using transformer-based denoising diffusion probabilistic model. *Physics in Medicine & Biology, 68*(10), 105004.

Park, K.-S., & Choi, H. (2024). How to harness the power of GPT for scientific research: A comprehensive review of methodologies,

applications, and ethical considerations. *Nuclear Medicine and Molecular Imaging, 58*(6), 323–331. https://doi.org/10.1007/s13139-024-00876-z

Partsol. (2025, March 31). World's first AI stem cell-engineered cognitive AI platform launched by Irish-based Partsol. *PR Newswire: Press Release Distribution, Targeting, Monitoring and Marketing*. https://www.prnewswire.com/news-releases/worlds-first-ai-stem-cell-engineered-cognitive-ai-platform-launched-by-irish-based-partsol-302415793.html

Pataranutaporn, P., et al. (2023). Influencing human–AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. *Nature Machine Intelligence, 5*, 1076.

Peel, M. (2025, March 18). Microsoft teams up with AI start-up to simulate brain reasoning. *Financial Times*. https://www.ft.com/content/37e44758-04a6-450b-abe3-f51f1d7d972a

Perez, E., et al. (2022). Red teaming language models with language models. In Y. Goldberg, Z. Kozareva, & Y. Zhang (Eds.), *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (pp. 3419-3448). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.225

Quintais, J. P. (2025). Generative AI, copyright and the AI Act. *Computer Law & Security Review, 56*, 106107. https://doi.org/10.1016/j.clsr.2025.106107

Radford, A., et al. (2018). Language models are unsupervised multitask learners. Retrieved from https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf

Rajpurkar, P., & others. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Medicine, 15*(11), e1002686.

Rajpurkar, P., et al. (2022). AI in health and medicine. *Nature Medicine, 28*(1), 31-38.

Raleigh, NC, USA, February 8-10, 2023. IEEE, 2023, pp. 339–364. https://doi.org/10.1109/SaTML54575.2023.00031

Ramachandran, K., Stewart, D., Hardin, K., & Crossan, G. (2024). As generative AI asks for more power, data centers seek more reliable, cleaner energy solutions. *Deloitte Center for Technology, Media & Telecommunications*.

Rawte, V., et al. (2023). The troubling emergence of hallucination in large language models - An extensive definition, quantification, and prescriptive remediations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 2541-2573). Association for Computational Linguistics. https://aclanthology.org/2023.emnlp-main.155/

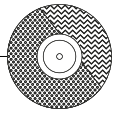
Reddy, S. (2024). Generative AI in healthcare: An implementation science informed translational path on application, integration and governance. *Implementation Science, 19*, 27. https://doi.org/10.1186/s13012-024-01357-9

Regulation (EU) 2020/852 of the European Parliament and of the Council of 18 June 2020 on the establishment of a framework to facilitate sustainable investment, and amending Regulation (EU) 2019/2088 (OJ L 198 22.06.2020).

Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/17.

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. https://eur-lex.europa.eu/eli/reg/2023/2854/oj

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) (Text with EEA relevance).

Rein, S., & Wierman, A. (2024). The uneven distribution of AI's environmental impacts.

Reina, V., & Griesinger, C. B. (2024). *Cyber security in the health and medicine sector: A study on available evidence of patient health consequences resulting from cyber incidents in healthcare settings*. European Commission: Joint Research Centre. https://publications.jrc.ec.europa.eu/repository/handle/JRC138692

Righi, R., et al. (2020). Academic offer of advanced digital skills in 2019-20. International comparison. Focus on Artificial Intelligence, High Performance Computing, Cybersecurity and Data Science (EUR 30351 EN). Publications Office of the European Union.

Roit, E., et al. (2023). Factually consistent summarization via reinforcement learning with textual entailment feedback. *arXiv*. Retrieved from http://arxiv.org/abs/2306.00186

Sala, A., Porcaro, L., & Gómez, E. (2024). Social media use and adolescents' mental health and well-being: An umbrella review. *Computers in Human Behavior Reports, 14*, 100404. https://doi.org/10.1016/j.chbr.2024.100404

Sallai, D., Cardoso-Silva, J., Barreto, M., Panero, F., Berrada, G., & Luxmoore, S. (2024). Approach generative AI tools proactively or risk bypassing the learning process in higher education. *LSE Public Policy Review, 3*(3), 7. https://doi.org/10.31389/lseppr.108

Sandfort, V., Yan, K., Pickhardt, P. J., & Summers, R. M. (2019). Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific Reports, 9*(1), 16884.

Santos, A. M., Molica, F., & Torrecilla-Salinas, C. (2025). EU-funded investment in Artificial Intelligence and regional specialization. *Regional Science Policy & Practice, 17*(7).

Saraswat, D., et al. (2022). Explainable AI for Healthcare 5.0: Opportunities and challenges. *IEEE Access, 10*, 1–10. https://doi.org/10.1109/access.2022.3197671

Schäfer, M. S., Kremer, B., Mede, N. G., & Fischer, L. (2024). Trust in science, trust in ChatGPT? How Germans think about generative AI as a source in science communication. *JCOM, 23*(09), A04. https://doi.org/10.22323/2.23090204

Schmidgall, S., et al. (2024). General-purpose foundation models for increased autonomy in robot-assisted surgery. *Nature Machine Intelligence*, 1–9.

Schmidgall, S., Su, Y., Wang, Z., Sun, X., Wu, J., Yu, X., Liu, J., Liu, Z., & Barsoum, E. (2025). Agent Laboratory: Using LLM Agents as Research Assistants. https://doi.org/10.48550/arxiv.2501.04227

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education, 57*(4), 620–631. https://doi.org/10.1111/ejed.12532

Selwyn, N. (2023). Constructive criticism? Working with (rather than against) the AIED backlash. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-023-00344-3

Shen, X., et al. (2024). "Do Anything Now": Characterizing and evaluating in-the-wild jailbreak prompts on large language models. In B. Luo et al. (Eds.), *Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1671–1685). ACM. https://doi.org/10.1145/3658644.3670388

Shokri, R., et al. (2017). Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP 2017)* (pp. 3–18). IEEE Computer Society. https://doi.org/10.1109/SP.2017.41

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature, 631*(8022), 755–759.

Signorelli, et al. (forthcoming). ATLAS: An Analytical Tool for Linking and Assessing Industrial EcoSystems.

Sollini, M., et al. (2019). Towards clinical application of image mining: A systematic review on artificial intelligence and radiomics. *European Journal of Nuclear Medicine and Molecular Imaging*. https://doi.org/10.1007/s00259-019-04372-x

Solove, D. (2025). Artificial intelligence and privacy. *Florida Law Review, 77*, 18.

Song, Z., Rehman, S. U., Ping Ng, C., Zhou, Y., Washington, P., & Verschueren, R. (2024). Do FinTech algorithms reduce gender inequality in bank loans? A quantitative study from the USA. *Journal of Applied Economics, 27*(1).

Stanceski, K., et al. (2024). The quality and safety of using generative AI to produce patient-centred discharge instructions. *npj Digital Medicine, 7*(1), 329.

Su, C., et al. (2025). Optimizing metabolic health with digital twins. *npj Aging, 11*(1), 20.

