Scientific Advice Mechanism
to the European Commission

April 2024

SAPEA evidence review report

Successful and timely uptake of artificial intelligence in science in the EU

- SAPEA. (2024). *Successful and timely uptake of artificial intelligence in science in the EU: Evidence review report*. Berlin: SAPEA.
- DOI 10.5281/zenodo.10849580
- Downloadable from https://scientificadvice.eu/advice/artificial-intelligence-in-science/

Version history

Version	Date	Summary of changes
1.0	15 April 2024	First published version

Scientific Advice Mechanism
to the European Commission

Successful and timely uptake of artificial intelligence in science in the EU

April 2024

SAPEA evidence review report

Table of contents

Foreword

Artificial intelligence (AI) has emerged as a revolutionary tool penetrating all parts of society, including science, with a remarkable acceleration in the past two years. AI brings forward powerful opportunities but also challenges and risks, emphasising the importance of adapted policies.

At this crucial time, the EU AI Act promises to be the first-ever legal framework on AI, positioning the continent in a leading role to address the risks of the technology. However, at the time of publication of this report, it appears that many of the challenges of AI used in scientific research will not fall under the regulations of the AI Act.

AI applications are rapidly permeating scientific research in practically all fields, accelerating scientific discovery and innovation through colossal datasets and access to extremely powerful computing infrastructures. Parallel to the enormous opportunities of AI for science, discussions and debates are ongoing in universities, weighing opportunities against the potential risks of AI for the reliability, reproducibility and transparency of scientific production and our future knowledge base.

In July 2023, the College of Commissioners asked the Scientific Advice Mechanism to the European Commission to provide evidence-based advice on how to accelerate a responsible uptake of AI in science. To address this question, SAPEA assembled an independent, international, and interdisciplinary working group of leading experts in the field, nominated by and selected from European academies and their respective networks. Between October 2023 and January 2024, the working group reviewed and compiled the latest evidence on the subject to create this thirteenth SAPEA Evidence Review Report. This report informs the accompanying Scientific Opinion of the Group of Chief Scientific Advisors, which contains the requested policy recommendations.

This project was coordinated by Euro-CASE acting as the lead network on behalf of SAPEA. We warmly thank all working group members for their voluntary contributions and dedication, and especially the co-chairs of the SAPEA working group, Professors Anna Fabijańska and Andrea Emilio Rizzoli. We would also like to express our sincere gratitude to all experts involved in the process of evidence-gathering and peer review, and everyone else involved in pulling this report together.

Finally, we would also like to express our sincere gratitude to the academies across Europe, for their contribution in bringing together the outstanding experts who formed the working group.

Tuula Teeri, Chair of the Euro-CASE Board
Patrick Maestro, Secretary General of Euro-CASE
Stefan Constantinescu, President of the SAPEA Board

Preface

The rapid advancement of artificial intelligence is driving transformative impact across numerous scientific fields. It has also opened new frontiers for research across various disciplines. From facilitating extensive experimental data analysis to generating novel scientific hypotheses from literature, AI has the potential to revolutionise scientific discovery, accelerate research progress and boost innovation.

As artificial intelligence continues its remarkable evolution, a deeper understanding of its potential impact on science is crucial for researchers and policymakers to ensure its responsible adoption and use.

This evidence review report contributes to ongoing debate on how artificial intelligence can be harnessed for scientific advancement while addressing potential challenges and risks associated with its adoption. It examines the issue of responsible and timely AI uptake in science in Europe. Specifically, it analyses the current landscape, examines existing challenges and opportunities associated with AI adoption in science, analyses the impact of AI on researchers' work environments and skills, and proposes policy options to address challenges identified.

This report draws upon a comprehensive evidence base established through an extensive literature review and further enriched by three expert workshops held between late 2023 and early 2024.

We extend our gratitude to all working group members, SAPEA's report writing team and contributors for their dedication and hard work in completing this report within a concise timeframe. Additionally, we thank the experts who participated in the workshops, providing valuable insights and expertise that greatly enriched the analysis presented in this report.

Anna Fabijańska, co-chair of the working group
Andrea Emilio Rizzoli, co-chair of the working group

Members of the working group

- Anna Fabijańska, Lodz University of Technology, Poland (co-chair)
- Andrea Emilio Rizzoli, Istituto Dalle Molle di studi sull'Intelligenza Artificiale (USI-SUPSI), Switzerland (co-chair from 19 September 2023)
- Paul Groth, University of Amsterdam, The Netherlands
- Patrícia Martinková, Czech Academy of Sciences, Czechia
- Arlindo Oliveira, Instituto Superior Técnico, Portugal
- Karen Yeung, Birmingham Law School, UK
- Virginia Dignum, Umeå University, Sweden (co-chair and working group member until 7 September 2023, involved in the selection committee)

The above experts were identified with the support of:

- The Academy of Engineering of Portugal
- The Academy of Sciences of Lisbon
- The Czech Academy of Sciences
- The Polish Academy of Sciences
- The Royal Netherlands Academy of Arts and Sciences
- The Royal Swedish Academy of Engineering Sciences
- The Swiss Academies of Arts and Sciences
- The Young Academy of the Polish Academy of Sciences

Executive summary

This SAPEA evidence review report gathers the relevant scientific evidence to analyse:

> How can the European Commission accelerate a responsible uptake of AI in science (including providing access to high-quality AI, respecting European Values) in order to boost the EU's innovation and prosperity, strengthen the EU's position in science, and ultimately contribute to solving Europe's societal challenges?

Specifically, the report approaches the topic through the lens of AI's impact on:

- scientific process, including the underlying principles upon which the scientific endeavour is organised
- people, including the skills, competencies, and infrastructure needed by scientists of tomorrow
- policy design, in the context of ensuring a timely, responsible, and innovative uptake of AI in science in Europe

In the rapidly-evolving field of AI, there is no universally accepted definition, nor a clear taxonomy outlining its various branches. Establishing such a definition would facilitate international collaboration among different countries. Therefore, recently, OECD countries have agreed to define an AI system as:

> a machine-based system 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. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

As AI applications permeate across many sectors, including in research, it is imperative that the EU takes hold of the opportunities, acts upon the challenges, and safeguards citizens from the risks that this fast-evolving technology can bring. As a companion effort to the EU's regulatory AI Act in progress, which is working to promote the uptake of human-centric and trustworthy AI while ensuring a high level of protection of health, safety and fundamental rights (per AI Act Recital 1) in Europe, the European Commission aims to understand the specifics of AI technology not only developed by science, but as applied to science; that is, AI in science. This report reviews current evidence and potential policies that could support the responsible and timely uptake of AI in science in the EU that may enhance the EU's innovation and prosperity. The scope of this report is confined to the takeup of AI in scientific research, rather than its implications for society more generally. In particular, it does not address the manifold and very significant challenges that arise from the growing and rapid deployment of AI technologies in specific social domains, and which fall outside the scope of this report and which the EU's AI Act is intended to address.

Landscape of AI in research and innovation

Computational power

The computational power ('compute') required for advanced machine learning systems has increased exponentially for many decades and in particular since 2010, ultimately leading to a divide between academia and industry regarding access to specialised software, hardware, and skilled workforce. Academic institutions released the most significant machine learning systems until 2014, but industry has now taken the lead. Generative AI models (especially large language models (LLMs) and diffusion models), convolutional neural networks for vision applications and models trained using deep reinforcement learning, widen the gap between industry and research due to the huge computational resources needed to train them. Governments are investing in computing capacity, but lag behind private sector efforts. Newcomers, startups and AI research laboratories frequently build on big tech's cloud services to train models and launch products.

Data

Besides compute, data is a crucial resource for AI development. However, the need to comply with copyright laws and research ethics requirements create challenges for public institutions in obtaining and processing data. While efforts are being made to provide fair and equal access while preserving privacy and ownership rights, issues remain unresolved.

Geopolitics of AI

The USA, China and the UK (ranking third, but lagging far behind the first two) dominate AI research globally, but other European countries contribute significantly. China leads in scientific domains where AI plays a prominent role, while the USA excels in health-related fields and the EU in social sciences and humanities. In 2022, the USA had more authors contributing to significant machine learning systems than other countries like China and the UK. The top-cited AI papers are from companies like Google, Meta, and Microsoft. The growth rate of AI-related publications is much faster than overall scientific production globally.

AI investment is surging globally, primarily driven by private sector investment, with the USA taking the lead. The EU has lost innovation leadership due to low research and development (R&D) investment and fewer startups. The problem is the commercialisation of R&D and scaling up. European efforts to boost AI development include programmes such as Horizon 2020, Horizon Europe, the Large AI Grand Challenge within the AI innovation package, and access to supercomputing resources through the EuroHPC JU network.

Regulatory landscape

Numerous AI-specific legal and regulatory measures are emerging worldwide, with many nations yet to enact comprehensive AI legislation. Most countries rely on existing frameworks for regulation, complemented by governance guidelines. As of October 2023, 31 countries had enacted AI laws, while an additional 13 countries are discussing potential regulations. The EU and China are leading in developing comprehensive AI regulations. China has been at the forefront of AI regulation, enacting specific measures for algorithmic bias, the responsible use of generative AI, and more robust oversight of deep synthesis technology. In the EU, considerable attention has been devoted to its AI Act, described by the European Commission as the most comprehensive AI legislation in the world. More recently, USA AI policy has seen the publication of several legally-mandated reforms.

Opportunities and benefits of AI in science

The increasing accessibility of generative AI and other machine learning tools for the analysis of large volumes of data has led scientists across various disciplines to incorporate them into their research. These tools facilitate the analysis of large amounts of text, code, images, and field-specific data, enabling scientists to generate new ideas, knowledge, and solutions. The number of scientific projects incorporating AI proliferates, with successful examples in protein engineering, medical diagnostics, and weather forecasting. Beyond facilitating groundbreaking discoveries, AI is also transforming the daily academic work of scientists, from supporting manuscript writing to code generation.

Accelerating discovery and innovation

AI's transformative potential extends to accelerating scientific discovery and innovation. The vast amount of research knowledge in natural language format is harnessed through literature-based discovery processes, using existing literature in scientific papers, books, articles, and databases to produce new knowledge. Researchers can now use LLMs to mine scientific publication archives to generate new hypotheses, develop research disciplines, and contextualise literature-based discovery. They can also use advanced search methods, such as those based on deep reinforcement learning, to comb vast search spaces, opening the way to AI-driven discoveries.

Scientific domains relying on large amounts of data seem to have taken up AI to a larger extent in their research processes. The generation of Big Data in these research fields presents a challenge that AI is well-equipped to address. AI algorithms analyse massive, complex, and high-dimensional datasets, enabling researchers to identify patterns and develop new insights. In fields like astronomy, particle physics, and quantum physics, where even a single experiment generates vast amounts of data, AI algorithms identify patterns at scale with increased speed, allowing scientists to discover never-before-seen patterns and irregularities. AI is becoming an indispensable tool for extracting knowledge from experimental data.

AI and machine learning tools can also help bridge the gap between diverse research fields, promoting cross-disciplinary collaborations. By incorporating Big Data analytics using AI, humanities researchers incorporate quantitative measures, diversifying their research and research questions. For example, some historians use machine learning tools to examine historical documents by analysing early prints, handwritten documents, ancient languages, and dialects. Furthermore, AI exhibits potential in advanced experimental control of large-scale complex experiments. For example, physicists are now incorporating AI systems that use reinforcement learning to gain better control over their experiments.

Automating workflows

Traditionally, researchers performed experiments manually, often involving labour-intensive tasks. However, technological advancements enable the automation of a significant portion of experimental workflows. AI is revolutionising experimental simulation and automation, opening up new possibilities for research.

Enhancing output dissemination

AI is also enhancing the dissemination of research outputs. AI-powered language editing empowers non-native English speakers to refine their scientific manuscripts, bridging communication gaps between experts and the public. AI can also simplify the publishing process for newcomers, potentially fostering more inclusive scientific discourse.

Challenges and risks of AI in science

In taking up AI, scientific researchers need to address bias, respect principles of research ethics and integrity, and deal responsibly with issues surrounding reproducibility, transparency, and interpretability.

Limited reproducibility, interpretability, and transparency

The use of AI in science compounds existing concerns about reproducibility, while the opacity of AI algorithms also poses significant challenges to scientific integrity, interpretability and public trust. AI algorithms can generate useful outputs, but their opaque nature makes it difficult to verify the accuracy and validity of research findings. The lack of transparency in AI algorithms hinders reproducibility, as researchers cannot replicate important discoveries without knowledge or understanding the underlying methodological processes. The opacity of AI algorithms raises concerns about accountability and trust, particularly in high-stakes applications such as healthcare.

The increasing prevalence of generative AI models and computer vision systems produced by industry raises concerns about their opacity and the lack of control over human evaluation by academic researchers. Insufficient access to scalable pipelines, large-scale human feedback, or data hinders academic researchers' ability to assess the safety, ethics, and social biases of machine learning models. The challenge of building

state-of-the-art AI models due to the scarcity of computational and engineering resources leads to a reliance on commercial models, limiting reproducibility and advancement outside of commercial environments. The monopolisation of AI capabilities by tech giants raises concerns about their control over the development and application of AI, potentially limiting scientific progress and ethical considerations.

Poor performance (inaccuracy)

Despite their remarkable capabilities, AI models are susceptible to performance issues arising from various factors. One such factor is the quality of training data. The model's predictions will inevitably suffer if the data used to train an AI model is biased, inaccurate, or incomplete. Additionally, AI models require ongoing updates to maintain their accuracy. Failure to retrain models with current data can lead to outdated algorithms generating inaccurate outcomes.

Another crucial aspect is the representativeness of training data. AI models often learn from data that does not accurately reflect real-world populations. This discrepancy can introduce biases into the model, resulting in erroneous predictions. Finally, the lack of adequate knowledge and training among researchers and developers contributes to performance shortcomings. AI models may be developed and deployed irresponsibly without proper expertise, leading to ethical and legal complications and concerns.

Fundamental rights protection and ethical concerns also arise. AI has the potential to perpetuate existing social biases and discrimination because AI systems trained on historically biased data and thus likely to reproduce these biases in their outputs. This can have a negative impact on people from marginalised groups, who may be unfairly discriminated against. AI systems can also introduce new forms of bias, such as visual perception bias. Machine vision systems may be biased because they are trained on datasets not representative of the real world.

In AI research, industry is now racing ahead of academia. Industry research has greater access to resources, such as data, talent, computing power, infrastructure, and funding, enabling them to take the lead over academia in developing sophisticated AI systems. This can disadvantage smaller institutions and academic researchers, making it more difficult for them to advance research.

AI systems can also raise privacy and data protection concerns since they often collect and process personal data and other, confidential information. There are several other challenges to advancing AI in science e.g. its adverse environmental impacts.

Misuse (malicious actors) and unintended harm

The misuse of AI in scholarly communication can lead to several significant social harms, including the proliferation of misinformation, the creation of low-quality outputs, and plagiarism. It constitutes research misconduct.

AI-generated content can be challenging to distinguish from human-generated content, increasing the risk of spreading misinformation. Predatory journals and paper mills can use AI to create fraudulent research papers. AI can make it easier to plagiarise content, potentially violating copyright and other intellectual property rights. The ease of producing AI-generated content may lead to an increase in the number of irrelevant papers. This can erode trust in scientific findings. AI-based tools can falsify information, which could lead to research misconduct. Using AI to generate content may lower the bar on the required scientific quality of the original work.

Using AI-based tools can automate specific tasks in the peer review process but, unlike human reviewers, cannot properly assess the novelty and validity of research findings reviewers can. As of today, AI still performs poorly in attempting to assess research quality, lacking human reviewers' deep knowledge, capability of grasping meaning, significance and human understanding. Using AI-based systems to evaluate scientific research may introduce bias and additional errors into the research assessment process.

Societal concerns

The advancement of AI has raised concerns about its potential impact on society. One concern is the unfair appropriation of scientific knowledge, as large tech companies increasingly leverage scientific talent from universities and volunteer developers' contributions from public code hosting and community platforms. Simultaneously, these companies hold patents and profits for themselves, while controlling access to computing and datasets. Additionally, using copyrighted material as training data for AI models raises concerns about copyright infringement, yet those whose IP rights may have been interfered with lack the capacity or resources to challenge purported infringement and seek redress, while identifying how IP law should apply in these contexts remains unsettled and uncertain.

Another concern is AI's potential to manipulate and spread misinformation at scale. Additionally, it may pose cybersecurity threats, including malware generation through unsafe code with bugs and vulnerabilities, advanced phishing attacks using LLMs for large-scale deployment, cybercriminals leveraging AI tools for malicious activities or deepfakes and voice cloning leading to impersonation, fraudulent digital content generation and realistic voice scams. Furthermore, AI may impact modern warfare and facilitate bioweapons development.

Impact on scientists and researchers

Research environments, literacy and training

AI can change the research context and environment, automating tasks, enhancing productivity, and liberating researchers from menial tasks. It can also amplify a researcher's expertise by personalising research tools and tailoring support and assistance to individual needs, preferences, and expertise. This transformation demands adaptation and the acquisition of new skills. To benefit fully from AI, universities

and researchers must invest in AI literacy and digital skills, foster a collaborative culture between humans and AI in the framework of human-centred AI, and embrace the dynamic interplay between human expertise and AI augmentation.

AI literacy involves understanding the concepts, abilities, and limitations of AI technology and being able to effectively communicate with it while evaluating its trustworthiness. Ethical awareness, critical thinking, value addition to AI output, and fact-checking are other crucial skills for successful AI integration in research. Adapting to the rapidly changing research environment is crucial for remaining competitive in the field. Several AI teaching programmes exist in Europe to address these needs. These aim to educate individuals in various aspects of AI, from technical knowledge to ethical considerations to help develop a skilled workforce capable of addressing the growing demand for AI expertise across industries and sectors, including research.

Inequalities and biases

AI's potential to transform research demands a conscious effort to address the geographical disparities in AI access and development and gender imbalances. Researchers should embrace a human-centred approach, mitigate biases, and collaborate with stakeholders to ensure AI's positive and equitable impact on society.

Impact on researchers

Adopting AI in research careers may lead to negative consequences, undermining mental well-being, increasing job insecurity, pressure, and unfair discrimination. Additionally, using AI for review and selection processes can erode a sense of belonging among researchers. It is important to address these challenges to ensure the appropriate implementation of AI in research.