Su, E., et al. (2024). Extracting memorized training data via decomposition. *arXiv:2409.12367 [cs.LG]*. Retrieved from https://arxiv.org/abs/2409.12367

Sun, L., Chan, A., Chang, Y. S., & Dow, S. P. (2024). ReviewFlow: Intelligent scaffolding to support academic peer reviewing. *Proceedings of the 29th International Conference on Intelligent User Interfaces*, 120–137. https://doi.org/10.1145/3640543.3645159

Sun, Y., et al. (2024). AI hallucination: Towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanities and Social Sciences Communications, 11*, 1278.

Tang, B., et al. (2022). Wafer-scale solution-processed 2D material analog resistive memory array for memory-based computing. *Nature Communications, 13*(1), Article 3037. https://doi.org/10.1038/s41467-022-30519-w

Tangi, L., van Noordt, C., & Rodriguez Müller, A. P. (2023). The challenges of AI implementation in the public sector: An in-depth case studies analysis. In *Proceedings of the 24th Annual International Conference on Digital Government Research* (pp. 414–422). Association for Computing Machinery. https://doi.org/10.1145/3598469.3598516

Team, Lcm, et al. (2024). Large Concept Models: Language Modeling in a Sentence Representation Space. https://arxiv.org/pdf/2412.08821

Teo, Z. L., Quek, C. W. N., Wong, J. L. Y., & Ting, D. S. W. (2024). Cybersecurity in the generative artificial intelligence era. *Asia-Pacific Journal of Ophthalmology, 13*(4). https://doi.org/10.1016/j.apjo.2024.100091

Teutloff, O., Einsiedler, J., Kässi, O., Braesemann, F., Mishkin, P., & del Rio-Chanona, R. M. (2025). Winners and losers of generative AI: Early evidence of shifts in freelancer demand. *Journal of Economic Behavior & Organization*, 106845.

The Royal Society. (2024). *Science in the age of AI - How artificial intelligence is changing the nature and method of scientific research*.

Thirunavukarasu, A. J., et al. (2023). Large language models in medicine. *Nature Medicine, 29*(8), 1930–1940.

TNO. (2025). Fair machine learning combats biases. Retrieved from https://www.tno.nl/en/technology-science/technologies/fair-machine-learning/

Tolan, S., Pesole, A., Martínez-Plumed, F., Fernández-Macías, E., Hernández-Orallo, J., & Gómez, E. (2021). Measuring the occupational impact of AI: Tasks, cognitive abilities, and AI benchmarks. *Journal of Artificial Intelligence Research, 71*, 191–236.

Tomašev, N., et al. (2019). A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature, 572*(7767), 116–119.

Tomczyk, P., Brüggemann, P., & Vrontis, D. (2024). AI meets academia: Transforming systematic literature reviews. *EuroMed Journal of Business*. https://doi.org/10.1108/EMJB-03-2024-0055

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56.

Tramer, F., et al. (2016). Stealing machine learning models via prediction APIs. In T. Holz & S. Savage (Eds.), *Proceedings of the 25th USENIX Security Symposium* (pp. 601–618). USENIX Association. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/tramer

Tu, T., et al. (2024). Towards generalist biomedical AI. *NEJM AI, 1*(3), AIoa2300138.

Tuomi, I. (2022). Artificial intelligence, 21st century competences, and socio-emotional learning in education: More than high-risk? *European Journal of Education, 57*(4), 601–619. https://doi.org/10.1111/ejed.12531

Tuomi, I., Cachia, R., & Villar Onrubia, D. (2023). On the futures of technology in education: Emerging trends and policy implications (JRC134308). Publications Office of the European Union. https://doi.org/10.2760/079734

UNESCO & International Research Centre on Artificial Intelligence. (2024). Challenging systematic prejudices: An investigation into bias against women and girls in large language models. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000388971

UNESCO. (2022). Recommendation on the ethics of artificial intelligence (No. HS/BIO/PI/2021/1). https://unesdoc.unesco.org/ark:/48223/pf0000381137

Vaccari, C., & Chadwick, A. (2020). Deep fakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society, 6*(1), 1–13. https://doi.org/10.1177/2056305120903408

Van Hartskamp, M., Consoli, S., Verhaegh, W., Petkovic, M., & Van de Stolpe, A. (2019). Artificial intelligence in clinical health care applications: Viewpoint. *Journal of Medical Internet Research, 21*(4), e12100. https://doi.org/10.2196/12100

Van Noorden, R., & Perkel, J. M. (2023). AI and science: What 1,600 researchers think. *Nature, 621*.

Vassilev, A., Oprea, A., Fordyce, A., Anderson, H., Davies, X., & Hamin, M. (2025). Adversarial machine learning: A taxonomy and terminology of attacks and mitigations (NIST Trustworthy and Responsible AI No. NIST AI 100-2e2025). National Institute of Standards and Technology.

Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008. https://doi.org/10.5555/3295222.3295349

Vinuesa, R., Azizpour, H., Leite, I., et al. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications, 11*, 233. https://doi.org/10.1038/s41467-019-14108-y

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wachter, S. (2019). Data protection in the age of big data. *Nature Electronics, 2*, 6–7.

Wang, P., Zhang, L. Y., Tzachor, A., et al. (2024). E-waste challenges of generative artificial

intelligence. *Nature Computational Science, 4*, 818–823.

Wei, A., Haghtalab, N., & Steinhardt, J. (2023). Jailbroken: How does LLM safety training fail? In A. Oh (Ed.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10–16, 2023*. Retrieved from http://papers.nips.cc/paper/fd6613131889a4b656206c50a8bd7790-Abstract-Conference.html

Wei, J., et al. (2025, April 16). BrowseComp: A simple yet challenging benchmark for browsing agents. *arXiv*. https://arxiv.org/abs/2504.12516

Weidinger, L., et al. (2025). Toward an evaluation science for generative AI systems. *arXiv:2503.05336*.

Whelehan, D. F., et al. (2020). Medicine and heuristics: Cognitive biases and medical decision-making. *Irish Journal of Medical Science, 189*(4), 1477–1484. https://doi.org/10.1007/s11845-020-02235-1

Wiley. (2025). ExplanAItions—An AI study by Wiley.

Williamson, B. (2021). Google's plans to bring AI to education make its dominance in classrooms more alarming. *Fast Company*. https://www.fastcompany.com/90641049/google-education-classroom-ai

Williamson, B., Gulson, K. N., Perrotta, C., & Witzenberger, K. (2022). Amazon and the new global connective architectures of education governance. *Harvard Educational Review, 92*(2), 231–256.

Woodring, J., Perez, K., & Ali-Gombe, A. (2024). Enhancing privacy policy comprehension through privacy: A user-centric approach using advanced language models. *Computers & Security, 145*, 103997.

Wornow, M., et al. (2023). The shaky foundations of large language models and foundation models for electronic health records. *npj Digital Medicine, 6*, 135. https://doi.org/10.1038/s41746-023-00879-8

Wright, G. (2023). Scientific method definition. Technical report. Retrieved from https://www.techtarget.com/whatis/definition/scientific-method

Wu, Y., et al. (2024). Automated data visualization from natural language via large language models: An exploratory study. *Proceedings of the ACM on Management of Data, 2*(3), 1–28.

Xu, N., et al. (2024). Cognitive overload: Jailbreaking large language models with overloaded logical thinking. In K. Duh, H. Gómez-Adorno, & S. Bethard (Eds.), *Findings of the Association for Computational Linguistics: NAACL 2024* (pp. 3526–3548). Association for Computational Linguistics. https://doi.org/10.18653/V1/2024.FINDINGS-NAACL.224

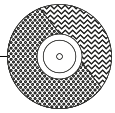
Xu, R., Sun, Y., Ren, M., Guo, S., Pan, R., Lin, H., Sun, L., & Han, X. (2024). AI for social science and social science of AI: A survey. *Information Processing & Management, 61*(3), 103665. https://doi.org/10.1016/j.ipm.2024.103665

Yamin, M. M., et al. (2024). Applications of LLMs for generating cyber security exercise scenarios. *IEEE Access*.

Yang, X., et al. (2022). A large language model for electronic health records. *npj Digital Medicine, 5*, 194. https://doi.org/10.1038/s41746-022-00742-2

Yang, Y., Zhang, H., Gichoya, J. W., Katabi, D., & Ghassemi, M. (2024). The limits of fair medical imaging AI in real-world generalization. *Nature Medicine, 30*(10), 2838–2848.

Yehudai, L., Eden, A., Li, G., Uziel, Y., Zhao, R., Bar-Haim, A., Cohan, M., & Shmueli-Scheuer, M. (2025). Survey on evaluation of LLM-based agents. *arXiv:2503.16416*.



Yu, J., & Qi, C. (2024). The impact of generative AI on employment and labor productivity. *Review of Business, 44*(1), 53–67.

Yu, J., et al. (2024). The shadow of fraud: The emerging danger of AI-powered social engineering and its possible cure. *arXiv preprint arXiv:2407.15912*.

Yuan, et al. (2024). GPT-4 Is Too Smart To Be Safe: Stealthy Chat with LLMs via Cipher. In *The Twelfth International Conference on Learning Representations, ICLR 2024*, Vienna, Austria, May 7–11, 2024. Retrieved from https://openreview.net/forum?id=MbfAK4s61A

Yuan, S., et al. (2025). Agent-R: Training language model agents to reflect via iterative self-training. *arXiv*. https://arxiv.org/abs/2501.11425

Zack, T., et al. (2024). Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: A model evaluation study. *The Lancet Digital Health, 6*(1), e12–e22. https://doi.org/10.1016/S2589-7500(23)00225-X

Zanfir-Fortuna, G. (2025). Why Data Protection legislation offers a powerful tool for regulating AI. *pdpEcho*. https://pdpecho.com/2025/02/26/why-data-protection-legislation-offers-a-powerful-tool-for-regulating-ai/

Zenil, H., et al. (2023). The future of fundamental science led by generative closed-loop artificial intelligence. *arXiv*. https://doi.org/10.48550/arXiv.2307.07522

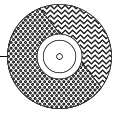
Zhang, Y., et al. (n.d.). The secret revealer: Generative model-inversion attacks against deep neural networks. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, Seattle, WA, USA, June 13–19, 2020. Computer Vision Foundation / IEEE, 2020, pp. 250–258. https://doi.org/10.1109/CVPR42600.2020.00033 Retrieved from https://openaccess.thecvf.com/content_CVPR_2020/html/Zhang_The_Secret_Revealer_Generative_Model-Inversion_Attacks_Against_Deep_Neural_Networks_CVPR_2020_paper.html

Zhou, L., et al. (2025). General Scales Unlock AI Evaluation with Explanatory and Predictive Power. *arXiv:2503.06378*.

Zhou, M., et al. (2024). Bias in Generative AI (Version 1). *arXiv*. https://doi.org/10.48550/arxiv.2403.02726

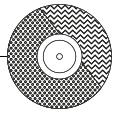
Zhu, R. J., et al. (2023). SpikeGPT: Generative pre-trained language model with spiking neural networks. https://doi.org/10.48550/arxiv.2302.13939

Zou, A., et al. (2023). Universal and transferable adversarial attacks on aligned language models. *arXiv:2307.15043 [cs.CL]*. Retrieved from https://arxiv.org/abs/2307.15043

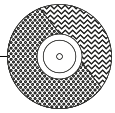


LIST OF ABBREVIATIONS AND DEFINITIONS

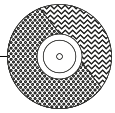
Accessibility	Extent to which products, systems, services, environments and facilities can be used by people from a population with the widest range of user needs, characteristics and capabilities to achieve identified goals in identified contexts of use (which includes direct use or use supported by assistive technologies).	HLEG AI, Assessment List for Trustworthy AI (ALTAI)
Accountability	This term refers to the idea that one is responsible for their action – and as a corollary their consequences – and must be able to explain their aims, motivations, and reasons. Accountability has several dimensions. Accountability is sometimes required by law. For example, the General Data Protection Regulation (GDPR) requires organisations that process personal data to ensure security measures are in place to prevent data breaches and report if these fail. But accountability might also express an ethical standard, and fall short of legal consequences.	Own elaboration based on: HLEG AI, Assessment List for Trustworthy AI (ALTAI)
Accountability (principle)	Data protection principle stipulating that a data controller (one who determines the purposes and means of processing of personal data) shall be responsible for, and be able to demonstrate compliance with data protection principles.	GDPR, Art.5(2)
Accuracy	The goal of an AI model is to learn patterns that generalize well for unseen data. It is important to check if a trained AI model is performing well on unseen examples that have not been used for training the model. To do this, the model is used to predict the answer on the test dataset and then the predicted target is compared to the actual answer. The concept of accuracy is used to evaluate the predictive capability of the AI model. Informally, accuracy is the fraction of predictions the model got right. A number of metrics are used in machine learning (ML) to measure the predictive accuracy of a model. The choice of the accuracy metric to be used depends on the ML task.	HLEG AI, Assessment List for Trustworthy AI (ALTAI)
Advanced AI Reasoning	Reasoning in artificial intelligence (AI) refers to the mechanism of using available information to generate predictions, make inferences and draw conclusions. It involves representing data in a form that a machine can process and understand, then applying logic to arrive at a decision.	What Is Reasoning in AI? | IBM
(Adversarial) Suffixes	Though nonsensical to humans, [adversarial suffixes] can manipulate strongly aligned LLMs into improperly responding to harmful prompts.	https://arxiv.org/html/2410.00451v2