Evidence-based policy options

Based on these findings, this report identifies five broad challenges that confront EU policymakers that may help to accelerate the responsible and timely uptake of AI in scientific and research communities, thereby supporting European innovation and prosperity. In this context, 'responsible' is taken to mean that accelerated uptake of AI should strive to be in accordance with the foundational commitments of scientific research and the foundational values underpinning the EU as a democratic political community and thus ruled by law, ensuring respect for the fundamental rights of individuals and the principles of sustainable development.

The primary challenge that must be addressed in order to accelerate the uptake of scientific research both in AI, and using AI for research, concerns resource inequality between public and private sector research in AI. To foster scientific uptake of AI responsibly, four further challenges must be addressed, concerning:

- scientific validity and epistemic integrity
- opacity

- bias, respect for legal and fundamental rights and other ethical concerns
- threats to safety, security, sustainability, and democracy

This report then sets out a suite of policy options which are directed towards addressing one or more of these challenges. These policy proposals include:

- founding a publicly funded EU state-of-the art facility for academic research in AI, while making these facilities available to scientists seeking to use AI for scientific research, thereby helping to accelerate scientific research and innovation within academia
- fostering research and the development of best practices, benchmarks, and guidelines for the use of AI in scientific research aimed at ensuring epistemic integrity, validity and open publication in accordance with law and conducted in an ethically appropriate manner
- developing education, training, and skills development for researchers, supplemented by the creation of attractive career options for early career AI researchers to facilitate retention and recruitment of talented AI researchers within public research institutions
- developing publicly-funded, transparent guidelines and metrics, using them as the basis for independent evaluation and ranking of scientific journals by reference to their adherence to principles of scientific rigour and integrity. The publication of these evaluations and rankings would be intended to provide a more thorough, rigorous, informed, and transparent indication of the relative ranking of scientific journals in terms of their scientific rigour and integrity than existing market-based metrics devised by industry, helping to identify predatory and fraudulent journals
- establishing an EU 'AI for social protection' institute, which engages in information exchange and collaborates with other similar public institutes concerned with monitoring and addressing societal and systemic threats posed by AI in Europe and globally, proactively monitoring and providing periodic reports and making recommendations aimed at addressing threats to safety, security, sustainability, and democracy

Chapter 1. Introduction

We are experiencing the impacts of the disruptive technology of AI permeating across many sectors. As a 'general-purpose technology', it is imperative that the EU takes hold of the opportunities, acts upon the challenges, and safeguards people from risks that this fast-evolving technology can generate. As a companion effort to the EU's regulatory AI Act in progress, which is working to to promote the uptake of human-centric and trustworthy AI while ensuring a high level of protection of health, safety, and fundamental rights (per AI Act Recital 1), the European Commission seeks to understand the peculiarities of AI technology not only developed by science, but as applied to and within science. Specifically, the EU seeks to understand how best to design research and innovation policies that will strengthen its research ecosystem and its competitive profile in a global context. As part of this effort, in July 2023, Margrethe Vestager, Executive Vice-President of the European Commission and acting Commissioner for Innovation, Research, Culture, Education, and Youth, asked the Group of Chief Scientific Advisors to deliver advice on the topic of the successful and timely uptake of AI in science in the EU.

This SAPEA evidence review report gathers the relevant scientific evidence to inform the Advisors' Scientific Opinion. It addresses issues described in a scoping paper which sets out the formal request for advice from the College of European Commissioners to the Advisors. The aim of this report is to analyse:

> How can the European Commission accelerate a responsible uptake of AI in science (including providing access to high-quality AI, respecting European values) in order to boost the EU's innovation and prosperity, strengthen the EU's position in science and ultimately contribute to solving Europe's societal challenges?

Specifically, the report focuses on the key areas provided in the scoping paper to approach the topic through the lens of AI's impact on:

- the scientific process, including the underlying principles upon which the scientific endeavour is organised and governed
- the people, including the skills, competencies, and infrastructure needed by scientists of tomorrow
- the policy design, with the aim of ensuring timely and responsible uptake of AI in science in Europe

What is scientific research?

> Research is the quest for knowledge obtained through systematic study and thinking, observation and experimentation. While different disciplines may use different approaches, they each share the motivation to increase our understanding of ourselves and the world in which we live. […] Research involves collaboration, often transcending social, political, and cultural boundaries, underpinned by

> the freedom to define research questions and develop theories, gather empirical evidence, and employ appropriate methods (ALLEA, 2023)[1]

The governance of scientific research is largely undertaken by academics, predominantly through peer-based norms and mechanisms rooted in their widely-shared cultural, political, and professional commitments to science as the quest for knowledge and understanding. Four core norms of scientific research were first introduced by American sociologist Robert Merton, which he called the "ethos of science" rooted in its ultimate institutional goal of extending "certified knowledge" (Merton & Sztomka, 1996). These norms, which Merton described as "institutional imperatives", are:

- **Universalism** refers to the impersonal nature of science, in which scientific truth claims are evaluated in accordance with pre-established criteria concerned with evidence and methodology that is independent of the character, identity, or status of those making such claims and consonant with previously-confirmed knowledge. In other words, everyone's scientific claims should be scrutinised and evaluated equally to establish their "epistemic soundness" (de Melo-Martín & Intemann, 2023), irrespective of the identity or status of the scientist.
- **Communism** (sometimes referred to as 'communalism') refers to the status of scientific knowledge as common property, in which scientific discoveries are collectively owned as the 'common heritage' of humanity, underpinning the obligation for scientists to communicate their findings publicly and openly. Common ownership is supported by the institutional goal of advancing the boundaries of knowledge (and by the incentive of recognition which is contingent on publication).
- **Disinterestedness** requires that scientists work only for the benefit of science, and reflected in their ultimate accountability to their scientific peers, and not for any organisational or other interest.
- **Organised scepticism** requires that the acceptance of all scientific work should be conditional on assessments of its scientific contribution, objectivity, and rigour (Merton & Sztomka, 1996).

This ideal model of scientific endeavour and its organisation is grounded in a deep commitment to epistemic integrity (de Melo-Martín & Intemann, 2023).[2] To this end, the scientific community has developed a set of more specific norms or "principles of research" to "define the criteria of proper research behaviour, to maximise the quality and robustness of research, and to respond adequately to threats to, or violations of research integrity", and have been enumerated in codes of conduct for research integrity, such as *The European code of conduct for research integrity* (ALLEA, 2023) and *Best practices for ensuring scientific integrity and preventing research misconduct* (OECD, 2007). These codes concern various matters including research misconduct, involving the "fabrication (making up results and recording them as if they were real), falsification (manipulating research materials or processes or changing, omitting or suppressing data or results without justification) or plagiarism (using other people's work and ideas without giving proper credit

[1] Originally developed by ALLEA (2011) and the European Science Foundation, as a living document to be revised every 3–5 years.

[2] Some claim this model is too limited, arguing that scientists and scientific practice should be "socially responsible", which includes epistemic integrity, but also recognises that scientific research has social implications for which scientists, and scientific practice, and for which scientists should be responsible (de Melo Martin, 2023).

to the original source", thus violating the rights of the original author(s) to their intellectual outputs (ALLEA, 2023).

In this report, we proceed on the basis that the 'responsible' uptake of science in accordance with European values should be aligned with and serve the institutional goal of science in extending the bounds of certified knowledge that is openly and universally communicated, owned in common and epistemically sound.

Who conducts research?

Although the traditional image of research is that of a university-based academic research group, much contemporary research is carried out in industry and other settings. The organisation of research has changed over time and differs between Europe and the USA (Carlsson et al, 2009). In the 19th century, interdependence emerged between the needs of the growing US economy and the contemporary rise of university education (Rosenberg, 1985). In Europe, the role of the universities was more oriented towards independent and basic research, as manifested by the Humboldt University in 1809. Although basic science was weak in the US until the 1930s and 1940s, research universities emerged after World War II, largely designed as a modified version of the Humboldt system entailing competition and pluralism. The beginning of the 20th century saw the development of the corporate lab, which also conducted basic research (the first corporate lab was set up in Germany in the 1870s).

The close links between industry and science, characterised by collaborative research and two-way knowledge flows, were thus reinforced. At that time, in-house corporate research was much higher in the USA than in Europe, with the employment of scientists and engineers growing tenfold in the US between 1921 and 1940. During the 1940s, there was a huge increase in R&D spending driven by the war, while the following decades saw a decrease in R&D relative to gross domestic product. Basic research diminished, while firms also reduced their R&D spending. In the USA, the situation was reversed during the 1980s, propelled by a number of institutional reforms directed towards IP rights, pension capital, and taxes. Entrepreneurial opportunities were created through scientific and technical discoveries which were paralleled by governmental policies, and which inserted a new dynamism in the US economy. A shift then followed away from large incumbent firms to small, innovative, skilled-labour intensive, and entrepreneurial entities (Braunerhjelm, 2010; Carlsson et al, 2009).

A large range of private entities, varying from small organisations to large and powerful tech and social media giants, are now continuously engaged in research, including AI research. While the motives of commercial research may be profit-driven, it is generally considered important for innovation and economic growth (Quinn, 2021). Accordingly, this report proceeds on the basis that research encompasses the work of a variety of different research communities. In the remainder of this document, we will consider the impact of AI on science, encompassing both the science of AI and the use of AI for scientific research more generally.

What is AI?

In order to proceed, it is necessary first to define what AI is. Unfortunately, there is no commonly-agreed definition of AI, nor a clear taxonomy describing its various branches. AI is a fast-moving field and its recent growth has challenged many of the definitions that have tried to frame it. Yet governments need a common definition to regulate it effectively, making it easier for different countries to work together. With this in mind, recently (November 2023), OECD countries agreed on the following definition:

> An AI system is a machine-based system 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. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

The above definition has been used in the EU's AI Act. It emphasises the reactive nature of AI algorithms, producing an output upon presentation of an input, while recent developments hint at the possibility of AI algorithms becoming active systems, acquiring 'awareness'. AI developments could include acquiring what might be understood as 'emotional intelligence' and multiskilling/tasking.

Due to its inherent versatility and broad applicability, AI is considered a general-purpose technology. It has therefore moved from a purely technical field to an interdisciplinary research domain, with multifaceted implications in terms of its use and uptake. Thus, any policies surrounding AI uptake will inevitably impact various domains, including scientific research, although the extent of their impact is likely to differ across domains and contexts.

The trajectory of modern AI started in the early 2010s with the advent of the computational capabilities needed to run deep neural network architectures with millions of parameters, able to process large datasets and extract knowledge from diverse sources, including audio and video signals, displaying for the first time superhuman capabilities, for example in image classification tasks (Krizhevsky et al, 2017). Since then, the frontier has been pushed constantly further ahead, and with transformer architectures (Vaswani et al, 2017) introduced in 2017, natural language processing has made a giant leap. These huge 'large languagd model' (LLM) systems, composed of billions of parameters, can now emulate the ability of a human in producing written text, and some researchers are now debating whether we are on course for the achievement of artificial general intelligence (AGI). A framework for the evaluation of AGI has been proposed (Ringel Morris et al, 2023), that is a "form of AI that possesses the ability to understand, learn and apply knowledge across a wide range of tasks and domains. AGI can be applied to a much broader set of use cases and incorporates cognitive flexibility, adaptability, and general problem-solving skills".

While AGI might still be out of immediate reach and LLMs might not be the way to achieve AGI, at least in their present form, the capabilities of LLMs continue to impress, especially when processing multimodal data (text, images, videos, sounds) and when they are integrated with other types of AI, such as generative AI models like stable diffusion, reinforcement learning, and more. The recent release of Sora, the OpenAI application that generates high quality videos from textual prompts is such an example. As a result, recent

years have witnessed a significant rise in LLM research and development activities and growing academic, scientific and public interest in the field.

Given this trajectory of exponential increase in the capabilities of AI, substantial impacts are expected in practically every domain where data can be digitised and fed into an AI system. In this report, particular relevance is therefore attributed to the impact of generative AI, with a specific focus on LLMs, in the domain of scientific advances and innovation in AI research, since enhancing the productivity of knowledge discovery returns a manifold of applications in all other domains.

Report structure

- In Chapter 2, we discuss the preconditions and the context for the application of AI to research, namely the availability of skilled researchers and the necessary infrastructure to develop state-of-the-art AI algorithms, including the access to trustworthy data for training AI models. We also analyse the geopolitical and economic context, conditions, the availability of access to human resources and computational infrastructures, and then we provide a brief examination of the current regulatory landscape, which is rapidly evolving.
- Against this background, Chapter 3 reviews the evidence that demonstrates the potential benefits and novel opportunities for the future use of AI in the scientific discovery process, such as the automation of scientific workflows and the AI-enhanced exploration of scientific literature.
- Chapter 4 then identifies potential challenges and risks, including potential misuses and abuses, besides the fundamentally unsolved issue of the accuracy and explainability of some AI research (and AI-enabled research), including the most novel AI architectures.
- In Chapter 5, the focus shifts to the people behind the scientific discovery process and how researchers are affected by the mounting wave of AI applications in their respective research areas: in particular, how can we promote a collaboration between the human and the machine avoiding the pitfall of relinquishing the driving seat to the latter? For this purpose, the role of education as the key to foster a synergistic collaboration is analysed.
- Finally, Chapter 6 identifies a suite of policy options aimed at addressing the challenges identified in this evidence review that currently hinder the successful and responsible uptake of AI in science.

At the end of each chapter, we gather the key findings from the evidence reviewed. These key findings are organised to highlight the level of uncertainty associated with the evidence gathering to support them.

Chapter 2. Landscape of AI research & innovation

Data and computational infrastructure

The growth in AI is enabled by increases in computing power (the hardware, also called compute), open-source software platforms, and the abundant availability of Big Data. In particular, the advent of generative AI has radically heightened the demand for training data and for computational infrastructure.

The development of generative AI

The introduction of transformer architecture has enabled the parallelisation of computational processes during AI training, enabling training on much larger datasets than previously possible, and thus the scaling up of AI language models.

However, training and running LLMs requires significant amounts of processing power. Not only does this require large amounts of capital to invest in or rent the necessary hardware, such as powerful graphics processing units (GPUs) or expensive purpose-built chips, but it also requires the professionals who possess the skills and the experience to operate complex neural networks on large clusters of hardware (Luitse & Denkena, 2021).

The release of Generative Pretrained Transformer-3 (GPT-3) by OpenAI in June 2020 sparked a significant surge in interest in LLMs, driven by its remarkable human-like language generation capabilities. Instead of releasing the model as open source, like its predecessors, OpenAI introduced an API through which accepted users can access it as a running system to generate textual output, introducing a pricing plan two months later following the conclusion of its 'beta' phase during which users could test the service free of charge. This marked a distinct move away from open-source release, in which OpenAI operates GPT-3 as a closed system and controls its accessibility. In this way, Mayer has described this as creating a model of "unique dependence" (Mayer, 2021) as it no longer allows developers to view, assess or build on top of GPT-3. However, despite the real power of models such as GPT-3, it took roughly three more years for the power of LLMs to widely reach public attention, something that happened only in late 2022 with the release of ChatGPT. 2023 was a breakout year for generative AI, with leading tech companies releasing their LLMs. This also includes a large number of open source LLMs, over 1000 of which were available on the HuggingFace platform.[3]

[3] https://www.stateof.ai/ (p.100)

Apart from text generation capabilities, leading LLMs have gained multimodal capabilities across image, audio, video, tabular data, and text understanding (Chui, 2023). These capabilities include, among others (Gemini Team et al, 2023; OpenAI et al, 2023; You et al, 2023):

- understanding and reasoning across multiple data modalities (text, images, video, audio, and programming code)
- automatic speech translation
- video question answering
- reasoning about user intent
- solving visual puzzles
- source code generation for specific tasks
- reasoning in maths and physics

Parallel to generative language models, image generative models have been actively developed, with state-of-the-art models coming from tech companies and AI startups. This group of models performs text-to-image translation (e.g. DALL-E, Midjourney) or text-to-video translation (Girdhar et al, 2023; Ho et al, 2022), which is a process of synthesising a photo-realistic image or a short video corresponding to a textual description (so-called 'prompt'). The use of generative models expands to audio as well. Solutions exist to generate sounds and music based on textual prompts and/or input melody (Copet et al, 2023; Kreuk et al, 2022).

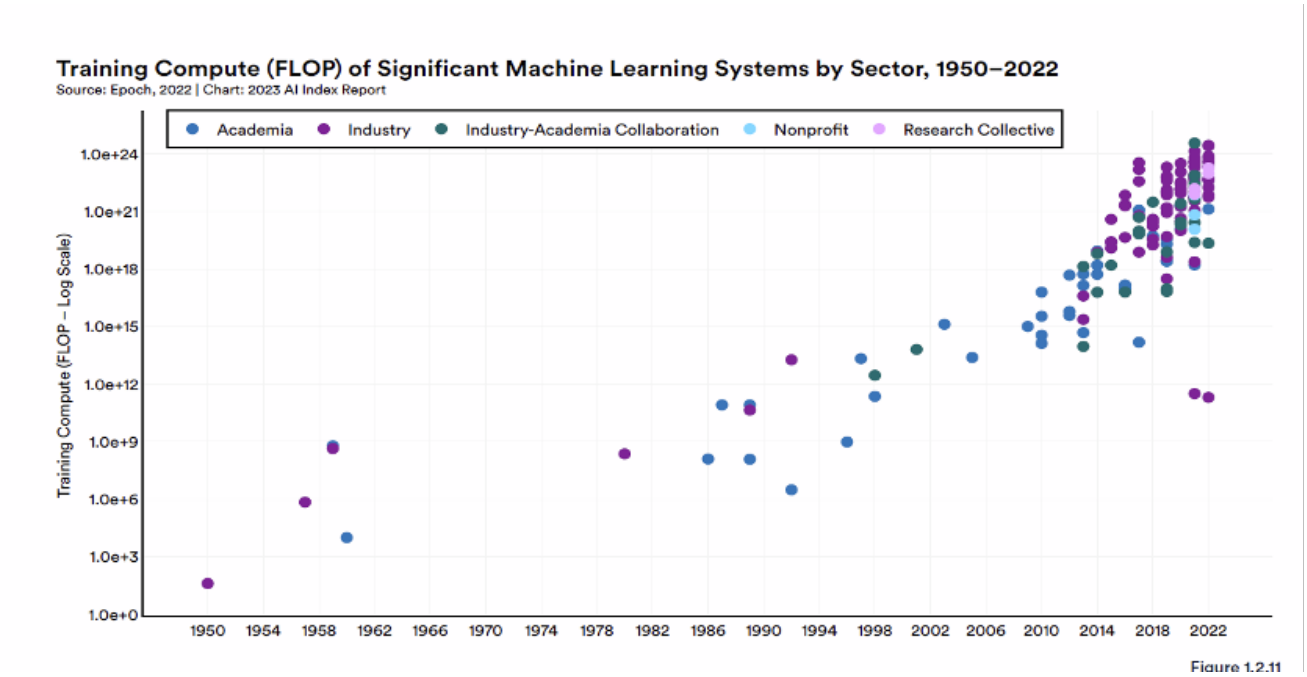
Hardware and software

The majority of AI developers, both in academia and industry, use open-source software frameworks to develop AI systems: PyTorch, Tensorflow, Keras, and Caffe can be used within Python and R to effectively develop and deploy advanced AI systems. But this is the only vertex of the software-data-hardware simplex where academia and industry stand on equal footing. The increasing compute needs of AI systems create more demand for specialised AI software, hardware, and related infrastructure, along with the skilled workforce necessary to use them. As government investments are constrained, compute divides between the public and private sectors can emerge or deepen. The massive expansion of the digital economy in the last two decades has become the object of social scientific research, including the "political economy of AI" (Srnicek, 2016; Zuboff, 2019). These studies investigate the dynamics of competition and the consolidation of power in the digital era. Although scholars have adopted a variety of theoretical frameworks, they all draw attention to the concept of the "digital platform", through which large tech firms position themselves as intermediaries in a network of different actors, allowing them to extract data, harness network effects and approach monopoly status (Luitse & Denkena, 2021).