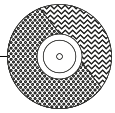
Adversarial Testing	Adversarial testing is a method for systematically evaluating an ML model with the intent of learning how it behaves when provided with malicious or inadvertently harmful input.	https://developers.google.com/machine-learning/guides/adv-testing
Adversial Attack	A malicious attempt which tries to perturb the input of a machine learning model (e.g. adding some noise imperceptible by humans) to cause the model to draw incorrect conclusions (e.g. a missclassification, or an error in the confidence of the classification).	Own elaboration, based on https://engineering.purdue.edu/ChanGroup/ECE595/files/chapter3.pdf and Goodfellow et all, 2015 (https://arxiv.org/pdf/1412.6572.pdf)
(AI) Training Dataset	Collections of data gathered for the purpose of training or fine-tuning machine learning and deep learning models. May for example consist of images, text, sound, and audiovisual content. AI training datasets are essential to AI development since they establish the boundaries of AI models and systems, and shape their capacity to recognize patterns, make predictions, and perform tasks.	Own elaboration
Algorithm	A formula or set of rules (or procedure, processes, or instructions) for solving a problem or for performing a task. In Artificial Intelligence, the algorithm tells the machine how to find answers to a question or solutions to a problem. In machine learning, systems use many different types of algorithms. Common examples include decision trees, clustering algorithms, classification algorithms, or regression algorithms.	"AI: A Glossary of Terms", Artificial Intelligence in Medical Imaging.
AI agent	AI agents are advanced AI systems designed to autonomously reason, plan, and execute complex tasks based on high-level goals.	https://www.nvidia.com/en-us/glossary/ai-agents/
Agentic AI	Class of AI systems that rely on one or more AI agents for their core functionality.	Own elaboration
Alignment (AI Alignment)	In the field of artificial intelligence (AI), alignment aims to steer AI systems toward a person's or group's intended goals, preferences, or ethical principles. An AI system is considered aligned if it advances the intended objectives. A misaligned AI system pursues unintended objectives.	https://en.wikipedia.org/wiki/AI_alignment
Article 29 Working Party (WP29)	A Working Party on the Protection of Individuals with regard to the Processing of Personal Data set up under the former Directive 95/46/EC. It was composed of a representative of the supervisory authority or authorities designated by each Member State and of a representative of the authority or authorities established for the Community institutions and bodies, and of a representative of the Commission.	Directive 95/46/EC, Art. 29



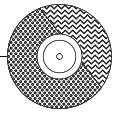
Artificial Intelligence	The capacity of computers or other machines to exhibit or simulate intelligent behaviour.	(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
Artificial Intelligence System	A machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.	AI act
Artificial Intelligence System Lifecycle	AI system lifecycle phases involve: i) 'design, data and models' which is a context-dependent sequence encompassing planning and design, data collection and processing, as well as model building ii) 'verification and validation' iii) 'deployment' and iv) 'operation and monitoring'. These phases often take place in an iterative manner and are not necessarily sequential. The decision to retire an AI system from operation may occur at any point during the operation and monitoring phase.	OECD
Attack	Attempt to destroy, expose, alter, disable, steal or gain unauthorized access to or make unauthorized use of an asset.	ISO/IEC 27000:2018(en)
Autonomy/ Autonomous	The ability of a person or artifact to govern itself including formation of intentions, goals, motivations, plans of action, and execution of those plans, with or without the assistance of other persons or systems.	(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
Availability	Property of being accessible and usable on demand by an authorized entity.	ISO/IEC 27000:2018
Balanced model	A balanced model maintains high performance during training, validation and testing.	Own elaboration
Benchmark	A set of tasks used to compare performance of several systems, models or components.	PR report
Bias	Bias is a systematic deviation from a true state. From a statistical perspective an estimator is biased when there is a systematic error that causes it to not converge to the true value that it is trying to estimate. In humans, bias can manifest itself in deviating perception, thinking, remembering or judgment which can lead to decisions and outcomes differing for people based on their membership to a protected group. There are different forms of bias, such as the subjective bias of individuals, data and algorithm bias, developer bias and institutionalized biases that are ingrained in the underlying societal context of the decision.	Tolan, 2018 https://arxiv.org/abs/1901.04730



	In science, computing, and engineering, a black box is a system which can be viewed in terms of its inputs and outputs (or transfer characteristics), without any knowledge of its internal workings. Its implementation is "opaque" (black).	
Black Box	In neural networking or heuristic algorithms (computer terms generally used to describe 'learning' computers or 'AI simulations'), a black box is used to describe the constantly changing section of the program environment which cannot easily be tested by the programmers. This is also called a "white box" in the context that the program code can be seen, but the code is so complex that it is functionally equivalent to a black box.	Wikipedia https://en.wikipedia.org/wiki/Black_box
Capability	Property of a systsem, usually a construct, that allows us to predict or explain performance.	PR report
Certainty / Uncertainty	Certainty: Dealing with entities that are entirely deterministic and certain. Uncertainty: Working with imperfect or incomplete information. There are many sources of uncertainty in a AI, including variance and noise in the specific data values, the (incomplete) sample of data collected from the domain, and in the imperfect nature of any models developed from such data.	Own elaboration
Chatbot	A computer program designed to simulate conversation with a human user, usually over the internet esp. one used to provide information or assistance to the user as part of an automated service.	Oxford English Dictionary
Computation	Computation is the integration of numerical simulation, mathematical modeling, algorithm development and other forms of quantitative analysis to solve problems that theorization, experimentation, and/or observation cannot.	(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
Compute	The amount of computational resources or computation needed to conduct the training of the ai model. Measured in floating point operations.	PR report
Confidentiality	property that information is not made available or disclosed to unauthorized individuals, entities, or processes.	ISO/IEC 27000:2018
Context	Refers to the information that enables the AI model to respond in a coherent and precise manner.	Own elaboration
Copyright	Copyright is a type of intellectual property that protects original works of authorship as soon as an author fixes the work in a tangible form of expression.	https://www.copyright.gov/what-is-copyright/
Court of Justice of the European Union (CJEU)	Interprets EU law to make sure it is applied in the same way in all EU countries, and settles legal disputes between national governments and EU institutions.	Treaty on European Union (TEU), Art. 19

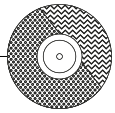


Cybersecurity	Things that are done to protect a person, organization, or country and their computer information against crime or attacks carried out using the internet.	https://dictionary.cambridge.org/dictionary/english/cybersecurity
Data	Any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audio-visual recording.	https://www.eu-data-act.com/Data_Act_Article_2.html
Data Annotation	The process of attaching a set of descriptive information to data without any change to that data. Note 1 to entry: The descriptive information can take the form of metadata, labels and anchors.	ISO/IEC DIS 22989(en). Terms related to Machine Learning
Data Condensation	Data condensation refers to the process of summarizing and simplifying a large amount of data into a more manageable form.	https://bcastudyguide.com/unit-1population-sample-and-data-condensation/
Data controller	The natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means [the "why" and the "how" personal data is processed] of the processing of personal data.	GDPR, Art. 4(7)
Data Leakage	Data leakage occurs when sensitive or confidential information is unintentionally exposed to unauthorized parties.	https://doi.org/10.1109/SP.2017.41
Data Poisoning	Data poisoning occurs when an adversarial actor attacks an AI system, and is able to inject bad data into the AI model's training set, thus making the AI system learn something that it should not learn. Examples show that in some cases these data poisoning attacks on neural nets can be very effective, causing a significant drop in accuracy even with very little data poisoning. Other kinds of poisoning attacks do not aim to change the behavior of the AI system, but rather they insert a backdoor, which is a data that the model's designer is not aware of, but that the attacker can leverage to get the AI system to do what they want.	HLEG AI, Assessment List for Trustworthy AI (ALTAI)
Data Protection Authority (DPA)	See "Supervisory Authority".	
Data Protection Impact Assessment (DPIA)	An assessment of the impact of the envisaged processing operations on the protection of personal data, to be carried out where a type of processing, in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons.	GDPR, Art. 35
Data quality	Data quality measures how well a dataset meets criteria for accuracy, completeness, validity, consistency, uniqueness, timeliness and fitness for the task.	Own elaboration from: https://www.ibm.com/think/topics/data-quality



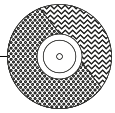
Data Sampling	The process to select a subset of data samples intended to present patterns and trends similar to that of the larger dataset (3.2.5) being analysed Note 1 to entry: Ideally, the subset of data samples will be representative of the larger dataset (3.2.5).	ISO/IEC DIS 22989(en). Terms related to Machine Learning
Data Subject Rights	A set of rights awarded by the General Data Protection Regulation to individuals whose personal data is processed by a data controller. These include the right of access to the individual's data, the right to rectification, the right to erasure, the right to restriction of processing, the right to data portability, the right to object, and the right not to be subject to any automated individual decision-making, including profiling.	GDPR, Articles 15 to 22
Database	The collection of organized according to a conceptual structure describing the characteristics of these data and the relationships among their corresponding entities, supporting one or more application areas.	ISO/IEC 20546:2019(en) Information technology - Big data - Overview and vocabulary
Datasets	Some AI systems rely on data sets to infer the logical mechanisms at play in the production of outcomes. Data sets are made of examples adapted to the task (e.g., pairs of inputs and labels for classification tasks), and are often divided into three parts, used in the three typical stages in the development of AI systems. Training set: used for learning, that is, fitting the learnable parameters of a model (e.g., the weights of a neural network), for example using optimization techniques. Validation set: used for validating the model, that is, in the context of AI development, providing an unbiased evaluation of the model after training and tuning the non-learnable parameters of the model and the learning process. The validation stage aims to prevent overfitting (the model begins to "memorize" training data rather than "learn" to generalize). The validation dataset can be a separate dataset or part of the training dataset, either as a fixed or variable split. Test set: used for testing the final model, that is, providing an independent evaluation of the model after training and validation. The test dataset must be independent from the training and validation datasets, that is, data in the test dataset should not be used in training or validation. Testing here is meant to be seen as an internal stage in the development of an AI system to ensure good performance and may not substitute the testing phase with regard to other obligations.	Own elaboration
Decision	An AI decision can be based on a prediction, a recommendation or a classification. It can also refer to a solely automated process, or one in which a human is involved.	ico.org.uk

Deep Fakes	Generated or manipulated image, audio or video content that appreciably resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful.	https://eur-lex.europa.eu/legal-content/
Deep Learning	<artificial intelligence> approach to creating rich hierarchical representations through the training (3.2.21) of neural networks (3.3.7) with many hidden layers.	ISO/IEC DIS 22989(en). Terms related to Neural Networks
Diffusion models	Diffusion models are generative models used primarily for image generation and other computer vision tasks. Diffusion-based neural networks are trained through deep learning to progressively "diffuse" samples with random noise, then reverse that diffusion process to generate high-quality images.	What are Diffusion Models? | IBM
Digital Commons	Digital information and technologies which are free and openly distributed and accessed. Examples include wikis, open-source software and licenses, online discussion forums, and digitized cultural heritage archives. Typically, content belonging to the digital commons is licensed through creative commons licenses or the GNU General Public License.	https://en.wikipedia.org/wiki/Digital_commons_(economics)
Digital Maturity	Digital maturity is a broad notion that covers a wide range of dimensions characterizing an entity.	Springer
Digital Technologies	Digital technologies refer to devices such as personal computers and tablets, tools such as cameras, calculators and digital toys, systems such as software and apps, augmented and virtual reality, and less tangible forms of technology such as the Internet.	https://digitalchild.org.au/defining-digital-technology/
Direct Prompt Injection	Direct prompt injections occur when a user's prompt input directly alters the behavior of the model in unintended or unexpected ways.	https://genai.owasp.org/llmrisk/llm01-prompt-injection
Discrimination	Differentiation for the purpose of separating persons to determine entitlements, rights, or eligibility.	(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
Disinformation	False information spread in order to deceive people.	https://dictionary.cambridge.org/dictionary/english/disinformation
Embeddings	A technique that allows machines to represent the meaning of words in such a way that complex relationships between words can be captured.	https://datos.gob.es/en/blog/understanding-word-embeddings-how-machines-learn-meaning-words

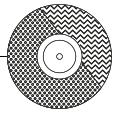


End User	An end-user is the person that ultimately uses or is intended to ultimately use the AI system. This could either be a consumer or a professional within a public or private organisation. The end-user stands in contrast to users who support or maintain the product, such as system administrators, database administrators, information technology experts, software professionals and computer technicians.	HLEG AI, Assessment List for Trustworthy AI (ALTAI)
Ethical AI	The development, deployment and use of AI that ensures compliance with ethical norms, including fundamental rights as special moral entitlements, ethical principles and related core values. It is the second of the three core elements necessary for achieving Trustworthy AI.	HLEG AI, Ethics Guidelines for Trustworthy AI
European Data Protection Board (EDPB)	EU body, composed of the head of one supervisory authority of each Member State and of the EDPS, or their respective representatives, tasked with ensuring the consistent application of the General Data Protection Regulation.	GDPR, Articles 68 to 70
Evaluation	A procedure to determine the value and qualitieis (capabilities, risks, etc.) of a system, model or component.	PR report
Explainability	Feature of an AI system that is intelligible to non-experts. An AI system is intelligible if its functionality and operations can be explained non technically to a person not skilled in the art.	HLEG AI, Assessment List for Trustworthy AI (ALTAI)
Fairness	A variety of ideas known as equity, impartiality, egalitarianism, non-discrimination and justice. Fairness embodies an ideal of equal treatment between individuals or between groups of individuals. This is what is generally referred to as 'substantive' fairness. But fairness also encompasses a procedural perspective, that is the ability to seek and obtain relief when individual rights and freedoms are violated.	HLEG AI, Assessment List for Trustworthy AI (ALTAI)
Fine-Tuning	To make very small changes to something in order to make it work as well as possible.	https://dictionary.cambridge.org/dictionary/english/fine-tune
Fingerprinting	Fingerprinting, or "fingerprinting", is a probabilistic technique designed to uniquely identify a user on a website or mobile application using the technical characteristics of their browser.	https://www.cnil.fr/fr/definition/fingerprinting
Free software	Free and open-source software (FOSS) is software available under a license that grants users the right to use, modify, and distribute the software – modified or not – to everyone free of charge.	https://www.gnu.org/philosophy/floss-and-foss.en.html https://en.wikipedia.org/wiki/Free_and_open-source_software