The expansion of private sector monopoly power can worsen the disparity between public and private sector research, because the public sector increasingly lacks the resources to train cutting-edge AI models. Industry, rather than academia, is increasingly providing and using the compute capacity and specialised labour required for state-of-the-art machine learning research and training large AI models. This trend points

to the need to increase access to facilities like high-performance computing and software to support the development of AI in public science.

For example, the Stanford AI Index Report 2023 draws attention to the increasing computational power needed for complex machine learning systems. Since 2010, language models have demanded the greatest use of computational resources. More compute-intensive models also tend to have more significant environmental impacts; training AI systems can be incredibly resource-intensive, although recent research has shown that AI systems can be used to optimise energy consumption (Maslej et al, 2023). Industry players tend to have greater access to computational resources than others, such as universities, as demonstrated in Figure 1, which highlights the amount of 'compute' by sector since the 1950s.

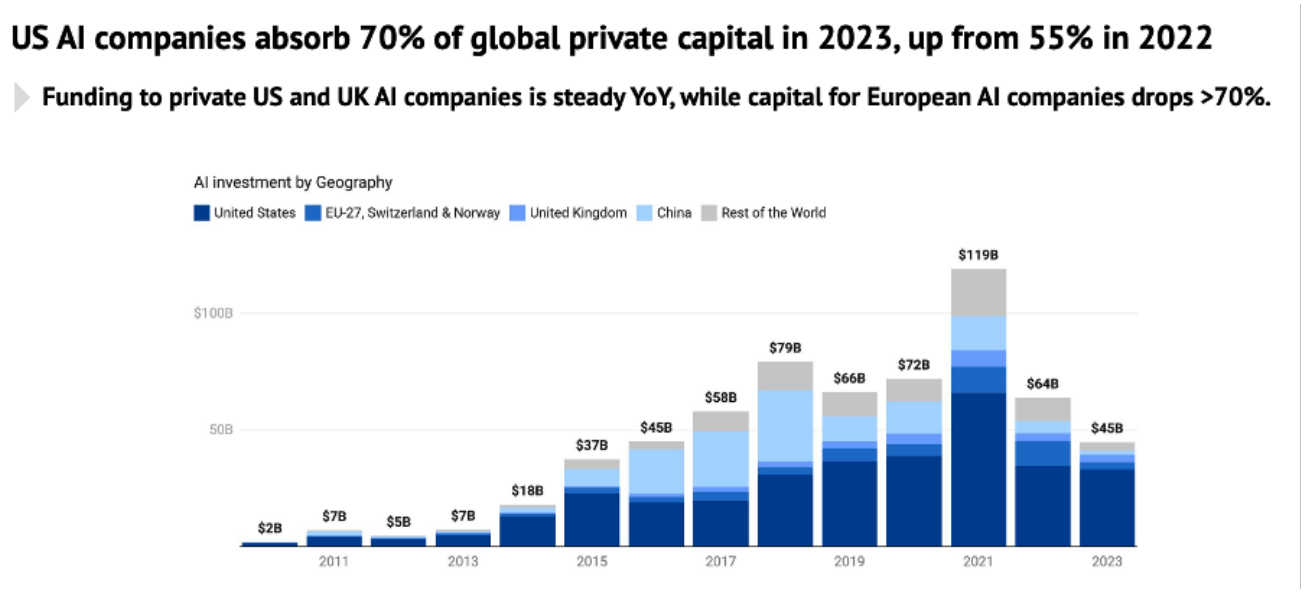


Training compute (FLOP) of significant machine learning systems by sector, 1950–2022, from the Stanford AI Index Report 2023 (Maslej et al, 2023). Source: Epoch, 2022 | Chart: 2023 AI Index Report

According to the Stateof.ai Report 2023, produced by Epoch AI:

> Governments are building out compute capacity but are lagging private sector efforts. Currently, the EU and the USA public research bodies are superficially well-placed, but Leonardo and Perlmutter, their national high-performance computing (HPC) clusters, are not solely dedicated to AI and resources are shared with other areas of research. Meanwhile, the UK currently has fewer than 1000 NVIDIA A100 GPUs in public clouds available to researchers.

The same report adds that companies such as Anthropic, Inflection, Cohere, and Imbue are shoring up NVIDIA GPUs and wielding them as a competitive edge to attract customers (Stateof.ai Report 2023, slides 131 and 72). This investment in compute is made possible by the enormous market capitalisation of AI startups, especially in the USA, as shown in Figure 2.



AI investment by geography, Stateof.ai Report 2023 on slide 111

Newcomers, startups, and even AI research laboratories rely on the computing infrastructure of Microsoft (Azure), Amazon (AWS), and Google (Google Cloud) cloud services to train their systems and use these companies' extensive consumer market reach to deploy and market their AI products (Kak et al, 2023; Rikap, 2023c, 2023d), thus reinforcing the model of unique dependence.

Data

Besides compute (hardware) and software, the most important resource for the successful development of AI systems is data.

Unlike software, much of which is open source, at least for what concerns the development frameworks for AI systems, a considerable body of valuable data is protected by copyright law. In addition, research undertaken at public research institutions is bound to comply with principles of research ethics, creating additional hurdles that must be met in order for that data to be accessed and processed by public research institutions. In the USA, for example, the New York Times has sued OpenAI for the alleged use of copyrighted material for training their GPT series LLMs, and we can expect further litigation of this kind. Although European law differs from US law, providing mechanisms by which copyright owners can reserve their rights for their copyright-protected work to be excluded from data mining by others, these procedures are criticised as unwieldy, impracticable, and difficult to enforce. Accordingly, the need to provide fair and equal access to data, while preserving privacy and ownership rights, remains an ongoing and fraught challenge. Although the EU is already active in this policy space, the issues remain contested and unresolved.

However, it is worth noting that even opening access to all currently available data might not be enough to address the needs of the most data-intensive AI algorithms, according to some claims. In particular, the

Stateof.ai Report 2023[4] claims that "we will have exhausted the stock of low-quality language data by 2030 to 2050, high-quality language data before 2026, and vision data by 2030 to 2060". Notable innovations that might challenge this claim are speech recognition systems such as OpenAI's Whisper that could make all audio data available for LLMs, as well as new optical character recognition models like Meta's Nougat.

Geopolitical economy of AI

To understand the context in which AI technologies are being taken up in research communities in Europe, it is helpful to understand the larger geopolitical and economic landscape and dynamics in which AI research and innovation are proceeding. For this, we briefly analyse the main actors in the development of AI systems and in the research about them, and then we review how such R&D efforts are funded. On this basis, we review the geopolitical implications.

Who develops AI systems?

AI research is on the rise across the board (affiliated with education, government, industry, non-profit, and other sectors), with the total number of AI publications more than doubling since 2010 (Maslej et al, 2023, pp. 24–28). Most significant machine learning systems were released by academia until 2014, but industry has since overtaken academics with 32 significant industry-produced machine learning systems compared to just three produced by academia in 2022. This could be attributed to the resources needed to produce state-of-the-art AI systems, which 'increasingly requires large amounts of data, computing power, and money: resources that industry actors possess in greater amounts compared to nonprofits and academia' (Maslej et al, 2023, p. 50). The Stanford AI Index Report 2023 estimates validate popular claims that large language and multimodal AI models are increasingly costing millions of dollars to train. For example, Chinchilla, an LLM launched by DeepMind in May 2022, is estimated to have cost $2.1 million; BLOOM's training is thought to have cost $2.3 million (Maslej et al, 2023, p. 62); and PaLM, one of the flagship LLMs launched in 2022 and around 360 times larger than GPT-2, is estimated to have cost 160 times more than GPT-2 at $8 million (Maslej et al, 2023, p. 23).

The highest number of notable machine learning systems originated from the US, totalling 16, followed by the UK with 8 and China with 3. Since 2002, the US has consistently surpassed the UK, EU, and China in terms of the overall quantity of significant machine learning systems produced (Maslej et al, 2023).

Who researches AI systems?

A bibliometric analysis of 815 papers published on AI and innovation in the areas of social science, business management, finance and accounting, decision science, and economics and econometrics as well as multidisciplinary areas revealed that 418 contributing authors were based in the USA, followed by China

[4] Slide 28.

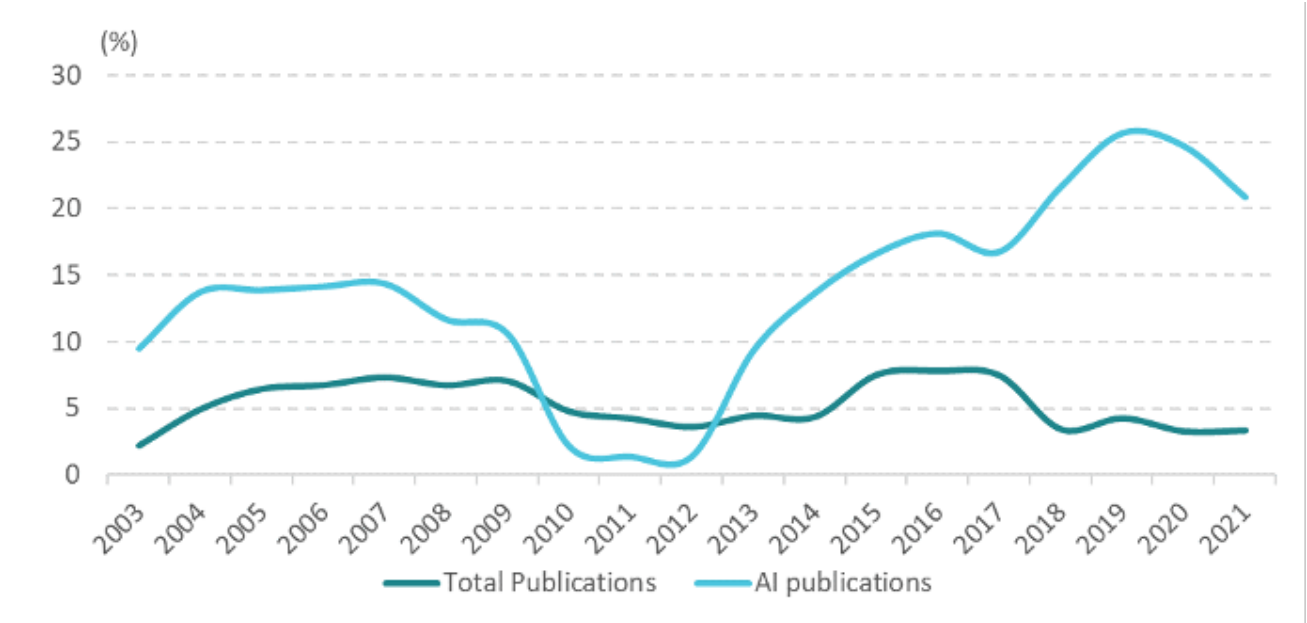
(229), the UK (125), and India (106) (Khurana, 2022). Cumulatively, authors from the EU and associated countries made a significant contribution, with 65 authors based in Italy and Spain each, 50 in Germany, 41 in Sweden, 33 in the Netherlands, 30 in France, 23 in Finland, 21 in Switzerland, and between one and 18 authors in other European countries.

Similar results were observed in another study on AI and innovation in business, management and accounting, decision science, economics, econometrics, and finance, where out of the 1448 identified records, 227 originated in the USA, 151 in China, 125 in Italy, 123 in Germany, and 119 in the UK (Mariani et al, 2023).

A bibliometric study of 5890 AI-related articles published in 2020 and 2021, during the COVID-19 pandemic, found that the top countries were China (2874 records), the USA (895 records), and the UK (430 records), closely followed by Australia (426 records). France and Germany were the countries of origin of 182 and 181 papers respectively (Soliman et al, 2023).

Similarly, authors based in the USA, the UK, and Canada produced the most research on AI in healthcare, followed by authors from Germany, Italy, France, and the Netherlands as well as China and India (Bitkina et al, 2023; Zahlan et al, 2023), while the USA, China, the UK, Canada, India, and Iran dominate research on AI in engineering, with a significant number of papers being produced in Germany and Spain (Su et al, 2022; Tapeh & Naser, 2023).

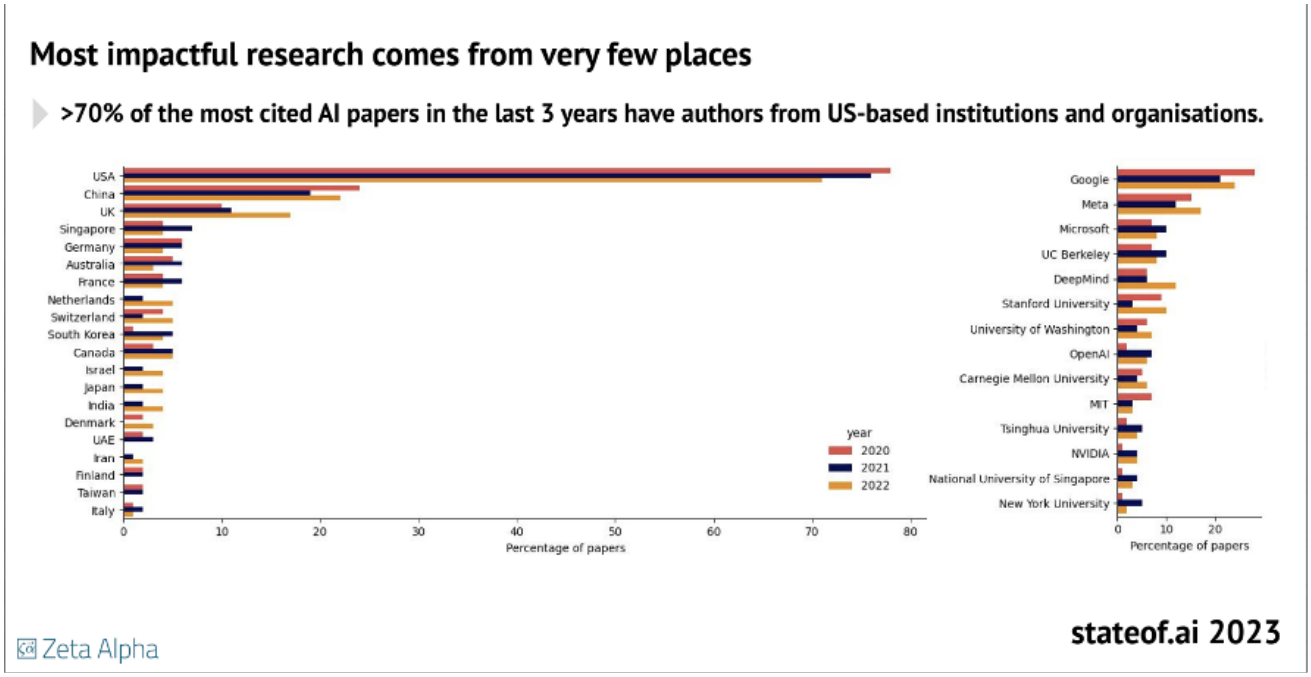
An important source of data is a bibliometric analysis undertaken by the European Commission's Directorate-General for Research and Innovation (European Commission, Arranz, et al, 2023). According to this analysis, the field of AI is growing at a faster rate than that of scientific production as a whole: global scientific activity has grown at around 5% per year between 2004 and 2021, while the annual growth rate of AI-related publications has been around or above 15%, except for 2010–2012 (Figure 3).



Growth in scientific activity, calculations based on Web of Science data. Annual growth calculated as a 3-year rolling average (from European Commission, Arranz, et al, 2023

In 2022, the USA led in the number of authors contributing to significant machine learning systems, boasting 285 authors. This figure is more than double the count in the UK and nearly six times that in China. In the last decade, the USA has outpaced both the EU and the UK, as well as China, in terms of private AI investment and the number of newly-funded AI companies. Since 2015, China has outpaced the EU in private AI investment (Maslej et al, 2023).

The Stateof.ai Report 2023 shows that over 70% of the most cited AI papers in the past 3 years have authors from USA-based institutions and organisations. The top 3 organisations with the most cited AI papers are also industry leaders, namely Google, Meta, and Microsoft, ahead of USA-based universities (Figure 4).



Most cited papers by country and by organisation, Stateof.ai Report 2023 on slide 68

Funding AI research and development

There is no systematic way of tracking AI funding across countries and agencies. However, a study by the European Commission Joint Research Centre's TechWatch estimated that 68% of AI investment in 2020 came from the private sector, with 32% from the public sector. Private sector investment was also growing at a faster rate (Dalla Benetta et al, 2021).

Europe is struggling to see the same level of private investments in AI as the US, and therefore is exploring a number of measures to kindle the development of AI companies and new AI products and innovations. European funding programmes such as Horizon 2020 and Horizon Europe support a large number of AI research projects. Since 2014, projects focusing on AI or using AI tools are estimated to have received €1.7

billion in EU funding. AI projects funded under the Horizon Europe programme are represented by academia and industry from multiple countries, aiming to boost collaboration in R&D across the EU.

The European Commission is also tackling the funding issue with a strategic approach outlined in the AI innovation package, which addresses the expected needs of startups and small and medium-sized enterprises. The first action of this programme has been to launch the Large AI Grand Challenge, a prize giving AI startups financial support and supercomputing access.

Finally, the EU is also funding access to supercomputing resources through the EuroHPC JU network, comprising eight supercomputers, two supercomputers still to be deployed, and six quantum computers.

AI as a geopolitical asset

AI has become a geopolitical asset. Building-blocks of AI technology act as geopolitical bottlenecks, impeding the development of the technology and its deployments in regions that do not have access. In particular, talents and computing power are strong limiting factors for the development of AI (Lazard, 2023).

Large AI models are attracting growing public investment. However, because of the growing power of these AI models, and because they lack transparency, it has become more and more difficult to balance out the power concentration (Miailhe, 2018). These limitations impact researchers on AI in the public sector who have limited or conditional access to the technology, and the uptake of AI in science, due to the unfair appropriation of scientific knowledge (see Chapter 4).

The Global AI Index ranks the performance of countries in AI, analysing absolute and relative measures including indicators on the implementation of AI systems, the level of innovation, and the amount of investments. The index delivers profiles for individual countries and an overall ranking arranged according to the final index scoring. The USA and China top the rankings, with EU countries following behind – but from the EU, only Germany, Finland, the Netherlands, France, Denmark, Sweden, Luxembourg, and Austria make the top 20. The UK is ranked fourth and Switzerland ninth.

A European Parliament Research Service report (EPRS, 2022) suggests that the EU has lost its innovation leadership due to low R&D investment and a small number of startups. The report attributes these issues to the commercialisation of R&D and scaling-up challenges. It offers an analysis of challenges that, if appropriately addressed, may help accelerate the responsible uptake of AI in science in accordance with European values, that might strengthen European science. To that end, it identifies policy options that are aimed at strengthening both scientific research in AI and the use of AI by European scientists more generally. These policy proposals are broadly concerned with:

- appropriate training, education, and development programmes
- the development of guiding principles and standards for the use of AI in science in accordance with the basic precepts of scientific rigour, integrity, and openness

- investment, infrastructure and institutional changes that may be necessary to establish the broader socio-technical, political and economic conditions for scientific research in and with AI to flourish in Europe.

As well as looking at state aid exemptions and corporate tax incentives for R&D expenditures, the report suggests, for example, promoting digital innovation hubs that act as 'one-stop shops' to provide services, from access to critical infrastructure and testing facilities to incubation and acceleration. These measures could promote the translation of research to commercial opportunities and the commercialisation of R&D by industry (EPRS, 2022).

The EU faces the challenge of researchers migrating to other regions, notably the US, leading to a brain drain. According to Khan (2021), five factors contribute to this outflow of human capital:

- attractive salaries outside Europe
- short-term fixed contracts for early career researchers
- unfair recruitment procedures
- appealing migration policies
- internationalisation policies that encourage permanent mobility

Though contributing to innovation across different fields, immigration's impact on STEM patent generation in the US appears particularly strong compared to arts and social sciences (Bernstein et al, 2022).

Regulatory landscape

AI-specific legal and regulatory measures are proliferating, reflected in the creation of laws and policy frameworks across many countries that are specifically directed at regulating AI. According to the Stanford AI Index Report 2023, which examined the legislative records from 127 countries, only one bill directly concerned with the regulation of AI passed into law in 2016, but by 2022, the number rose to 37 (Maslej et al, 2023).

The year 2016 marked a critical turning-point in public debate about AI, commonly referred to as the 'techlash' ("a strong and widespread negative reaction to the growing power and influence that large technology companies hold", a term first introduced into the Oxford English Dictionary in 2018). In particular, public revelations concerning the Russian use of social media platforms to interfere with the 2016 USA elections, Cambridge Analytica's misuse of Facebook data for political micro-targeting, and the opening of investigations against Google for alleged antitrust violations, highlighted how AI applications can damage vital individual rights and collective interests, threatening the integrity of democratic procedures. Until 2016, policymakers worldwide had largely accepted that self-regulation could be relied upon to address adverse impacts arising from AI applications, reflected in the multitude of 'ethical codes' promulgated by members of the tech industry, either individually or as various consortia (Rességuier & Rodrigues, 2020; Yeung et al, 2019).

Legislative proposals specifically concerned with the regulation of AI have subsequently emerged throughout the world, but the content and scope of these measures display considerable variation. Although the EU and China are taking the lead in developing comprehensive AI regulations, more recent US AI policy has seen the publication of a number of legally-mandated reforms. For example, on 30 October 2023, President Biden issued a sweeping executive order on AI with the goal of promoting the "safe, secure, and trustworthy development and use of AI", which applies to executive branch authorities only, and relies extensively on the US National Institute of Standards and Technology to develop guidelines and best practices. Several weeks later, the AI Research, Innovation, and Accountability Act of 2023 was introduced, supported by key members of the Senate Commerce Committee from both parties. It seeks to:

- introduce legislative initiatives to encourage innovation, including amendments to open data policies, research into standards for detection of emergent behaviour in AI, and research into methods of authenticating online content
- establish accountability frameworks, including key definitions, reporting obligations, risk-management assessment protocols, certification procedures, enforcement measures, and a push for wider consumer education on AI

China has also introduced AI laws, comprised of a series of more targeted AI regulations, enacting specific measures for algorithmic bias, the responsible use of generative AI, and more robust oversight of deep synthesis technology (synthetically generated content) arising by the Algorithmic Recommendation Management Provisions (2021), Interim Measures for the Management of Generative AI Services (2023), and the draft Deep Synthesis Management Provisions (2022). These measures are regarded as laying the intellectual and bureaucratic groundwork for a comprehensive national AI law that China is expected to release in the coming years. These regulations aim to prevent manipulation, protect users, and ensure AI's responsible development and use while enabling Chinese regulators to develop their bureaucratic know-how and regulatory capacity.

In the EU, considerable attention has been devoted to its AI Act, described by the European Commission as the most comprehensive AI legislation in the world. The text of the Act that will form the basis of the vote in the Permanent Representatives Committee on 2 February 2024 was leaked on 22 January 2024 and was officially released to Member State delegations on 24 January 2024. Its stated purpose is to establish a uniform legal framework for the development and deployment of AI systems in the EU in conformity with European values to "promote the uptake of human-centric and trustworthy AI while ensuring a high level of protection for health, safety, fundamental rights enshrined in the Charter including democracy, the rule of law and environmental protection against the harmful effects of AI systems in the Union and to support innovation" (AI Act, Recital 1, paragraph 11). It adopts a so-called "risk-based" approach, such that the higher the risks associated with the AI system in question, the proportionately more demanding legal requirements. To this end, it classifies "AI systems" into four classes:

- prohibited practices, considered to pose an unacceptable risk to the EU's values and principles, such as those that manipulate human behaviour or exploit vulnerabilities, and are therefore banned

- "high risk" AI systems that are used in critical sectors or contexts, such as health care, education, law enforcement, justice, or public administration
- "general-purpose AI models", defined as AI models, including those trained with a large amount of data using self-supervision at scale, that "display significant generality and are capable to competently perform a wide range of distinct tasks". These systems, and the models upon which they are based, must adhere to transparency requirements, including drawing up technical documentation, complying with EU copyright law and disseminating detailed summaries about the content used for training. If these models meet certain criteria, their developers will have to conduct model evaluations, assess and mitigate systemic risks, conduct adversarial testing, report to the Commission on serious incidents, ensure cybersecurity, and report on their energy efficiency
- "limited or minimal risk" AI systems, which are subject to transparency obligations, such as informing a person of their interaction with an AI system and flagging artificially generated or manipulated content

AI systems that do not fit within these categories fall outside the scope of the Act.

The Act also seeks to establish new institutional and administrative reforms, including:

- **An AI Office** within the Commission. It will oversee the most advanced AI models, help develop new standards and testing practices, and oversee the enforcement of common rules in all EU member states. Some commentators anticipate that its role will become equivalent to the AI Safety Institutes that have recently been announced in the UK and the US.
- **A scientific panel of independent experts** to advise the AI Office about general-purpose AI models, and to contribute to the development of methodologies for evaluating the capabilities of foundation models and monitor possible material safety risks related to foundation models, when high-impact models emerge.
- **An AI Board**, which comprises EU member states representatives, to remain as a coordination platform and an advisory body to the Commission while contributing to the implementation of the AI Act (e.g. designing codes of practice).
- **An advisory forum for stakeholders**, to provide technical expertise to the AI Board.

Despite these institutional innovations, the entire foundation of the regime is based on the EU's existing 'New Legislative Framework' approach to product safety. Although developers and deployers of these systems must ensure that their systems comply with the 'essential requirements' specified in the Act (for 'high risk' systems, these concern data quality, transparency, human oversight, accuracy, robustness, security, and the maintenance of suitable 'risk management' and 'quality management' systems), the Act provides that European standardisation bodies (notably CEN/CENELEC) may establish "harmonised standards" for AI. This standard-setting work is currently underway. If formally adopted by the European Commission (through notification in the Official Journal), firms that voluntarily comply with these standards will benefit from a presumption of conformity with the Act's essential requirements.

Yet these technical standards are not, and will not be, publicly available on an open-access basis: because CEN/CENELEC are non-governmental voluntary organisations through which technical standard-setting is undertaken by volunteer experts, the standards are protected by copyright. Nor is the technical standard-setting process subject to the conventional legal procedures of democratic consultation and oversight. Thus, although civil society bodies have broadly welcomed the EU's initiative, they have expressed significant criticisms of what they regard as deficiencies in the opportunities it provides for democratic participation while failing to provide meaningful and effective protection for the protection of fundamental rights, consumer safety, and the rule of law (Ada Lovelace Institute, 2023a; ANEC, 2021; BEUC, 2022; Micklitz, 2023).

However, it is important to situate the EU's AI Act within its broader digital strategy. The EU has been at the forefront of establishing a legal framework to protect individuals' personal data, with the General Data Protection Regulation as its centrepiece. Its 'digital strategy' is intended to supplement the EU's data and digital framework with a new set of rules aimed at fostering data flow, data access, and the data economy introduced by the Data Act and the Data Governance Act, which will apply to both personal and non-personal data, including machine and product data. It also introduces enhanced legal obligations and user protections for online platform services, online hosting services, search engines, online marketplaces, and social networking services under the Digital Markets Act and the Digital Services Act. Accordingly, the AI Act must be understood within this broader digital policy landscape, particularly given the role of data as a critical input for AI development and technologies. A discussion of these elements is beyond the scope of this report.

We have already noted active contestation and uncertainty about the scope, role, and limits of copyright law, given that many of the latest 'foundation models' used for generative AI applications rely on ingesting massive volumes of data scraped from the internet, including works that are ostensibly subject to copyright protection. However, IP law is a highly specialised field of legal protection. Accordingly, identifying and applying the content and contours of copyright law has become increasingly fraught as the size, significance, and sophistication of digital technologies and the digital economy has grown in recent years, particularly since IP laws are typically jurisdiction-specific. It is widely recognised that IP laws serve an important purpose that enables innovation to flourish, by conferring property rights on the creators of original works which are legally enforceable. Yet IP legal theorists also recognise that the goal of IP law should be to strike an appropriate balance between the interests of authors of original content in the temporary monopoly which IP rights create (thus recognising their interests and investments in their own creation, the personality of the authors, etc.) and the protection of certain other interests, such as public access to knowledge and information. In this way, copyright can foster creativity, innovation, and socioeconomic welfare. Thus, various European laws contain specific carve-outs to allow for scientific and other kinds of research. For example, the General Data Protection Regulation provides special provisions for the processing of personal data for "archiving purposes in the public interest, scientific or historical research purposes or statistical purposes" (GDPR article 89), while the EU's Digital Single Market Directive allows two text/data mining exceptions. Article 3 introduces a mandatory exception under EU copyright law which exempts acts of reproduction (for copyright subject matter) and extraction (for the sui generis database right) made by "research organisations and cultural heritage institutions" in order to carry out text and data mining for the purposes of scientific research, while Article 4 mirrors Article 3 with one major difference: it is

available to any type of beneficiaries for any type of use, but these can be overridden by express reservation via right holder 'opt-out' (see, for example, Margoni & Kretschmer, 2022).

Most countries do not have comprehensive AI legislation, laws, or policies tailored explicitly to AI. Instead, AI operates within existing legal and regulatory frameworks, complemented by governance frameworks, supporting acts, and guidelines (Australia, India, Israel, Japan, New Zealand, Saudi Arabia, Singapore, South Korea, United Arab Emirates, UK, USA). As of October 2023, 31 countries have enacted AI legislation, while an additional 13 countries are currently discussing and deliberating on AI laws.

Key findings

Little uncertainty

These key findings are supported by a large body of evidence and systematic analyses. There is little uncertainty.

- AI research is characterised by a strong leadership of AI research activities and infrastructure development by industry. This has implications for the practice of research itself.
- AI research and research using AI require large amounts of infrastructure. The largest AI infrastructures are located outside Europe.

Some uncertainty

There is some evidence to support these key findings, but some uncertainty exists.

- Across the globe, the regulatory landscape around AI is highly dynamic. In Europe, the EU AI Act aims to become the most comprehensive AI legislation in the world.
- AI research and the use of AI in research are highly impacted by the strong economic and geopolitical interests in AI.

Chapter 3. Opportunities and benefits of AI in science

This chapter outlines currently available evidence on the uses of AI to support or enhance research work, through applications of AI across the scientific process. The evidence was curated to include the most relevant and successful uses of AI in science at this time and exclude controversial evidence that still needs to be better understood and investigated. Finally, this chapter highlights that the advances in AI science and technology will lead to further potential uses of AI in research, and that there are currently no comprehensive evaluation studies about the impact of AI on the science system as a whole.

AI is increasingly used throughout science

The use of AI in science is not new. However, scientists everywhere are incorporating generative AI and machine learning tools in their research due to increased accessibility. With tools to analyse large quantities of text, code, images, and field-specific data, AI technologies facilitate the generation of new ideas, knowledge, and solutions. For example, in the last three decades, the percentage of AI-related publications has increased from less than 0.5% to 4% of all publications in health and life sciences, social sciences and humanities; from less than 1% to 10% in the physical sciences. Meanwhile, 30% of computer science publications were AI-related in 2022 (Hajkowicz et al, 2022).

The number of scientific projects incorporating AI is growing, with successful examples in protein engineering, medical diagnostics, humanities and weather forecasting. AlphaFold, an AI tool developed by DeepMind, predicts structures of thousands of proteins with more than 90% accuracy, tremendously accelerating scientific productivity: it previously took years of study for a PhD student to explore the three-dimensional structure formation of a single protein (Jumper et al, 2021). Recently, a new family of antibiotics was found with the help of AI technologies, a significant advance in drug design (Wong et al, 2023).

Weather forecasts are becoming increasingly more accurate and reliable with AI. Multiple algorithms forecast precise weather conditions in just a few minutes, as data-driven AI models surpass in speed and accuracy physics simulation models that run on supercomputers, used until now almost exclusively by weather agencies (Voosen, 2023).

In the humanities, AI helps historians to trace the history and heritage of smells over the past 400 years in text and images (van Erp et al, 2023). Recent advances in deep learning have helped historians to pick out patterns in large and complicated datasets of poorly handwritten documents or early print, early languages and dialects, as shown by the ITHACA project for the case of ancient Greek inscriptions (Assael et al, 2022).

In environmental, earth and agricultural sciences, a wide range of AI technologies, including machine learning, neural networks, big data, robotics and image processing, exist to meet the modern demands of urban and rural assessment, planning and smart practice including sustainability and climate change mitigation (de Oliveira & de Souza e Silva, 2023; Lazzeretti et al, 2023; OCDE, 2023; Păvăloaia & Necula, 2023; Qazi et al, 2022).

In biology, human limitations in areas such as data collection and integration have resulted in the field splitting into highly specialised subdisciplines, while AI technologies have the potential to help researchers integrate existing knowledge (Hassoun et al, 2022). AI can learn how to translate vast amounts of existing data into accessible formats (Zhang et al, 2022). This would provide researchers with more complete and up-to-date information on the problem they are investigating and increase the efficiency of research.

Beyond making discoveries, scientists are also incorporating AI tools in their academic duties supporting literature search, summarisation, manuscript writing, and code development. For example, text-based generative AI tools can help non-English-speaking researchers write papers with grammatical accuracy.

While AI transforms scientific research in numerous ways, its use is restricted in several aspects, such as evaluating scientific literature to maintain academic integrity. For example, the European Research Council warns that using AI in grant and peer-review assessments goes against good scientific and professional conduct. According to its survey, 90% of European scientists agree on AI's ability to accelerate the scientific process, whereas the consensus is weak for AI-based publications and reviews (European Commission, 2023).

AI can accelerate scientific discovery and innovation

Automated idea generation from the literature

An overwhelming amount of research knowledge is available in text format. Literature-based discovery processes use existing literature in the form of scientific papers, books, articles, and databases in an automated or semi-automated manner to produce new knowledge. Using LLMs, researchers began mining large databases of scientific publications (such as Scopus or Web of Science) to generate new hypotheses, develop new research disciplines, and contextualise literature-based discovery (Henry & McInnes, 2017; Q. Wang et al, 2023). Some specific examples of the use of AI in the production of new knowledge are:

- Semantic Scholar, developed by the Allen Institute for AI, is a free, AI-driven search and discovery tool indexing over 200 million academic papers available to the global research community. The algorithm extracts meaningful connections within papers to discover and understand research.
- Researchers in the materials science field used natural language processing to capture complex concepts, such as the nature of the periodic table and the structure-property relationships in

materials. Using such an approach, scientists are already recommending materials for functional applications (Tshitoyan et al, 2019).

- Physicists developed a similar semantic network for quantum physics called SemNet using 750 000 scientific papers and knowledge from books and Wikipedia. An artificial neural network model then predicted future research trends generating personalised, out-of-the-box ideas (Krenn & Zeilinger, 2020).