General Data Protection Regulation (GDPR)	Regulation (EU) 2016/679; the EU legal act laying down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data.	GDPR, Art. 1(1);
General Purpose AI (GPAI) / Foundational models	Also known as "foundational models". Large AI models trained on a vast quantity of data (generally unlabeled data and using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks.	Center for Research on Fundation Models (CRFM), "On the Opportunities and Risks of Foundation Models"
General Purpose AI System	An AI system which is based on a general-purpose AI model and which has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems.	AI act
Generative Adversarial Networks (GANs)	Generative Adversarial Networks, or GANs for short, are an approach to generative modeling using deep learning methods, such as convolutional neural networks. Generative modeling is an unsupervised learning task in machine learning that involves automatically discovering and learning the regularities or patterns in input data in such a way that the model can be used to generate or output new examples that plausibly could have been drawn from the original dataset.	What does "Generative Adversarial Network (GAN)" mean? – Legal definition – CyberLaws
Generative AI	Generative AI is a subset of AI that uses specialised machine learning models designed to produce a wide and general variety of outputs, capable of a range of tasks and applications, such as generating text, image or audio.	Adapted from: https://www.edps.europa.eu/system/files/2024-06/24-06-03_genai_orientations_en.pdf
Governance	"The process of collective decisionmaking and policy implementation, used distinctly from government to reflect broader concern with norms and processes relating to the delivery of public goods" (McLean and McMillan 2016).	(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
GPT	Generative pretrained transformers (GPTs) are a family of large language models (LLMs) based on a transformer deep learning architecture.	What is GPT (generative pre-trained transformer)? | IBM
Guardrail	A safeguard that is put in place to prevent (AI) from causing harm…created to keep people safe and guide position outcomes.	https://www.techopedia.com/definition/ai-guardrail
Hallucination	Phenomena where AI algorithms invent information that sounds plausible but is not factual.	https://curia.europa.eu/jcms/upload/docs/application/pdf/2023-11/cjeu_ai_strategy.pdf
High-Risk	Applied to AI systems or models that are likely to negatively affect health, safety or fundamental rights. High risk, in the EU AI Act, is placed between unacceptable risk, and hence forbidden, and limited risk, with some requirements of transparency. Most of the analysis and regulation focuses on high-risk systems.	PR report

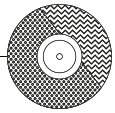


Human Oversight	Human oversight helps ensure that an AI system does not undermine human autonomy or causes other adverse effects.	HLEG AI, Assessment List for Trustworthy AI (ALTAI)
Human-Centric AI	The human-centric approach to AI strives to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights, including those set out in the Treaties of the European Union and Charter of Fundamental Rights of the European Union, all of which are united by reference to a common foundation rooted in respect for human dignity, in which the human being enjoy a unique and inalienable moral status. This also entails consideration of the natural environment and of other living beings that are part of the human ecosystem, as well as a sustainable approach enabling the flourishing of future generations to come.	HLEG AI, Ethics Guidelines for Trustworthy AI
Hyperparameter	<machine learning> characteristic of a machine learning algorithm (3.2.10) that affects its learning process Note 1 to entry: Hyperparameters are selected prior to training and can be used in processes to help estimate model parameters. Note 2 to entry: Examples of hyperparameters include number of network layers, width of each layer, type of activation function, optimization method, learning rate for neural networks the choice of kernel function in a support vector machine number of leaves or depth of a tree the K for K-means clustering the maximum number of iterations of the expectation maximization algorithm the number of Gaussians in a Gaussian mixture.	ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence
Indirect Prompt Injection	Indirect prompt injections occur when an LLM accepts input from external sources, such as websites or files. The content may have in the external content data that when interpreted by the model, alters the behavior of the model in unintended or unexpected ways.	https://genai.owasp.org/llmrisk/llm01-prompt-injection
Inference	The step in which a system generates an output from its inputs, typically after deployment.	OECD publishing: EXPLANATORY MEMORANDUM ON THE UPDATED OECD DEFINITION OF AN AI SYSTEM
Intellectual Property (IP)	Someone's idea, invention, creation, etc., that can be protected by law from being copied by someone else	https://dictionary.cambridge.org/dictionary/english/intellectual-property
Integrity	Property of accuracy and completeness	ISO/IEC 27000:2018



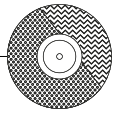
Interpretability	Interpretability refers to the concept of comprehensibility, explainability, or understandability. When an element of an AI system is interpretable, this means that it is possible at least for an external observer to understand it and find its meaning.	HLEG AI, Assessment List for Trustworthy AI (ALTAI)
Jailbreak Prompts	Jailbreaking in the context of language models refers to techniques used to bypass a model's built-in restrictions or safety protocols. These restrictions are generally in place to prevent harmful, unethical, or unsafe content generation. Jailbreaking prompts are crafted to trick the model into ignoring its guidelines, allowing the model to respond in ways it normally wouldn't, such as giving controversial opinions, providing restricted information, or performing unfiltered actions.	https://oecd.ai/en/about-air
Knowledge management	The collection, storage, curation, dissemination, archiving and destruction of documents, images, drawings and others sources of information.	https://www.apm.org.uk/book-shop/apm-body-of-knowledge-8th-edition/
Label	<machine learning> the target variable assigned to a sample.	ISO/IEC DIS 22989(en). Terms related to Machine Learning
Lawfulness (principle)	Data protection principle stipulating that any processing of personal data needs to be made in a lawful manner, in accordance with specific provisions foreseen in article 6 GDPR.	GDPR, Articles 5(1) and 6
Legitimate interest	Legal basis stipulated in article 6(1)(f) GDPR which is applicable to processing that necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.	GDPR, Article 6(1)(f)
Literacy	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.	AI act
LLM	A model that captures the distribution of a language, such as Catalan, or several languages at a time, natural or artificial, such as English and Python, usually expressed as a stochastic model assigning probabilities to next words or tokens, given the previous text. These probabilities can be used to generate text.	PR report
Machine Learning	The process of optimizing model parameters (3.1.28) through computational techniques, such that the model's behaviour reflects the data or experience.	ISO/IEC DIS 22989(en). Terms related to Machine Learning

	A mathematical construct that generates an inference (3.1.22), or prediction (3.2.12), based on input data.	
Machine Learning model	Note 1 to entry: A machine learning model results from training based on a machine learning algorithm (3.2.10). EXAMPLE: If a univariate linear function (y = θ0 + θ1x) has been trained using linear regression, the resulting model can be y = 3 + 7x.	ISO/IEC DIS 22989(en). Terms related to Machine Learning
Malicious Actor	See “Threat Actor”.	
Membership Inference Attack	Membership inference attacks occur when an attacker manipulates the model’s training data in order to cause it to behave in a way that exposes sensitive information.	https://owasp.org/www-project-machine-learning-security-top-10/docs/ML04_2023-Membership_Inference_Attack.html
Metadata	A structured description of the contents or the use of data facilitating the discovery or use of that data.j	https://www.eu-data-act.com/Data_Act_Article_2.html
Mitigation	Limitation of any negative consequence of a particular incident.	ISO 22300:2021(en)
Model Extraction / Model Theft	Unauthorized access and exfiltration of LLM models by malicious actors or APTs. This arises when the proprietary LLM models (being valuable intellectual property), are compromised, physically stolen, copied or weights and parameters are extracted to create a functional equivalent.	https://genai.owasp.org/llmrisk2023-24/llm10-model-theft
Model Inversion	Model inversion attacks occur when an attacker reverse-engineers the model to extract information from it.	https://owasp.org/www-project-machine-learning-security-top-10/docs/ML03_2023-Model_Inversion_Attack
Model Poisoning	A security threat where an attacker manipulates training data or the learning process to compromise an AI model’s behavior while potentially maintaining model accuracy on normal inputs.	https://www.gpt-privacy.com/dictionary/model-poisoning
Multi-modal AI	Multimodal AI refers to artificial intelligence systems that are able to process and integrate information from multiple types of input data, such as text, images, audio and video (referred to as modalities), to produce more comprehensive and nuanced outputs. Traditional AI models typically focus on a single modality, such as text-based natural language processing (NLP)[i] or image recognition. In contrast, multimodal AI systems combine different types of data to enable more sophisticated and versatile interactions.	Multimodal artificial intelligence | European Data Protection Supervisor