Speeding up simulations, facilitating Big Data analysis

Many research fields generate Big Data, where massive, complex, high-dimensional datasets are ready to be analysed. Researchers can use AI and machine learning algorithms to identify patterns in the data and develop new insights. Simulating physics and mathematical problems remains challenging for high-dimensional data. Moreover, solving problems with complex physics is often expensive, requiring different formulations with elaborate computer codes. Machine learning offers a promising alternative, where deep neural networks trained on physical laws offer better accuracy and faster training (Karniadakis et al, 2021). For example:

- Humanities researchers work on Heritage Image Databases and identify recurring patterns across a vast collection of heterogeneous images to better understand cultural evolution across Europe. With AI, art historians analyse and annotate images to identify common and subtle patterns among images to derive new insights (Gefen et al, 2020).
- In material sciences and engineering, any material's structure, property, and performance information comes from atomic to macrostructure level, making it challenging to establish connections between materials. Deep learning can be used to extract meaningful information from unstructured data. By developing the Materials Genome Initiative, scientists have established a large, open-source data repository that researchers use for finding linkages between materials through deep learning algorithms (Choudhary et al, 2022).
- Quantum mechanics is an important field for developing better sensors and secure communication channels, and for enhancing computational power. Performing experiments to understand the foundations of quantum mechanics requires generating specific quantum states or efficiently performing quantum tasks. For a long time, people designed these experiments. However, automated computer algorithms and AI can create numerous experiments and predict putative outcomes through simulations. Researchers developed new AI-based frameworks and discovered highly-entangled quantum states, quantum measurement schemes, and quantum communication protocols while optimising the properties of quantum experiments and states (Ruiz-Gonzalez et al, 2023).

New ways of performing research and opening up new fields of research inquiry

AI and machine learning tools facilitate cross-disciplinary collaborations among diverse research fields promoting new ways of doing research. This is true in almost all domains of modern science and is likely to increase as the tools become more effective.

For example, Digital Humanities is a field that brings humanities scholars into conversation with computer and data scientists (Ekpenyong, 2021). Humanities research largely relies on qualitative analysis. By incorporating Big Data analytics using AI, humanities researchers incorporate quantitative measures to diversify research areas and questions (Gefen et al, 2020).

Similarly, historians use machine learning tools to examine historical documents by analysing early prints, handwritten documents, ancient languages and dialects (Donovan, 2023). For example, Time Machine Europe is a project that aims to enliven Europe's rich past with digital technologies to create a comprehensive map of the European economic, social, cultural, and geographical evolution across time (Kaplan & di Lenardo, 2017). By bringing cultural heritage and machine learning together to simulate large data of the past, this digital humanities project addresses how Europeans lived in the past and what their cultural values were.

Advanced experimental control

Physics experiments are often complex and large in scale. Therefore, physicists need to use algorithms to precisely control their experiments. Reinforcement learning is one of the most effective machine learning paradigms for control and sequential decision-making, deriving a control strategy that operates in a dynamic environment and makes beneficial decisions. Scientists are now incorporating AI systems that use reinforcement learning to exert better control on their experiments.

Specific examples of the application of AI in advanced experimental systems include the tokamak plasma control for nuclear fusion (Degrave et al, 2022), the control and manipulation of quantum systems (Reuer et al, 2023), and the calibration of scaled-up experiments in quantum computers (Ares, 2021).

Discoveries from experimental data

In certain fields like astronomy and quantum physics, even a single experiment produces large amounts of data that is difficult to analyse manually. Finding patterns in such data is like finding a needle in a haystack. AI algorithms identify patterns in such data at scale with increased speed, allowing scientists to find never seen before patterns and irregularities. Specific examples include:

- Machine learning facilitates Earth-like exoplanet characterisation in large astronomy datasets. Researchers recently used a neural network model to detect exoplanets in noisy time series data with a greater accuracy than other previously described methods (Pearson et al, 2017).

- Scientists are training AI algorithms to explore subtle signals in the mega dataset from the Laser Interferometer Gravitational-Wave Observatory to discover gravitational waves (Cuoco et al, 2021).
- In large sequencing databases, such as the Human Genome Project, AI and machine learning algorithms allow scientists to discover genetic alterations (Alharbi & Rashid, 2022), design new modified organisms, and discover pathways for developing new therapies (Drew, 2023; Lewis, 2023).
- In solid-state chemistry, identifying novel functional materials enables technological developments from clean energy to information processing. A neural network model discovered more than 2.2 million stable structures of inorganic crystals from agglomerated datasets encompassing computational and experimental structures (Merchant et al, 2023).
- In high-energy physics, AI-based systems are being increasingly used to identify the most interesting patterns for further analysis by physicists and other specialised algorithms (Calafiura et al, 2022).

AI can help automate scientific workflows

Traditionally, researchers perform experiments manually and these are often labour-intensive. With technological advancements, many experimental workflows can be automated using AI-based control. For example:

- AIA-Lab is an autonomous laboratory for the solid-state synthesis of inorganic powders that uses computations, historical data from the literature, and machine learning to plan and interpret the outcomes of experiments performed using robotics in synthesising novel materials (Szymanski et al, 2023).
- Coscientist is another example of an AI system driven by GPT-4 that autonomously designs, plans, and performs complex experiments and accelerates research across several different tasks. Using this tool, researchers successfully used robotic liquid handlers in biology and drug discovery applications (Boiko et al, 2023).

AI can improve the dissemination of research outputs

A global survey on the role and future of AI in academic publishing was conducted in 2021 (Thomas, Bhosale, Shukla & Kapadia, 2023). 212 universities in 54 countries generated 365 individual responses. Half of the participants suggested AI would support plagiarism checks. About 40% of participants suggested that AI could support language enhancement and over 30% of respondents suggested it could offer opportunities for text analysis, text summarisation, and grammar checks. Finally, about 20% of participants also suggested AI systems could support content extraction and creation, translation, and copyright checks.

Only a few participants thought that AI systems would develop into bots that write manuscripts. Participants in the survey seemed inclined to think AI can help automate repetitive tasks rather than replace scientists in the activities of communication. However, they also raised concerns about AI systems, putting these applications in perspective, such as lack of understanding of AI, infrastructure access and integration, access, and dependence on AI experts (Thomas et al, 2023).

AI can also support the review process of scientific papers, as shown by the AI assistant of publishing house Frontiers, which makes some background checks on the suitability of a paper before sending it out to human reviewers.

Future perspectives for AI systems, new approaches and techniques

Tapping into the full potential of AI in research is contingent on the continuing progress in the field and equipping researchers with the knowledge and skills to use AI appropriately, recognising its strengths and limitations. In a rapidly growing field that has the potential to affect all aspects of human lives, there is an urgent need to determine AI's limitations and risks. Developing AI research and technology by understanding it in socio-technical contexts can enhance its validity as a tool.

There are several key areas in AI development for potentially enhancing its usefulness for research.

- Retrieval-augmented generation is proposed as a remedy for the shortcomings of current LLMs. The tendency to 'hallucinate', i.e. to generate seemingly random and irrelevant responses, is a major challenge. This generates misinformation without providing logical reasoning for the output. Retrieval-augmented generation suggests connecting any LLM to external databases and other symbolic engines, for additional data that the model can use to mitigate the generation of false information (Y. Gao et al, 2023; Lewis et al, 2020; Li et al, 2022).
- The tremendous size of the most popular LLM is as much a weakness as a strength. The sheer number of these models' parameters makes them essentially inscrutable, and means that teaching them new skills is very expensive. However, combining LLM with smaller, purpose-made models promises to changing this. As an example, Bansal et al (2024) report improvements ranging from 13% to 40% in the performance of specific tasks after an LLM has been augmented with a smaller model (Bansal et al, 2024).
- One approach attempts to treat a LLM as a building-block of a bigger structure with the aim of eliminating LLM's known imperfections and in the hopes of developing altogether new abilities, such as enhanced reasoning, agency, and meta-cognition. Apart from the LLM itself, other elements that may comprise such megastructures include multiple neural models, discrete knowledge and reasoning modules, and external knowledge sources (Ahn et al, 2022; Karpas et al, 2022; Oliveira et al, 2023).

- LLMs can display causal reasoning abilities, but while they perform better than existing algorithms on a pairwise causal discovery task and even counterfactual reasoning task, they also fail unexpectedly (Kıcıman et al, 2023). The HuggingFace AI collaboration platform has published a specific dataset (Jin et al, 2024) to assess the causal inference performance of LLMs. More broadly, work on causal representation learning looks to discover causal variables in low-level observational data (Schölkopf et al, 2021).
- Systems such as Auto-GPT, which gained a lot of traction among the general public shortly after the release of GPT4, may also significantly enlarge the domain of applicability of LLMs. In principle, it should be possible to create a system that uses several LLMs, and then prompt them to interact with each other and with the external world, toward a specific goal. If done correctly, carefully, and responsibly, such a scheme would result in semi-autonomous digital entities, able to perform complex tasks with relatively little direct human supervision. While the prospect is definitely enticing and has managed to fire the imagination of many a techno-enthusiast, the current state of the technology limits its broader applications (Firat & Kuleli, 2023).
- Neuro-symbolic AI is a promising area of research integrating the symbolic and the neural approach to AI. Symbolic AI engines, based on the explicit representation and execution of the rules of logic, have been dominant in AI until the advent of deep neural networks in the early 90s of the past century, but nowadays the integration of generative AI models trained on symbolic engines is producing remarkable results, such as AlphaGeometry, an AI algorithm able to demonstrate Euclidean geometry problems approaching the performance of an average International Mathematical Olympiad gold medallist (Trinh et al, 2024).
- While language processing using LLMs is the most strongly transformative technology, computer vision technologies using convolutional neural networks are also having a significant impact in scientific fields. The development of multimodal LLMs such as GPT-4 and Gemini, which integrate text, images, and other modalities of information, will likely lead to much more capable systems with capacities that far exceed today's publicly available LLMs.

Key findings

Little uncertainty

These key findings are supported by a large body of evidence and systematic analyses. There is little uncertainty.

- AI is increasingly used throughout areas of research and throughout the research process.
- However, the applications and uptake of AI in research are unevenly distributed across scientific domains. There are currently many examples that highlight the potential of AI to support the research process, in particular in scientific domains relying on large amounts of data.

High uncertainty

There is little evidence and no systematic analysis to support these key findings.

- We are missing comprehensive evaluation studies about the impact of AI on the science system as a whole.
- Potential opportunities for AI uptake in qualitative and theoretical development research, in the humanities and social sciences, may develop. No systematic evidence of those opportunities is currently available.

Chapter 4. Challenges and risks of AI in science

Progress in science often requires new technological innovation. With the increasing use of AI in the scientific discovery process, scientists have begun to discuss the core issues raised by AI technologies.

To improve AI technology itself and promote its lawful and ethical use, developers and users must abide by laws and clear guidelines to avoid issues related to bias, ethics, reproducibility, transparency, and interpretability (H. Wang et al, 2023). These issues transcend scientific disciplines and require serious attention from policymakers.

Limited reproducibility, interpretability and transparency

The 'crisis' of reproducibility

Although AI software is being embraced across scientific disciplines, this has led to the publication of scientific research papers that fail to meet conventional standards of scientific validity. Accordingly, various scientific commentators express concern that the takeup of AI is exacerbating an already-existing scientific "reproducibility crisis" (Ball, 2023; Heaven, 2020; Hutson, 2018b).

Broadly understood, 'reproducibility' refers to the ability of independent researchers to achieve the same (or similar) results as a previous study using the same (or similar) methods, thereby demonstrating the study's validity (CCA, 2022). Although there has been no systematic evaluation of error in scientific papers, a recent editorial in *Nature* quotes several scientists who claim that "error-strewn AI papers are everywhere", who describe the problem as "widespread" in many communities beginning to adopt machine learning methods due to "lack of rigour" in developing these models (Ball, 2023).

A number of studies have sought to investigate how problems of reproducibility arise. For example, a recent survey of scientific papers (Kapoor & Narayanan, 2023) examined reproducibility in scientific papers, in which a research finding was defined as reproducible if the code and data used to obtain the finding were available and the data correctly analysed. They found that more than 300 published manuscripts were affected by errors due to 'data leakage', referring to a spurious relationship between the independent variables and the target variables that arises as an artefact of the data collection, sampling, or pre-processing strategy and which usually leads to inflated estimates of model performance. Other reproducibility failures can arise when machine learning methods are used, including lack of quality in the dataset, or inappropriate use of metrics for evaluation, exacerbated by the lack of standard modelling and evaluation procedures, so that reproducibility problems can arise even when standard datasets are used (Bender et al, 2021; Kapoor &

Narayanan, 2023; Paullada et al, 2021). Without understanding the input data and iteration process in the AI model, researchers cannot reproduce important discoveries (Lazzeretti et al, 2023; Mukhamediev et al, 2022; Tapeh & Naser, 2023).

Further challenges arise from the use of AI in scientific research that affect the validity and epistemic integrity of the resulting research findings, due to the need for cross-disciplinary knowledge and skills to ensure that AI is employed in scientific research in a domain-appropriate, context-sensitive manner. As the Council of Canadian Academies has observed, the use of AI in science is pushing disciplinary boundaries, collaboration and coordination towards a "transdisciplinary future" (CCA, 2022). Leonelli's investigations and analysis of the practice of scientific research using Big Data in science demonstrates that one size does not fit all. Instead, she highlights the importance of discipline-specific, localised judgements involved in conceptualising how data is understood as evidence for the purposes of scientific inquiry, including the practices and workflows through which machine learning techniques are employed to transform data into scientific findings (Leonelli, 2020). At the same time, notions of reproducibility may vary according to discipline (Leonelli, 2018). So, for example, research in medicine, history and the social sciences adopts observational methods rather than laboratory controlled-experiments, and thus rely on sensitive human judgement rather than mechanical objectivity (Daston & Galison, 2007). Yet given the novelty of AI tools and the rate at which they are advancing, there are grounds for concern that the kind of cross-disciplinary skill, knowledge and sensitivity needed to employ AI in accordance with the demands of epistemically integrity are currently lacking.

Compared with their industry counterparts, academic researchers have less access to large-scale human feedback, reinforcement learning, and human plausibility to test for AI safety, ethics, and social bias at scale (Casper et al, 2023). For example, in the natural language processing field, researchers typically benchmark their results through human feedback. However, many academics validate their results using ChatGPT, making the benchmarking process obscure while relying on commercial services to advance their work (Saphra et al, 2023). Further, such reliance on AI models to execute other AI tools contributes to reproducibility issues that can affect all domains where AI is used (Lee et al, 2023; Rogers et al, 2023).

The problem of opacity

One main concern with many modern AI methods is opacity. Lack of transparency makes it challenging to interpret results generated by AI algorithms. While AI allows scientists to identify new patterns and extract new insights, it is challenging to verify the accuracy and validity of many new AI-derived concepts (Cranmer et al, 2020; Iten et al, 2020; Krenn et al, 2021; Liu & Tegmark, 2022). The lack of transparency in how AI algorithms operate also contributes to challenges for reproducibility. This is further compounded when putting AI models into practice. Interviews with over 50 AI practitioners indicate that many experience "data cascades", where data issues propagate through AI systems causing negative results in domains ranging from wildlife conservation to public safety and health (Sambasivan et al, 2021).

There is a stark contrast between the AI research coming from publicly funded academic institutions and for-profit, private tech giants such as Google, Meta, and Microsoft. For example, in the UK, a huge share (70%) of

leading publications on AI are generated solely by DeepMind (European Commission, Arranz, et al, 2023). This distributional inequality can compound the problem of opacity, due to commercially-created opacity that arises from the assertion of IP rights over research produced by commercial firms rather than making their findings and methods openly and publicly available.

Poor performance

Due to poor data quality

Building an AI model requires input data to train the system. Poor quality data can generate bad models. Some of the main data quality issues include (Duan et al, 2022; Gill et al, 2022; Hassoun et al, 2022; Kumar et al, 2023; OCDE, 2023; Păvăloaia & Necula, 2023):

- accuracy of data
- faulty labelling
- accessibility of all the data
- data interoperability in different tasks

Researchers demand quality checks for data and results. Further, bias in datasets perpetuate existing data biases such as gender-based prejudices, making models trained on those datasets harmful for society. For example, many AI models that select candidates or recommend jobs mirror existing pay disparities between men and women (Bied et al, 2023; Gallegos et al, 2023).

In addition to the input data and output quality check, personal data protection issues may arise depending on the domain of application, especially in health and medicine. Tools supporting medical sciences require enormous amounts of high-quality input data that must comply with current standards, such as HL7 Fast Healthcare Interoperability Resources. Further, AI requires vast volumes of training data, leading to concerns about how data is collected and handled (Mukhamediev et al, 2022; Păvăloaia & Necula, 2023).

Due to failure to update the model

As data is dynamic, many AI tools must be retrained periodically. Tools trained a few years ago might quickly become obsolete if not retrained. The rate of error of AI tools inevitably increases in time, so their value will also decrease over time if they are not periodically retrained (Valavi et al, 2022). As yet, researchers are unclear about how well LLMs perform on real-world data because they cannot provide the evaluation data at a fast enough rate (Villalobos et al, 2022).

Due to differences between training data and real world population

For AI to be broadly applicable without artificially-created, unintended bias, the training data must incorporate real world population data and reflect real world scenarios as much as possible[5]. The rapid spread of AI technologies is generating growing concerns about data quality and privacy. AI relies on existing data for training and the quality and representativeness of such data determines how well the technology is able to operate (Lazzeretti et al, 2023; Mukhamediev et al, 2022). For example, individual biases from clinicians may be transferred onto an AI diagnostic tool (Kumar et al, 2023).

Depending on the data quality, correlation-based models can be very weak, making the results untrustworthy (Bied et al, 2023). Moreover, the success of using AI in survey work in social sciences depends on algorithmic fidelity of the trained data. AI-assisted research will depend on AI being able to accurately mirror the perspectives of diverse demographic groups. For now, pretrained models are known to capture sociocultural biases present in society (Bircan & Salah, 2022).

Similarly, AI can further create more bias for a specific group of people. For example, every disability is unique and may pose challenges for algorithms. As a result, algorithms may discriminate against individuals with facial differences or asymmetry, different gestures, gesticulation, speech impairment, different communication styles or used assistive devices. The most affected group – people with disabilities, cognitive and sensory impairments, or autism spectrum disorders – can be [excluded and unfairly discriminated against](https://www.psu.edu/news/information-sciences-and-technology/story/trained-ai-models-exhibit-learned-disability-bias-ist/>\ and <https://www.psu.edu/news/information-sciences-and-technology/story/ai-language-models-show-bias-against-people-disabilities/) (Goggin & Soldatić, 2022; Packin, 2021; Welker, 2023a, 2023b; Whittaker et al, 2019).