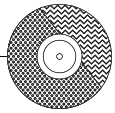


Multi-Omics	The “omics” notion refers to the fact that all or nearly all instances of the targeted molecular space are measured in the assay, and therefore they provide holistic views of the biological system. Initially, omics experiments used to concentrate on one type of assay (i.e. transcriptomics) and provide single-omics data. However, more recently researchers have combined multiple assays from the same set of samples to create multi-omics datasets.	https://www.nature.com/articles/s41597-019-0258-4
Natural Language	language which is or was in active use in a community of people, and the rules of which are mainly deduced from the usage. Note 1 to entry: Natural language is any human language, which can be expressed in text, speech, sign language etc. Note 2 to entry: Natural language is any human language, such as English, Spanish, Arabic, Chinese, or Japanese, to be distinguished from programming and formal languages, such as Java, Fortran, C++, or First-Order Logic.	ISO/IEC DIS 22989(en). Terms related to Natural Language Processing
Natural Language Processing (NLP)	<system> information processing based upon natural language understanding (3.5.11) and natural language generation (3.5.8). <discipline> discipline concerned with the way computers process natural language data.	ISO/IEC DIS 22989(en). Terms related to Natural Language Processing
Neural Network (NN) / Artifical Neural Network (ANN)	Network of two or more layers of neurons (3.3.8) connected by weighted links with adjustable weights, which takes input data and produces an output. Note 1 to entry: Whereas some neural networks are intended to simulate the functioning of biological neurons in the nervous system, most neural networks are used in artificial intelligence as realizations of the connectionist model (3.1.12).	ISO/IEC DIS 22989(en). Terms related to Neural Networks
Online Platform	Means a provider of a hosting service which, at the request of a recipient of the service, stores and disseminates to the public information, unless that activity is a minor and purely ancillary feature of another service and, for objective and technical reasons cannot be used without that other service, and the integration of the feature into the other service is not a means to circumvent the applicability of this Regulation.	REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC
Open License	A license which grants permission to freely use, modify and share copyright protected works. Examples includes Creative Commons Licenses (CC), open source licenses, and open data licenses. Open licenses exist in many different forms and may for example grant full or partial permission to reuse works for commercial purposes, or impose conditions and requirements for making attributions to original authors.	https://en.wikipedia.org/wiki/Open-source_license

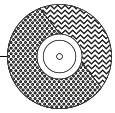
Open Source	Source code that is produced in a decentralized manner and made freely and openly available for reuse, modification, and distribution. Is commonly licensed through open licenses.	https://en.wikipedia.org/wiki/Open_source
(Model) Parameter	<machine learning> internal variable of a model (3.1.26) that affects how it computes its outputs. Note 1 to entry: Examples of parameters include the weights in a neural network, or the transition probabilities in a Markov model.	ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence
Performance	Measurable result. Note 1 to entry: Performance can relate either to quantitative or qualitative findings. Note 2 to entry: Performance can relate to managing activities, processes, products (including services), systems or organizations.	ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence
Personally identifiable data	Information which can be linked to a single person.	(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
Personal Data	Any information relating to an identified or identifiable natural person. An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. Sensitive Personal Data are personal data, revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership data concerning health or sex life and sexual orientation genetic data or biometric data. Its processing is prohibited, unless the derrogations foreseen in the Regulation apply.	Personal data - Regulation (EU) 2016/679 (General Data Protection Regulation), Article 4(1); Sensitive Personal data - Regulation (EU) 2016/679 (General Data Protection Regulation), Article 8(1);
Personalized learning paths/ experiences	Services that are tailored to individual users' interests and preferences.	https://www.sciencedirect.com/topics/computer-science/personalized-service
Pre-trained model	A pre-trained model is a machine learning (ML) model that has been trained on a large dataset and can be fine-tuned for a specific task. Pre-trained models are often used as a starting point for developing ML models, as they provide a set of initial weights and biases that can be fine-tuned for a specific task.	Pre Trained Model Definition | Encord
Prediction	<machine learning> output of a machine learning model when provided with input data.	ISO/IEC DIS 22989(en). Terms related to Machine Learning



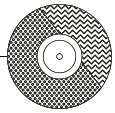
Privacy	"The protection of select information through the use of mechanical or statistical masking mechanisms for the purpose of protecting individual or group dignity, desire for seclusion or concealment, property, secrets, or freedom of choice".	(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
Privacy (Right to) (also known as right to a private life)	The fundamental right stipulating that everyone has the right to respect for his or her private and family life, home and communications.	Charter of Fundamental Rights of the European Union, article 7
Processing (of personal data)	Any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.	Regulation (EU) 2016/679 (General Data Protection Regulation), Article 4(2)
Prompt Injection	Prompt injection is a method used to manipulate an LLM's behavior by embedding specific instructions within a prompt. This approach exploits the model's tendency to follow instructions within the prompt sequence, even if those instructions are unintended or malicious. Prompt injection can be used to alter the model's response style, retrieve hidden or restricted data, or disrupt intended interactions.	https://oecd.ai/en/about-air
Provider	Means a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge.	AI act
Public Domain	Creative works which are exempt from intellectual property rights. For example, this may be the case since no one holds exclusive rights, or because rights have expired, been forfeited, or explicitly waived. As a result, the works can be legally used by anyone. Examples include the works of Cervantes, William Shakespeare, and Leonardo da Vinci.	https://en.wikipedia.org/wiki/Public_domain
Public Value	Public Value is value for and from the public. The new look associated with Public Value (PV) is viewing impacts on values in society as value creation. This perspective puts concurrent ideas of "public interest," "common good" or "common welfare" into a more managerial and entrepreneurial perspective.	Springer.com



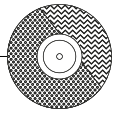
Reasoning Model	Reasoning models are designed to break down complex problems into smaller, manageable steps and solve them through explicit logical reasoning (This step is also called "thinking"). Unlike general-purpose LLMs which might generate direct answers, reasoning models are specifically trained to show their work and follow a more structured thought process.	https://techcommunity.microsoft.com/blog/azuredevcommunityblog/how-reasoning-models-are-transforming-logical-ai-thinking/4373194
Recommendation	A suggestion that something is good or suitable for a particular purpose or job.	https://dictionary.cambridge.org/dictionary/english/recommendation
Recommender System	Fully or partially automated system used by an online platform to suggest in its online interface specific information to recipients of the service, including as a result of a earch initiated by the recipient or otherwise determining the relative order or prominence of information displayed.	REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC
Red Teaming	Red teaming is the practice whereby a red team or independent group challenges an organisation to improve its effectiveness by assuming an adversarial role or point of view. It is often used to help identify and address potential security vulnerabilities.	HLEG AI, Assessment List for Trustworthy AI (ALTAI)
Reinforcement Learning from Human Feedback	A common mechanism in large language models and other kinds of AI models that alters the generated content to make the model more instructable, agreeable, safe or palatable by using human feedback.	PR report
Reliability	Property of consistent intended behaviour and results.	ISO/IEC DIS 22989(en). Terms related to Trustworthiness
Reproducibility	(Different team, different experimental setup) The measurement can be obtained with stated precision by a different team and a different measuring system, in a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using artifacts that they develop completely independently.	Association for Computing Machinery (ACM)
Resilience	The ability of a system to recover operational condition quickly following an incident.	ISO/IEC DIS 22989(en). Terms related to Trustworthiness
Responsibility	Capability of fulfilling an obligation or duty The quality of being reliable or trustworthy The state or fact of being accountable for actions Liability for some action.	(Ordinary Languege) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
Retrieval-Augmented Generation	Retrieval-augmented generation is a technique for enhancing the accuracy and reliability of generative AI models with information fetched from specific and relevant data sources.	What Is Retrieval-Augmented Generation aka RAG | NVIDIA Blogs