Due to inadequate knowledge and training

Ethics and legal requirements (lawful and ethical data governance)

Researchers will need to establish skills in, and guidelines for, the ethical use of LLMs and other AI based systems in research, addressing concerns related to data privacy, algorithmic fairness, replicability and the potential misuse of LLM-generated findings (Grossmann et al, 2023) and computer vision applications (Fabbrizzi et al, 2022). While there is a consensus among researchers on the immediate need for ethical guidelines, their ongoing absence may already be creating legal and ethical problems.

Cross-disciplinary expertise

Typically, every academic group drives its research on a limited set of questions fairly independently. With burgeoning growth in LLM service providers, there is a major identity crisis in the natural language processing field (Duan et al, 2022). For example, LLM growth follows Moore's Law, where the field moved

[5] However, real world population data may also reflect bias that is historically entrenched in social structures.

from 1 billion models to 500 billion models with increasing performance in the last two years. While this allows researchers to solve a large number of natural language processing tasks, they need to reckon with what kind of research questions the AI academics should focus on (Ignat et al, 2023; Li et al, 2023; Saphra et al, 2023; Togelius & Yannakakis, 2023).

Access to digital infrastructure and the capacity of individual infrastructure varies significantly, resulting in needs-based variation across individual disciplines and fields of scientific inquiry, indicating that support may be best targeted based on specific needs. A particular problem area is technical infrastructure for the arts and humanities. Humanities researchers often lacked quantitative or digital skills, whereas researchers with technical skills often lack awareness of ethical risks of AI. Therefore, interdisciplinary cooperation that cuts across traditional disciplinary boundaries in both public and private sectors offers a way forward to develop better AI, but proper training in the sensitivities and challenges of good quality cross-disciplinary research is required (Procter et al, 2020).

Fundamental rights protection and ethical concerns

Inequality, unjustified bias & unfair discrimination

Social-cultural bias reflected in underlying datasets

Some of the main areas of socioeconomic outcomes where AI fairness may cause concerns include housing, hiring, educational opportunities, and the court system (Mehrabi et al, 2019; Morse et al, 2022). Implicit bias is another complication where the current pipelines of training sets and their creation are influencing research findings and citation practices. If AI is trained on data reflecting white male, Western perspectives, one of its clear dangers is strengthening and reinforcing systematic discrimination against women and the hegemony of Western science while undermining work from the academics in the Global South.

Despite AI's promise to optimise and expedite research processes, there are a number of issues with its implementation. One of the most important challenges to consider is the transfer of existing biases onto automation tools. If the data used to train AI algorithms are biased – for instance, in how they were collected or curated – then AI will continue to promote those biases (Chubb et al, 2022; Lund et al, 2023). This might include, for instance, racism, misogyny, and ableism, subtly perpetuating discriminatory attitudes through microaggressions, dehumanisation, and sociopolitical framing within language and decision-making (Bender et al, 2021)

New forms of 'machine bias'

Machine learning could introduce new kinds of bias or outright falsifications into the historical record (Donovan, 2023). For example, machine vision systems may be inherently biased because they not only rely on biased datasets but their way of representing the visual world gives rise to a new class of bias called perceptual bias (Offert & Bell, 2021). Generative AI can essentially make it easy to create such content,

inadvertently reinforcing existing biases and stifling promotion of diversity. Moreover, the lack of transparency and black box nature of AI systems, the decision-making process can introduce bias and obscurity, eroding trust in scientific findings (Flanagin et al, 2023)

Inequality between well funded and poorly funded research

There is an expanding gap between entrenched, visible, popular research endorsed by rich sponsors, and marginalised, invisible, unpopular research which however is crucial to addressing global challenges. Technologies continue to be used as proxy for the quality control of data, where they amplify already popular research lines and exacerbate lack of confidence by low-resourced researchers or with those with lower skills (Leonelli, 2023b).

Studying human subjects using AI is also potentially problematic, with potential for bias and discrimination. Participants need to be aware of such risks. The support of AI to research ethics reviewers is also challenging, since AI systems show limited abilities for ethical and moral positioning (Pournaras, 2023).

Data privacy

Privacy and ethical concerns are at the forefront of risks of AI technology. For example, in healthcare, since data is privacy-sensitive, few public datasets are available and they are often used in research, resulting in the overfitting of models to specific datasets, which can hinder their generalisability (McDermott et al, 2021). On the other hand, many AI systems use copyrighted data without proper consent or respecting IP rights. Datasets flagged due to personal (e.g., biometric) information infringement may resurface through backchannels or derivative versions and be used to train AI months later (Paullada et al, 2021).

Challenges in advancing AI in science

Most AI researchers have limited access to the available computing power. With a few paid services monopolising LLMs, academics will depend on them while losing infrastructure control at scale (Lee et al, 2023). There are several reasons for academics falling behind industry in making advances in AI science. First, academia does not have scalable pipelines as the industry does to process large datasets required to build, train, and implement state-of-the-art AI models. The problem is exacerbated by private companies selling computational and engineering resources as a paid service (Ahmed et al, 2023; Lee et al, 2023).

AI model development, training, testing, and deployment have enormous computational and energy consumption costs. GPUs perform heavy computations at high speed; however, to build AI at scale, the number of required GPUs is very large. Therefore, all the stakeholders have a responsibility to sensibly evaluate the carbon footprint and other environmental impacts associated with the use of AI (Tamburrini, 2022). In addition to assessing carbon footprints, the environmental and societal costs of AI are so far unclear (Ligozat et al, 2022). Researchers ought to foster better practices of raw material capturing, efficient and inclusive AI systems (Schwartz et al, 2020), and sustainable practices (Jagannadharao et al, 2023; Kaack et al, 2022; OCDE, 2022).

Other ethical concerns

Experts advocate that more attention is needed to boost the uptake of AI technologies in a manner that respects human rights and values and earns public trust (European Commission, Arranz, et al, 2023). Improving understanding of this technology at every stakeholder level is particularly critical to achieving this. For example, a recent trend in surveys and crowdsourcing evaluations indicates that they are unusable because crowdsourcing workers often use ChatGPT to create summaries rather than writing them themselves, and they therefore fail to qualify as human evaluation (Veselovsky et al, 2023).

Another critical overview highlights some of the ethical dilemmas posed by generative AI and language models for knowledge, epistemology and research practice. It identifies risks of copyright infringement, deskilling of researchers in writing, research conduct, security, misinformation, and data quality (Pournaras, 2023). Within research design, developing a research hypothesis or research question, generative AI can be used, either as a research instrument or as a research subject, along with human subjects. Inappropriate reliance on these kinds of tools by researchers may result in a loss of critical thinking skills and confirmation bias, along with diminished accountability and transparency. AI can also diminish skills and competencies if the researchers overly rely on them for all aspects of their academic work, including data analysis, literary review writing, research assessment, etc. and other research skills. These trends give rise to fears of large-scale loss of human competences in specific fields as AI systems take over and fully replace humans in tasks. Researchers in medicine, law, chemistry, and many other fields are raising alarms of these possibilities if AI systems are extensively used to replace humans in these fields (Chiang, 2000).

Misuse and unintended harms – Misinformation and poor quality information

Predatory journals and fraudulent papers

Predatory publishing is already a challenge in scholarly communication because predatory journals or 'paper mills' create fraudulent content. Generative AI can essentially make it easy to create such content, inadvertently reinforcing existing biases and stifling promotion of diversity and making the fight even more difficult against paper mills that churn out fake research (Liverpool, 2023).

Proliferation of low-quality outputs

An increase in the number of irrelevant, low-quality papers is difficult to control as articles are becoming easier to produce. This puts a strain on the peer review process, where researchers simply cannot validate all the studies that are published (Park et al, 2023). Furthermore, the lack of transparency and black box nature of AI systems affects decision-making in scholarly communication and peer-review process by introducing bias and obscurity, eroding trust in scientific findings (Flanagin et al, 2023). For example, an OECD series on AI points out that humans are becoming less capable of differentiating AI from human-generated content,

thereby increasing risks of mis- and dis-information (Lorenz et al, 2023). There is a proliferation of scientific misinformation, as true, untrue, and fabricated data becomes more difficult to distinguish (C. A. Gao et al, 2023).

While AI is somewhat useful in enhancing the English readability of scientific papers, it is unreliable in assessing rigour, novelty, and impact of research papers. Evaluating these key attributes still requires expert human review. A recent study tested the potential of AI in peer review by evaluating a large number of conference papers using an AI model (Checco et al, 2021). Researchers showed that it has the potential to accelerate the peer review process by automating certain tasks such as plagiarism checks, manuscript formatting, quality control for fraudulent and erroneous data, testing the validity of statistical tests, and many more. However, AI cannot assess the novelty and validity of research findings better than researchers who are experts in their fields. Moreover, AI can introduce machine bias by focusing on authors instead of the content.

There are efforts to train AI in certain peer-review tasks to augment the process without relying on human quality checks. For example, a partial or complete automation of some publishing-related tasks, such as suggesting appropriate journals for an article, providing quality control for submitted papers, finding reviewers for submitted papers or grant proposals, reviewing, and review evaluation can be done by AI. In one case study (Kousha & Thelwall, 2023), researchers used provisional peer review scores for thousands of articles in different research areas submitted to the UK Research Excellence Framework and trained AI models to evaluate research quality using the available peer review scores. This is the only large-scale study using AI to predict research quality scores for journal articles. Each research output was scored on a four-point scale given by field-specific experts. The results were then used to assign £16 billion of research funding. Scientists who scored less than 3 points were not given any funds. Researchers predicted individual paper's quality rating with AI using paper and journal citation rates, title text, keywords, and other quantifying measures. The researchers found that in arts, humanities, and social sciences papers, AI results underperformed with results similar to random guesswork. In natural health sciences, biological sciences, and economics, AI performed well but never showed more than 75% accuracy, which counts as poor ranking in many research areas. These results indicate that AI is not accurate in assessing research quality. According to many experts, it is almost impossible to have an accurate AI system because human reviewers have lifetime expertise in their research. In that regard, AI has shallower knowledge (Thelwall et al, 2023).

Automation is useful for helping to find reviewers and it can sometimes help with initial quality control of submitted manuscripts. However, the value of AI to support reviewing has not been clearly demonstrated. While peer review text and scores can theoretically have value for research assessment exercises, it is not yet widely enough available to be a practical evidence source for systematic automation (Kousha & Thelwall, 2023).

Plagiarism and research misconduct

The emergence of readily available and usable AI tools has provoked discussions about what students learn and how they can falsify information (Offert & Bell, 2021). AI tools should not be used without careful

oversight from knowledgeable human researchers. For example, ChatGPT has been criticised for reporting factual inaccuracies, having weaknesses in the logical flow of its arguments, being uncritical in its selection of and elaboration on data, and lacking originality (Dwivedi et al, 2023). When researchers asked ChatGPT to produce a conference abstract, it created a well-written abstract while following given instructions; however, one of the references was completely made-up (Babl & Babl, 2023). Such gross oversight leads to spread of misinformation where the content generation process is not checked and validated for accuracy and epistemic validity.

AI's ability to synthesise and rephrase existing content makes it easy to plagiarise anything. AI can cause copyright and IP infringement by using text and images from research papers that are copyrighted. Therefore, AI lowers the bar on the required scientific quality of the original work and increases the risk of plagiarism (Elali & Rachid, 2023). In the third version of the AI Act by the European Parliament, AI providers are only required to document and disclose summary of the training data that is protected by copyright but these summaries are not enough to identify all the resources, papers, and outputs processed, making it challenging to assess copyright and patent protection (D. C. European Commission et al, 2020).

Societal concerns

Unfair appropriation of scientific knowledge

In addition to losing control over human-led benchmarking, AI researchers are also seeing a dramatic decrease in collaborative public datasets. For example, as the popularity of commercial LLMs like ChatGPT rises, contributions to public platforms like Wikipedia, Stack Overflow, etc. continues to decline (del Rio-Chanona et al, 2023).

There is also a problem called 'code capture', where computer code in public platforms are monetised. Numerous people contribute to public and private projects on a daily basis, and share their code freely with the community on GitHub. GitHub contributions typically come from people interested in a project, and programmers working for private companies on specific projects. Because many projects belong to Big Tech, free labour from non-employee contributors can be monetised by turning their work into complementary services (Rikap & Lundvall, 2022):

Top projects on GitHub by 2018/2019	Big Tech contributors on GitHub (i)	Total contributors to these projects (ii)	Code capture = (ii–i)/ii
Microsoft vs code	7700	19 000	59%
Facebook react-native	1700	10 000	83%
Google Tensorflow	5500	9300	41%

Meanwhile, US Big Tech has adopted strategies to profit from AI and dominate the AI innovation frontier. In these strategies, knowledge inflows from academia are maximised, while minimising outflows through secrecy (Rikap, 2023c). Healthy competition can incentivise better products and thus lead to innovation; but

instead of co-existing, co-evolving and co-producing innovation, firms are developing intellectual monopolies (Rikap, 2023b). Another way to control innovation is the increasing cloud service provider market share held by Amazon, Meta and Google in private, academic, and startup sectors (Rikap, 2023b, 2023c).

Violations of copyright

AI models are trained on vast quantities of published material, much of which falls under copyright protection. There is an ongoing debate about establishing ownership of AI-generated content. One argument suggests that the use of copyrighted works as training sets for AI does not interfere with copyright, and therefore it should be considered as 'fair use' or 'fair dealing'. On the other hand, many argue against this view. There is no international standard and little consensus globally on how or whether to extend copyright protection for AI-generated works (Rallabhandi, 2023).

The lack of standardised international guidelines for attribution of copyright authorship in AI-generated works has implications for literary and artistic works such as music, articles, and artwork. Such creative projects without human authors or creators can be regarded as free of copyright and placed in the public domain to be used freely by anyone (Rallabhandi, 2023). However, many argue against the value of such work that was not created by a human. With ongoing court proceedings on copyright violations by AI-generated material, this issue will require academic, legal, ethical, and stakeholder debate and consultation to satisfactorily resolve.

Security threats

Manipulation and misinformation at scale

AI's capability to generate fake information at scale including counterfeit representations of people poses a threat to humanity. Mass armies of automated bots can tip the fragile balance between information and disinformation and can also be weaponised to manipulate, control, and disrupt societies, for political or economic gain.

AI chatbots are becoming increasingly sophisticated, and there are challenges and opportunities in detecting them to mitigate the harmful effects of AI-generated conversations and behaviours (Ferrara, 2023). For example, Italian QAnon supporters designed and maintained an "infrastructure of disinformation" spanning multiple social media platforms, messaging apps, online forums, alternative media channels, and content creation platforms. Researchers found that the longer platforms remain functional, the harder it is to eradicate infrastructures: they become more sophisticated over time, get more traction, and develop a critical mass of loyal followers (Pasquetto, Olivieri, et al, 2022).

Researchers are devising new strategies to combat chatbot misinformation generated using LLMs (Chen & Shu, 2023). For example, researchers leveraged social ties among groups to maximise the re-sharing of debunking messages, such as those accessed by WhatsApp users. They found that debunking messages

received in the format of audio files generated more interest and were more effective in correcting beliefs than text-based or image-based messages. In addition, they found that users re-share debunking messages at higher rates when they receive them from people close to them (Pasquetto, Jahani, et al, 2022).

Another major security threat comes from impersonation and fraudulent digital content generation through personal information. Deepfakes and voice cloning are such state-of-the-art forgeries that most people cannot distinguish them whether they are human-generated or machine-generated (Frank et al, 2023). Bad actors can use deepfake technology to [generate non-consensual pornography.[6] AI voice scams are reportedly increasing; in a worldwide survey, 70% of people said they could not confidently tell the difference between a cloned voice and the real one.

In high-quality, natural language text output generation, researchers have developed a model to effectively preserve text utility at large scale. Using watermarks, they can decode the model while hiding its presence from adversaries. The model is also robust against a range of attacks (Abdelnabi & Fritz, 2020; Kirchenbauer et al, 2023). Similarly, model inventors are advised to fingerprint their models carried by generated samples that can be faithfully detected and attributed to the source (Yu et al, 2020).

However, LLMs protected by watermarking can still be vulnerable against attacks because humans can insert hidden LLM text signatures. With malicious intent, someone can add AI-generated text to human-generated text, causing the text to be misidentified as AI-generated, potentially damaging the author's reputation. Therefore, the AI developer community needs an open and honest conversation on using AI-generated text ethically and reliably (Sankar Sadasivan et al, 2023).

Bio-weapon development

With rapid developments in the use of AI in life sciences for automation and robotics, scientists are developing new biological materials and engineering new living systems and organisms. While most of these applications truly benefit humans by creating vaccines, biotherapeutics, and carbon-capturing microbes, the use of AI could also accidentally or deliberately cause significant harm. Given the threat of a global biological catastrophe, government, biologists, industry leaders and biosecurity experts must proactively identify emerging risks and develop strategies to prevent such threats (Carter et al, 2023).

Many viruses and biological agents, such as mousepox, H5N1, and botulinum toxins, already carry hazardous biological elements that can be further enhanced with easy access to recombinant DNA technology (Lewis et al, 2019). AI technologies can potentially inform the creation of deadlier and more virulent agents (Dybul et al, 2023). Therefore, regulators such as the EU and the US-FDA have proposed regulations for AI developers and deployers, requiring oversight and checkmarks at every stage of AI model development in life science applications (Ada Lovelace Institute, 2023b).

[6] https://www.wired.com/story/deepfake-porn-is-out-of-control and https://www.bbc.com/news/world-europe-66877718

Cybersecurity, fraud, hacking

Because LLMs and AI are trained on massive codebases for code generation, they lack an inherent awareness of security, and frequently produce unsafe code with bugs and vulnerabilities (He & Vechev, 2023). Researchers have also successfully used popular AI algorithms to create objectionable material (Zou et al, 2023).