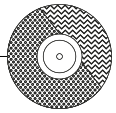
Risk	Possible loss or harm.	(Ordinary Languege) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
Robustness AI	Robustness of an AI system encompasses both its technical robustness (appropriate in a given context, such as the application domain or life cycle phase) and as well as its robustness from a social perspective (ensuring that the AI system duly takes into account the context and environment in which the system operates). This is crucial to ensure that, even with good intentions, no unintentional harm can occur. Robustness is the third of the three components necessary for achieving Trustworthy AI.	HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL
Safety	Prevention of accidents.	(Ordinary Languege) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
Sandbox	A controlled framework set up by a competent authority which offers providers or prospective providers of AI systems the possibility to develop, train, validate and test, where appropriate in real-world conditions, an innovative AI system, pursuant to a sandbox plan for a limited time under regulatory supervision.	AI act
Stakeholders	By stakeholders we denote all those that research develop, design, deploy or use AI, as well as those that are (directly or indirectly) affected by AI – including but not limited to companies, organisations, researchers, public services, institutions, civil society organisations, governments, regulators, social partners, individuals, citizens, workers and consumers.	HLEG AI, Ethics Guidelines for Trustworthy AI
Subject	For the purpose of real-world testing, means a natural person who participates in testing in real-world conditions.	AI act
Supervisory Authority (for data protection)	Independent public authority responsible for monitoring the application of Regulation (EU) 2016/679 (GDPR), in order to protect the fundamental rights and freedoms of natural persons in relation to processing and to facilitate the free flow of personal data within the Union.	GDPR, Article 51(1)
Supply Chain Attack	A supply chain attack uses third-party tools or services — collectively referred to as a "supply chain" — to infiltrate a target's system or network.	https://www.cloudflare.com/learning/security/what-is-a-supply-chain-attack
Synthetic data	AI generated data which is created to mimic the characteristics of data made by humans. Exists in all modalities (such as text, images, and sound) and can for example be used to train AI models, test hypothesis, or evaluate AI model performance.	Own elaboration / https://en.wikipedia.org/wiki/Synthetic_data



System Prompt	System prompts define how the AI should behave across all interactions, establishing the tone, ethical guidelines, and general approach.	https://www.regie.ai/blog/user-prompts-vs-system-prompts
Target	The target variable is the feature of a dataset that you want to understand more clearly. It is the variable that the user would want to predict using the rest of the data in the dataset.	https://ai-terms-glossary.com/item/target-variable/
Text generation	Task (3.1.37) of converting data carrying semantics into natural language (3.5.7)	ISO/IEC DIS 22989(en). Terms related to Natural Language Processing
Threat	Potential cause of an unwanted incident, which can result in harm to a system or organization.	ISO/IEC 27000:2018
Threat Actor	Threat actors, also known as cyberthreat actors or malicious actors, are individuals or groups that intentionally cause harm to digital devices or systems. Threat actors exploit vulnerabilities in computer systems, networks and software to perpetuate various cyberattacks, including phishing, ransomware and malware attacks.	https://www.ibm.com/think/topics/threat-actor
Training Data	A subset of input data samples used to train a machine learning model.	ISO/IEC DIS 22989(en). Terms related to Machine Learning
Transformer	A neural network that learns context and thus meaning by tracking relationships in sequential data like the words in this sentence.	https://blogs.nvidia.com/blog/what-is-a-transformer-model/
Transparency	Easily seen through, recognized, understood, or detected (OED). Sufficient illumination to confer comprehension.	(Ordinary Languege) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
Trustworthy AI	Trustworthy AI has three components: (1) it should be lawful, ensuring compliance with all applicable laws and regulations (2) it should be ethical, demonstrating respect for, and ensure adherence to, ethical principles and values and (3) it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm. Trustworthy AI concerns not only the trustworthiness of the AI system itself but also comprises the trustworthiness of all processes and actors that are part of the AI system's life cycle.	HLEG AI, Assessment List for Trustworthy AI (ALTAI)
Unseen Dataset	See "Testing Data".	
User	A natural or legal person that owns a connected product or to whom temporary rights to use that connected product have been contractually transferred, or that receives related services.	https://www.eu-data-act.com/Data_Act_Article_2.html



Variational Autoencoders	Variational autoencoders (VAEs) are generative models used in machine learning (ML) to generate new data in the form of variations of the input data they're trained on. In addition to this, they also perform tasks common to other autoencoders, such as denoising.	What is a Variational Autoencoder? | IBM
Validation	Like all autoencoders, variational autoencoders are deep learning models composed of an encoder that learns to isolate the important latent variables from training data and a decoder that then uses those latent variables to reconstruct the input data.	(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
Vulnerability	Weakness of an asset or control that can be exploited by one or more threats.	ISO/IEC 27000:2018
Watermarking	The process of embedding into the output of an artificial intelligence model a recognisable and unique signal (i.e. the watermark) that serves to identify the content as AI-generated. In practice, AI watermarking creates a unique identifiable signature that is invisible to humans but algorithmically detectable and that can be traced back to the AI model. Different watermarking techniques have been developed for text, image, video and audio content.	https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI(2023)757583_EN.pdf
Web-Scraping	An automated system for browsing and collecting data from the internet, using tools such as "bots" or "crawlers". Is currently central to the gathering of data for AI training datases, and is for example also used for market surveillance purposes and news monitoring.	Own elaboration
Wiki	An online, hypertext publication which can be collaboratively edited and is non-hierarchically managed by collaborators who can access and edit the publication through a web browser. A wiki can either be fully open for anyone to contribute, or shared within a limited group or organization.	https://en.wikipedia.org/wiki/Wiki
(Model) Workflow	The workflow of an AI model shows the phases needed to build the model and their interdependencies. Typical phases are: Data collection and preparation, Model development, Model training, Model accuracy evaluation, Hyperparameters' tuning, Model usage, Model maintenance, Model versioning. These stages are usually iterative: one may need to reevaluate and go back to a previous step at any point in the process.	HLEG AI, Assessment List for Trustworthy AI (ALTAI)



LIST OF FIGURES

LIST OF TABLES

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:
— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
— at the following standard number: +32 22999696,
— via the following form: european-union.europa.eu/contact-eu/write-us_en.

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

Open data from the EU

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

Scan the QR code to visit:
The Joint Research Centre: EU Science Hub
https://joint-research-centre.ec.europa.eu

Publications Office
of the European Union