AI tools such as ChatGPT pose cybersecurity threats, such as introducing malware. Traditional security solutions leverage multi-layer, data intelligence systems to tackle threats; however, sophisticated and automated systems could prevent these systems from working. Things can get even worse if AI-generated polymorphic malware becomes available to bad actors. Moreover, many ChatGPT users report creating ransomware with the tool with moderate success, suggesting AI will get better at creating such dangerous cybersecurity threats with ease in future. Further, researchers developed advanced phishing attacks and automated their large-scale deployment with ChatGPT (Begou et al, 2023). AI tools such as WormGPT are used by cybercriminals for malicious activities.

In response to emerging threats in cybersecurity, researchers created new AI tools to address security concerns with ChatGPT. Penetration testing is a crucial industrial practice to ensure system security. PentestGPT is an LLM-powered automatic penetration testing tool that leverages the abundant domain knowledge present in LLMs. It not only outperforms LLMs in task completion, but also proves effective in tackling real-world penetration testing challenges (Deng et al, 2023). Another tool, BurpGPT, enhances precision and efficiency of application security testing.

To counteract these malicious uses of AI, researchers are creating new models to find problematic prompts. For example, *Prompting4Debugging* is a debugging tool that automatically finds problematic prompts and tests the reliability of a deployed safety mechanism (Chin et al, 2023). Generative adversarial networks are successful at generating photorealistic images, and researchers are analysing the model's footprints to detect fake images generated by it (Yu et al, 2018). Similarly, Intel's real-time deepfake detector analyses video pixel qualities to give results in milliseconds with 96% accuracy.

Military AI applications

AI's applications are numerous in all areas. There are growing concerns about its use in military applications, potentially changing modern warfare. In addition to ethical and humanitarian risks, the major concerns are centred around the reliability, fragility, and security of AI systems. With biased or otherwise manipulated strategic intelligence, the possibility is undeniable that AI will increase the likelihood of war, escalate ongoing conflicts, and proliferate to malicious actors.

Despite ongoing United Nations discussions, an international ban or other regulation on military AI is unlikely. Broader consensus on the vital importance of human accountability in the use of weaponry for military applications is vital. Policymakers must mandate human involvement across the entire life cycle of each military AI system, including the development, regulation, and deployment. In addition, stakeholder

consultation based on mutual agreement for risk-reduction must be prioritised for the general public, hostile nations, and other states and non-state actors attempting to develop military AI (Morgan et al, 2020).

Because of the dangers of military AI, scientists have a great responsibility to facilitate discussions regarding the appropriate use of science. Researchers must engage with the general public to address societal issues and concerns and be vigilant in decision-making about preserving democratic values (Bird, 2014).

Key findings

Little uncertainty

These key findings are supported by a large body of evidence and systematic analyses. There is little uncertainty.

- State-of-the-art AI models and systems lack transparency, and commercially-created opacity adds complexity in reaching transparent and reproducible results and creates dependency for academics on industry-provided models and services.
- Many current AI models perform poorly because poor data was used to train them. This is due to low input data quality, failure to update the model, and inherent differences between training data and real-world population.
- Potential opportunities may develop for AI uptake in qualitative and theoretical development research, in the humanities and social sciences. No systematic evidence of those opportunities is currently available.
- Social-cultural bias reflected in underlying datasets are also reflected in the AI systems outputs. Additional new forms of 'machine bias' stemming from the system itself have also been observed.
- Popular and lucrative sciences and the researchers working on them tend to benefit from more funding, thus casting aside other crucial research. Additionally, deep inequalities exist in funding and access to infrastructure between industry, leading AI research efforts, and public research.
- AI tools are not yet able to reliably perform peer reviews or assess research grants. However, these tools are adding to the strain of the scientific publication system through the generation of automated misinformation and their potential to be used to create paper mills or predatory journals.
- AI Big Tech companies have adopted strategies to profit from AI and dominate the AI innovation frontier. In these strategies, knowledge inflows from academia are maximised while minimising outflows through secrecy.

Some uncertainty

There is some evidence to support these key findings, but some uncertainty exists.

- Current research on AI has shown its potential to lead to manipulation and misinformation at scale, bio-weapon development, cybersecurity, fraud, hacking, deepfake, and military AI applications. Researchers developing AI systems are aware and somewhat monitoring these threats, but currently lack guidelines on regulations and governance.

Chapter 5. Impact on scientists' and researchers' work environments, careers, skills and education

As demonstrated by the evidence presented in Chapter 3, AI is being taken up increasingly across many areas of research and applications within the research process. There are still many challenges and risks that come with the use of AI in research, as emphasised in Chapter 4.

In this chapter, based on current knowledge of the applications of AI in research and its limitations, we address the impact and requirements of AI on the education and careers of the scientists and researchers of today and tomorrow, including the relevant skills, competencies and tools, and the impact on the research workforce.

AI impact on research jobs and careers

Academic research careers under pressure

In order to implement AI in research careers appropriately, there is first a need to consider the present challenges in research careers. There is currently a high level of perceived mental wellbeing problems in research professions: 32%-42% of academic employees are at risk of developing common psychiatric disorders (Levecque et al, 2017), with still very low awareness of these issues (Guthrie et al, 2017; Kismihók et al, 2019; Mattijssen et al, 2020).

These issues stem from working conditions which include:

- **Unattractive career prospects:** There has been a continuous decline in the number of permanent academic positions per researcher at universities, and increasing dependency of researchers on short-term, third-party funding (Glausiusz, 2019). In 2018, 70% of doctoral candidates, postdocs, and tenure track researchers had to seek employment outside of academia (Woolston, 2018). In addition, when opportunities to stay in academia do arise, they are often coupled with significant job insecurity due to short fixed-term contracts (Kismihók et al, 2019).
- **Limited funding opportunities:** Conducting meaningful research is expensive both in terms of time and resources. At the same time, competition for research funding and resources is increasing

(Kismihók et al, 2019). While funding systems are moving away from core basic funding for research institutions to allocation via competitive mechanisms, researchers are facing funding variability over time, stagnation in funding, and future uncertainty leading to a demand for flexibility in staffing and less permanent positions (OCDE, 2021). The success rate of funding for research grant proposals for Horizon Europe in 2022 was 15.9%. The European Commission reports that 77.1% of high-quality proposals do not receive funding.

- **Highly demanding (but not rewarding) diversification of skills:** Competition for funds puts a premium on researchers who can demonstrate high performance against easily measured indicators, such as citations for publications, and ability to attract research funds (OCDE, 2021). However, researchers are also expected to be adaptive and capable communicators, experts in research and (open) data management, effective networkers, able to manage stressful situations in their research, and at the same time remain open, innovative, and constantly mobile. Academics call for more emphasis on transferable skills training and recognition (Kismihók et al, 2019).

Internal conflicts, work-life balance and financial problems in some regions further compound the challenges (Kismihók, 2021).

Against this backdrop, a Declaration on Sustainable Researcher Careers was published in 2019, calling on research institutions, funding bodies and governments to ensure sustainable researcher careers (Kismihók et al, 2019). The Researcher Mental Health Manifesto (2021) also contains evidence of these issues, as does further research (Kismihók et al, 2022).

These challenges must be seriously considered while developing the uptake of AI in scientific processes. AI tools influence the research process, from ideation to publication, and researchers need to adapt to this new work environment to remain competitive.

AI could support rather than replace researchers

A large number of reports address issues around the impact of technology, including AI, on work and the workforce. The evidence specifically on researchers and scientists is much smaller, and potential effects of AI technologies on the research job market are not yet well understood.

AI has made most progress in its ability to perform non-routine, cognitive tasks such as ordering information and memorisation, mostly impacting on high-skilled occupations. The OECD employment outlook (OECD, 20230 emphasises that we should not be led by "technological determinism" where technology shapes social and cultural changes, but rather we should ask what AI can do for us.

Kabashkin et al (2023) proposed a rethinking of the university model in the context of the emergence of AI. They suggest that, in the era of AI, human competencies that cannot be replaced by AI, or those necessary to develop and apply AI, will grow in importance. A recent structured systematic review investigating how automation technologies affect employment suggests that science and academia are occupations with a low probability of full automation, because they involve non-routine work activities and require specialised

knowledge, analytic thinking, creativity, and imagination (Filippi et al, 2023). It is suggested that about 40% of knowledge workers' time is spent on activities that bring little personal satisfaction and could be delegated to others, and AI could potentially take over such activities (Wade in Dwivedi et al, 2023). Therefore, some argue that AI technologies can free researchers' time to focus on activities that cannot be delegated to computers, while relieving them of repetitive and menial tasks (Laumer in Dwivedi et al, 2023; Jardim et al, 2022; Rahman & Watanobe, 2023).

For example, in the future these may include AI writing code for statistical analysis; analysing and classifying large amounts of data; simulating and testing complex procedures; and assisting in writing and formatting manuscripts for submission (Burger et al, 2023; Esplugas, 2023). AI could also potentially conduct systematic reviews, and AI tools have been used to facilitate processes such as screening records, classifying studies, and assessing risk of bias, thus reducing the time needed to produce a review and decrease costs (Jardim et al, 2022). However, these potential applications in which AI relieves researchers of menial tasks should be put in context of the challenges and risks associated with the use of AI in research, as presented in Chapter 4.

A study of an international news database spanning 956 articles from 122 newspapers published in 2020 suggests that, in research and development, AI is primarily being used to enhance human work rather than to replace humans, thus transforming job roles and skill requirements, but not necessarily endangering a large number of jobs (Johnson et al, 2022).

Public-private partnerships in AI impact researcher careers

On the one hand, a UNESCO conference report on AI and education, *Planning education in the AI era: Lead the leap* (2019) identifies areas of opportunity where public-private partnerships could impact the progress of using AI in education, including cloud infrastructure, data storage, computational resources, apps, services, development and licensing, hardware and in-school devices, infrastructure, communication and access, operations, security, maintenance and cybersecurity protection; expertise, evaluation, usage training and efficacy measurements; and AI education support such as courses, resources, competitions and incentives for learning about AI. An open-source platform could be established to which all can contribute, to help leverage open resources, translation, and adaptation to local contexts. AI algorithms would also be shared (UNESCO, 2019).

On the other hand, a study of the career paths of AI researchers highlighted the interplay between academic and industry research in the field and presented evidence regarding a potential brain-drain from the public sector. The researchers demonstrated that the increasing involvement of the private sector in research of AI and research using AI coincided with a noticeable migration of researchers from academia to industry, particularly within technology companies like Google, Microsoft, and Facebook. Survival analysis indicated that researchers working with deep learning techniques, driving recent advances in AI systems, had a significantly higher likelihood of transitioning to the industry. This finding supports the idea that the private sector is actively building capabilities in state-of-the-art AI systems, raising concerns about the ability of 'public interest' deep learning research to keep pace, especially since industry tends to attract influential and high-impact researchers (Jurowetzki et al, 2021).

Public private partnerships also lead to unbalanced recognition of the work from the academic researchers (see Unfair appropriation of scientific knowledge).

Facilitating human-machine collaboration and co-creativity

AI can play an important role in facilitating human-machine collaboration, while helping with the execution of tedious tasks (Lane & Saint-Martin, 2021).

A report by ESIR (2023) highlights that the approach of "Industry 5.0" puts a stronger emphasis on "good jobs", meaning a human-centric, resilient and sustainable approach to transformation. This involves paying due attention to human-machine cooperation and redefining job quality, taking a new approach to skills and sustainability. Human-centricity means that AI and other technologies are not taken as the ultimate goal, but instead as a means to empower and enhance work. In short, not everything that *can* be automated *must* be automated. Policy should focus on the development of technologies that can be trusted and are compatible with workers' rights and wellbeing. Workers should be able to concentrate on intellectually stimulating tasks, rather than repetitive ones. It also means that humans should be able to oversee the functioning of machines and that workers should be able to co-design systems. This requires the setting down of transition pathways within organisations (ESIR, 2023).

Aarhus University has developed workshops for organisations to experience and explore the potential of AI in their workflows, its impacts and potential to reshape the workflow patterns. To accept the technology, a crucial aspect is that end users need to be engaged in the design process. This calls for better understanding of what determines their buying into designing and training these tools (Mao et al, 2023). Applied to the field of science, these workshops propose research challenges in the form of games, to understand the flow between intuitive and computational interactions. The people who helped the most to develop these interfaces were the researchers themselves, because it allowed them to think differently about their complex tasks, and to present and pose them in completely new ways. This can be investigated in the domain of citizen science (Rafner, Gajdacz, et al, 2022). Another experiment also showed that, in a game, AlphaZero combined with detailed methodologies of quantum researchers in a hybrid approach was the most efficient (Dalgaard et al, 2020). So these hybrid interfaces should be investigated much more systematically.

With such an approach, where humans and AI are considered as integrated, workflows in the real world (in companies or in the research workflow) can be addressed differently and the technology can be adapted. To reap the benefits of AI tools and especially generative AI, the non-AI experts need to be considered as AI-innovators and the challenge is to get them to become innovators. That cannot be done by merging AI and human-centred AI only, but by bringing in all the fields of research (Rafner, Bantle, et al, 2022) and by moving towards an interdisciplinary framework of hybrid intelligence, involved from development to deployment. For example, in the management sector, this would entail the implementation of a pact between leadership and employees, stating that a goal of transformation is not necessarily optimisation of processes, but rather the upskilling of employees. This creates a psychological safe space which opens up opportunities for creating integrated solutions. Hybrid intelligence is therefore not only an interface, but

also a structure that allows to focus on upskilling and empowerment, and on the rethinking of the entire value generation stream. There could therefore be great value in understanding more deeply the human role in all aspects of a job-cycle, at the risks of deskilling and the potential for upskilling (Elrod & Tippett, 2002; Rafner et al, 2021). There is great potential in thinking about integrating AI into the organisational context, so that end users also become co-creators and co-developers (Sherson et al, 2023).

There remains a challenge of understanding human-AI co-creativity. Creativity is essential in a researcher's career, but currently research in the area of human computer interactions and AI is lacking insights from psychological sciences, so there is a need to develop a theory of human-AI co-creativity (Rafner et al, 2023) that is strongly informed both by psychological sciences as well as by human-AI interaction.

AI impact on researchers and research environments

AI may cause worker deskilling

Although automation and machine support can increase efficiency and lower costs, it can also, as an unintended consequence, de-skill workers, who lose valuable skills that would otherwise be maintained as part of their daily work. Through deskilling come risks of lacking critical thinking skills and confirmation bias in academics, along with the lack of accountability and transparency of AI systems (Rafner et al, 2021) (see Other ethical concerns).

With increasing AI applications, worker profiles might change, for example to the profit of non-academically-trained workers who can work side by side with AI (Xue et al, 2022).

AI may create larger gaps: potential inequalities

Gender gap

Workers with AI skills tend to be disproportionately male and tertiary-educated (OECD, 2023). CEDEFOP (2023), a decentralised EU agency that supports the development of vocation and educational training, states that in almost all countries, the share of female adults with at least basic digital skills is lower than the share observed for males. There is also a gender imbalance in information and communication technology employment: in 2021, only 19.1% of those employed by companies in the IT sector were women. Male-dominated developer teams design AI systems and applications (CEDEFOP, 2023). A lack of diversity in AI development teams will tend to reproduce and perpetuate gender biases in the technologies they develop. Cultural, societal and political values are inherent in AI systems. AI may exacerbate gender inequality in labour markets, including increasing the pay gap (Gomez-Herrera & Koeszegi, 2022).

According to a report by Bruegel (Gomez-Herrera & Koeszegi, 2022), the sudden disappearance of a percentage of jobs and the creation of a new set of jobs could affect different groups of people unevenly,

with a more substantial impact on groups at risk of exclusion. There could be a mismatch between skills and technologies in the short and medium terms. The situation may favour some geographical regions and population groups over others. This may create or enlarge existing gender, inter-regional, generational and income inequalities.

Geographical disparities

A report by CEP (2021) highlights some of the particular issues faced by Central and Eastern European countries. These countries have a particularly high share of new firms, which are too small to develop their own technologies and are dependent on external technologies and services. Data supply is also an issue. Access to AI technology is therefore key, but there exist major inequalities in access across regions and sectors (Mihai et al, 2023)

People with disabilities. Despite advancements, groups with disabilities are still limited on institutional and technological levels which affect not only their lives, but available access to research, data, algorithms and systems (Whittaker et al, 2019).

AI may perpetuate existing biases and widen the gap between academic and industry research due to unequal access to computational resources and infrastructure (see Fundamental rights protection and ethical concerns).

AI and digitisation may negatively impact workers' mental health

In a system that is already under pressure, researchers may be particularly affected by the uptake of AI in the workplace. According to EMPL (2022), only a small number of studies focus on the direct impact of new technologies on working conditions or health outcomes. A growing body of literature indicates an association between digital transformation and mental workload and, in particular, a trend towards borderless work settings and different forms of work. Job insecurity and fears of unemployment tend to increase when a job has a high potential of being performed by machines.

'People analytics' operate in human resources, from recruitment and hiring practices using psychometric tests to digitalised interviews. This may allow employers to increase control over their workers and the workplace, incorporate rating systems or other metrics into performance evaluation, improve workers' performance and productivity, rationalise the organisation of work and production, reduce the cost of monitoring and surveillance, profile workers, influence their behaviour, discipline them, and improve HR management (STOA 2020, cited in EMPL, 2022). The OECD stresses that, in regular work environments, the use of workers' data to reward or penalise them could lead to job insecurity and stress. The use of automated decision-making systems for personnel selection entails a risk of discrimination as AI has the potential to produce results that are inaccurate or biased, and therefore lead to unfair and discriminatory decisions (OECD, cited in EMPL, 2022).

AI literacy and competencies

Skills and competencies for users and developers of AI

Evidence suggests that many new jobs will be created for workers with AI skills, or who have the necessary skills to work with AI. There is also evidence that workers with AI skills may earn a substantial wage premium and have improved job quality (OECD, 2023). In research, new roles that are emerging include many types of 'data scientist', some of whom are supporting research and others who are actively involved in conducting research (OECD, 2020).

The OECD (2023) Employment Outlook lists skills needed in the age of AI and digitisation. The report differentiates between the skills needed by AI professionals, who will develop and maintain AI systems, and those needed by users and workers who will interact with AI systems. AI researchers and technologists should aim to develop skills in AI (machine learning, models, tools) and data science (analysis, software, programming, visualisation, and cloud computing), along with creative problem-solving abilities and social and management skills. AI users should aim to acquire elementary knowledge of AI such as the principles of machine learning along with analytical skills, judgement and creativity (OECD, 2023).

For both users and developers, using AI is likely to require critical thinking skills, including the ability to challenge and interrogate knowledge, and identify hidden or encoded biases. This needs to be combined with ethical awareness, a consideration of how AI should be regulated and constrained, and how specific algorithmic behaviours might cause harm to and benefit different groups. AI is also likely to disrupt traditional career paths, requiring workers to operate in a way that is more self-reliant, particularly in terms of personal and professional development (Brown, 2023).

Urgency in skilling AI workers

The set of skills required for AI jobs is strongly related to mathematics (statistics, calculus, algebra, algorithms, probability), science (physics, cognitive learning theory, language processing) and computer science (data structures, programming). Expanding and developing the science, technology, engineering, and mathematics workforce is a critical issue for governments (Gomez-Herrera & Koeszegi, 2022). CEDEFOP (2023) reports that 70% of EU companies reported a lack of adequate digital skills.

The European Commission's digital targets for 2030 aim to train 20 million ICT specialists by 2030, or to achieve basic digital skills across at least 80% of the European population. The workplace increasingly demands AI skills, especially for highly skilled jobs (OECD, 2023), but for some jobs, it is currently unclear whether they will disappear or be boosted if an AI tool can be used by the employees.

What does AI literacy involve?

'AI literacy' is defined as the broad general knowledge and skills of individuals who interact with AI technology (Schleiss et al, 2023); as the ability to understand and use AI concepts for learning about and evaluating the real world (Kong et al, 2022); or as the ability to effectively communicate with AI and evaluate the trustworthiness of its output (Pretorius, 2023).

Upshall (2022) identifies five essential skills for the assessment of AI tools, namely being able to:

- distinguish between tools that do and do not use AI
- analyse differences between human and AI
- identify AI-based technologies
- distinguish between general and narrow AI
- identify the types of problems that AI can solve easily and those that it struggles with

Alternatively, a conceptual framework of AI literacy suggests that AI literacy consists of three dimensions (Kong et al, 2023):

- **cognitive**, whereby people need to be educated about basic concepts of AI, such as machine learning and deep learning, and learn how to use them to enhance their understanding and evaluation of the world
- **affective**, empowering people to gain confidence to participate in the digital world
- **sociocultural**, concerning the ethical use of AI, such as not violating human autonomy, ensuring that the benefits of AI outweigh the risks, and distributing the benefits and risks equally among people

In addition to these dimensions, AI literacy should also include awareness and knowledge about availability and use of computing infrastructure and awareness of the evaluation of the tools available (for example, whether they are open-source).

Digital skills and competencies for researchers

What digital skills do researchers need?

The DigComp framework was designed as a digital skills competency framework for citizens rather than researchers, but it includes a full introductory curriculum for AI (Rina et al, 2022, Appendix A2) that includes, in some detail:

- What do AI systems do and what do they not do?
- How do AI systems work?
- When interacting with AI systems
- The challenges and ethics of AI
- Attitudes regarding human agency and control

An evaluation of the competencies encompassed in the European Commission's Digital Competence Framework for Citizens (DigComp, cited in OECD, 2020) was undertaken to assess the relevance and adequacy of these competencies for science. DigComp consists of five areas of digital competence, and each area contains general competencies that can be usefully applied to research. These are likely to vary, according to the discipline or field. They are:

- **Information and data literacy**: Browsing, searching and filtering data; critically evaluating credibility and reliability of data sources; organising and storing data. In addition, understanding of statistics and the requirements for reproducibility are important.
- **Communication and collaboration**: Sharing data; knowing about referencing and attribution practices; using digital tools and technologies for collaborative processes; protecting one's reputation. In addition, following open science principles and referencing/attribution practices are important.
- **Digital content creation**: Creating new, original and relevant content and knowledge; understanding copyrights and licences; programming and software development. In addition, the visualisation of data and information is important.
- **Safety**: Protecting personal and sensitive data.
- **Problem solving**: Customising digital environments to personal needs; using digital tools to create knowledge and innovate processes; identifying digital competence gaps and seeking opportunities for self-improvement.

Skills needed for the responsible and ethical uptake of AI in research

Some of the skills that are likely to be required to successfully integrate AI into research are critical thinking and the ability to add value to AI output (Laumer in Dwivedi et al, 2023). Fact-checking is another important skill because researchers implementing AI in their work need to be able to verify the claims made by it (Ahuja in Dwivedi et al, 2023).

As for the hard skills required for research and how they may be affected by AI, coding is one example. There are a number of tools that can generate code for statistical analysis, including ChatGPT and rTutor.ai (Wright & Sarker in Dwivedi et al, 2023; Merow et al, 2023). However, at this moment, tests indicate that such tools still require oversight from researchers skilled at coding. The output of generative AI still relies on the knowledge and skillset of its user, who is responsible for providing prompts and assessing the accuracy of the output, so being able to communicate with such tools effectively is an important ability, and part of AI literacy (Pretorius, 2023).

The complete research process is affected from ideation to publication, and a number of AI tools are already available to assist (see Chapter 3). Researchers rapidly need to adapt their work environments and their competencies to remain competitive.

More research needs to be done about the impact of AI on soft skills which likely become vital (see AI could support rather than replace researchers), such as negotiation, intercultural dialogue, equality and diversity management, mediation, mentorship, and others.

Education and training in the digital era

Challenges of AI for academic education

A report by EPRS (2022) lists potential gaps and barriers for digital transformation within the EU, including a shortage of digitally skilled workers. The JRC's report (López Cobo et al, 2019) states that there is a persistent ICT skills gap in Europe, and the 'offer' from universities lags behind the market. Potential actions include increasing mobility of the workforce, consideration of other types of training including massive online open courses and vocational training, analysis of data from job websites, and improving the connection between industrial needs and the educational offer. In its 2023 report (Tuomi et al, 2023), the JRC looks at other alternatives to formal education, including digital credentials and modular training. It suggests that AI can become a 'learning partner'. It highlights the need for non-epistemic competencies, and recommends the interlinking of educational, digital, environmental, and industrial policy.

Training is needed to develop, use and reflect on AI (OECD, 2023), focusing not only on technical skills but also on skills to adopt, use and interact with AI applications. However, this poses a challenge if researchers need to acquire technical, transversal and domain-specific skills at the same time.

A significant challenge is the need to teach AI methods balancing breadth and depth in AI competencies and skills, covering diverse topics while incorporating interdisciplinary and ethical aspects. National expert organisations should be involved when designing curricula, along with experts from the European Association for Artificial Intelligence and Confederation of Laboratories for AI in Europe. Promoting Europe-wide cooperation in the field of AI skills and competencies, as well as AI tools, should be based on principles of cooperation and openness, including existing open source and open educational resources. Rather than starting from scratch, efforts should identify existing resources that have successfully supported AI skills development in member states.

UNESCO's report on the effects of AI on working lives of women (cited in Gomez-Herrera & Koeszegi, 2022) puts forward as a major finding the need for reskilling and upskilling women workers. It is crucial that women are not left out of the increased demand for professionals in science, technology, engineering and mathematics, and in AI specifically. Algorithms may perpetuate exclusion and discrimination in the education of people with disabilities further due to the lack of access to data for target populations, unconscious or conscious bias, or existing social practices. But AI could also provide access to more inclusive digital solutions.

Examples of current AI training programmes

The OECD's (2023) report finds that the AI workforce in OECD countries is still relatively small but growing rapidly. The report also finds that demand for AI workers is strong, and there is some evidence that the supply of workers with AI skills may not be keeping up with demand in many OECD countries. It is therefore crucial that adequate training programmes be developed and provided.

Based on an analysis of 11365 Coursera descriptions, it is estimated that 369 courses on AI are already available, which is encouraging, with many bottom-up projects in the field, such as Online Open Learning Recommendations and Mentoring Towards Sustainable Research Careers (OSCAR) and eDoer, an open community-based AI-driven learning platform. There are hackathon formats that help researchers prepare for to this new working environment, such as OEduverse. On research data literacy, one can mention the Data Literacy Alliance. There are also initiatives such as European Skills, Competences and Occupations which show the emergence of AI in occupations and skills across the EU.

Some relevant examples of existing AI teaching programmes in Europe are described below.

- **Finland**: To scale out and up the AI teaching, courses can be automated, such as in Elements of AI. For example, the University of Helsinki in Finland created an online course with the goal of educating 1% of the Finnish population, which was achieved in one year. The scope broadened to the world, and Sweden was the first case study. Courses and examinations can be started and completed at any time, challenging the traditional 'university' concept. Over 1 million students from over 170 countries have signed up for the Elements of AI course. About 40 % of course participants are women, more than double the average for computer science courses.
- **Germany**: Germany has an initiative called AI Campus which provides broad, accessible AI courses and resources, promoting cross-institutional cooperation and addressing diverse educational needs. These resources are open source and used in many German-speaking countries in the EU, spanning from 'Explainable AI for Engineers' to 'Data Literacy for Primary Schools'.
- **Sweden**: The Wallenberg AI Autonomous Systems and Software Programme has a budget of €600 million over 15 years and a goal of educating at least 600 PhD students. In the area of AI autonomous systems and software, the programme currently has around 450 PhD students, and so far, approximately 100 PhDs have been produced. It is a large-scale effort with a strong focus on education and competence development in this area. In the WASP-ED AI Curriculum (Lindgren & Heintz, 2023), the technical core remains related to core AI functionality, but the curriculum was broadened to become more socially and individually embedded (e.g. human-AI collaboration) and to include societal perspectives (e.g. ethical, legal, social, economic, cultural aspects, trustworthy AI). The programme also addresses socially and physically constrained infrastructures (distributed and edge AI, robotics, control and autonomy systems), applied AI in research and society, the history and futures of AI and fundamentals of knowledge. Events and seminars are regularly organised, thus building a community around the programme.
- **Czechia**: prg.ai was founded in 2019 as a coordinated effort of academics from the Czech Technical University, Charles University, and the Czech Academy of Sciences, with a significant contribution

provided by the city of Prague. They launched prg.ai Minor, a programme designed for students who aim for in-depth interdisciplinary understanding of AI and practical skills to apply it in various areas of interest. They are a national partner of the world-leading Elements of AI course and are administering its Czech version.

- **Netherlands**: The Masters specialisation AI at Leiden University offers future-oriented topics in computer science with a focus on machine learning, optimisation algorithms, and decision support techniques. It allows students to pursue careers in research or industrial environments.
- **Ireland**: The Masters in AI course from the National College of Ireland aims to educate graduates who will become leading practitioners in the field of AI. It contains modules covering the complete development lifecycle of AI, including fundamental and specialised AI topics as well as topics related to the operationalisation and application of AI to solve real-world problems, including evaluation, ethics and governance.
- **Italy**: An online course from Politecnico Milano, available in 20 languages, provides an overview of AI, including ethical and legal issues, machine learning tools, and automated learning approaches.
- **UK**: The National Institute for Health Research offers an introduction to AI for clinical researchers, including an overview of the current landscape of AI and machine learning technologies, key AI definitions and concepts, and evaluation metrics to assess performance, overview of organisations able to support AI projects, practical challenges, regulatory and ethical requirements.
- **Switzerland**: Many universities offer Masters degrees in AI, the first being the AI Master of Università della Svizzera Italiana in 2017. Universities of applied sciences also offer bachelor programmes to educate professionals to use AI tools and methods.

Many additional publicly available courses are also available from the private sector, for example from Google or IBM.

Strategies for integrating AI education into existing scientific curricula

A paper by the OECD (2020) examines human resource requirements for data-intensive science. It recommends that resources need to be appropriately curated, made interoperable and preserved. Digitally skilled researchers need a set of foundational digital skills, coupled with domain-specific specialised skills. Suggested actions for universities to strengthen digital workforce capacity and skills for data-intensive science include:

- providing training for scientists and research support staff
- developing new career paths with appropriate evaluation, recognition and reward mechanisms

Some of these actions can be built on existing structures; for example, university libraries can support the development of data management skills, and computing departments can help to propagate software and coding skills. Other organisations and entities, including science associations, academies, research institutes, research infrastructures and the private sector, will have roles to play. International collaboration and the sharing of best practices are important (OECD, 2020)

Kabashkin et al (2023) proposed a model of universities for the age of AI, whose main objectives are fostering innovation, creativity, and critical thinking. According to this work, the role of universities will shift to forming human competencies that cannot be replaced by AI, or those that are required for the development and application of AI.

While the AI education norms are changing, universities and educators are faced with an immediate issue of the availability and use of tools such as ChatGPT in academic environments, despite the current uncertainties and limitations associated with these tools (see Chapter 4). Universities have taken a range of different approaches to this issue, from sanctions and reverting to pen and paper, to fully embracing the technology. Some suggest that the steps to be taken remain unclear; for example, universities may create publicly funded LLMs in collaboration with open, stakeholder-led initiatives like the BigScience project. This situation calls for the development of common standards and norms by educators in Europe (Milano et al, 2023).

Educating researchers in the age of AI will seek to broaden technical knowledge and skills, and beyond this, academics are recognising the need to teach future scientists how to think through ethical, cultural, and social dilemmas associated with the development and use of AI (Dignum, 2020). More transdisciplinary approaches will support the adaptations of research funding and education policies to continue to promote ethics, equity, diversity, and inclusion in science and engineering (CCA, 2022).

Beyond training, the Final Report of the Open Science Policy Platform (European Commission, 2020) suggests moving towards an academic career structure that fosters outputs, practices and behaviours to maximise contributions to a shared research knowledge system. In the context of AI, one could infer that shared training, literacy and skills practices and standards will support the development of an AI-in-research research system that is inclusive, diverse and responsible.

Key findings

Some uncertainty

There is some evidence to support these key findings, but some uncertainty exists.

- Research careers and jobs will be impacted by AI. Current evidence shows that additional digital skills and AI literacy will be required for most researchers.
- Different skills will be required for users and developers of AI, with the common need to understand the underlying ethical and governance requirements of the technology.
- Public-private partnerships could benefit the landscape of AI education and literacy, but in the current landscape, these partnerships could also be harmful to recognition of the knowledge provided by the academics.

- AI systems and tools have the potential to enhance rather than replace humans, and in particular researchers, through human-machine collaborations fostering upskilling and creativity.

High uncertainty

There is little evidence and no systematic analysis to support these key findings.

- Additional requirements for academics to acquire digital skills and AI literacy may add onto the already high-pressure academic environments.
- Education and training in AI are being built into university curricula, and increasingly in demand. As they develop, there are risks that inequalities might leave some groups behind in the process of digitalisation.

Chapter 6. Evidence-based policy options

The preceding chapters have demonstrated that AI is already having a significant impact on scientific research, both in the field of AI research, and in scientific research more generally. As a powerful tool for knowledge discovery, AI has already enabled major advances in scientific knowledge, facilitating the automation of a number of tasks involved in the process of scientific research, although the takeup of AI across scientific disciplines is uneven. In the absence of peer-reviewed scientific studies, this report has identified numerous examples in which AI has enabled scientific research across a wide range of disciplines, particularly in the natural and mathematical sciences, and in computational methods within sub-disciplines in the social sciences and the humanities. These examples demonstrate how AI has considerable potential to transform science in ways that are beneficial to the scientific endeavour and to society more widely. At the same time, there are a number of threats, risks and concerns associated with advances in AI research, and the takeup of AI software by academic researchers.

This report is focused primarily on the implications of AI in scientific research, rather than its implications for society more generally. In particular, it does not address the manifold and very significant challenges that arise from the growing and rapid deployment of AI technologies in specific social domains, and which fall outside the scope of this report. Yet science cannot be separated from society: it is a product of social practices and human communities, and it is ultimately a shared human endeavour that is institutionally committed to the pursuit of epistemically sound knowledge and understanding.

While this is true of all scientific knowledge, it is especially true of AI research for at least two reasons:

- Firstly, AI is a general-purpose technology. Accordingly, AI can be used in an extraordinarily wide and varied range of purposes and domains, and to perform a wide range of tasks, deployed by actors motivated by considerations that may be noble, self-serving, or malicious. In other words, AI applications range far beyond that of the research laboratory, with AI-driven innovation now driving digital transformation worldwide.
- Secondly, AI operates through software powered by computational systems that can be deployed rapidly, automatically, at scale, operating in real-time, all made possible by the global data infrastructure which the internet has become. Thus, its capacity and reach are formidable while its interaction with the social world is non-linear and complex, often producing unintended consequences that may not have been anticipated or readily foreseeable.

To describe AI as “revolutionary” and akin to the discovery of fire is not hyperbolic: it is an apt depiction of its transformative power (Buchanan & Imbrie, 2024) evoking both its power and promise, and its potentially destructive capabilities. Hence the vital importance of learning how to handle AI, both collectively and individually, wisely, and well.

Chapter 6. Evidence-based policy options

This chapter reflects on the key findings set out in the preceding chapters, taking account of its current and anticipated benefits, and the risks and dangers associated with the takeup of AI in research, in light of the sociopolitical and economic context in which AI research now occurs. In so doing, we draw on existing literature and the views and insights of invited experts who participated in our Expert Workshops, in particular the Policy Design Workshop held on 10 January 2024, convened with the aim of responding response to the guiding question set out in the Scoping Paper for key area 4:

> How should the Commission (through policy initiatives, regulation, communication and outreach) facilitate responsible and timely AI uptake by the scientific and research communities across the EU (including providing access to high quality AI, respecting European Values)?

Based on these findings, this report identifies five broad challenges that confront EU policymakers that may help to accelerate the responsible and timely uptake of AI in scientific and research communities, thereby supporting European innovation and prosperity. In this context, 'responsible' is taken to mean that accelerated uptake of AI should strive to be in accordance with the foundational commitments of scientific research and the foundational values underpinning the EU as a democratic political community and thus ruled by law, ensuring respect for the fundamental rights of individuals and the principles of sustainable development.

The primary challenge that must be addressed in order to accelerate the uptake of scientific research both in AI, and using AI for research, concerns resource inequality between public and private sector research in AI. To foster scientific uptake of AI responsibly, four further challenges must be addressed, concerning:

- scientific validity and epistemic integrity
- opacity
- bias, respect for legal and fundamental rights and other ethical concerns
- threats to safety, security, sustainability, and democracy

This report then sets out a suite of policy options which are directed towards addressing one or more of these challenges. These policy proposals include:

- founding a publicly funded EU state-of-the art facility for academic research in AI, while making these facilities available to scientists seeking to use AI for scientific research, thereby helping to accelerate scientific research and innovation within academia
- fostering research and the development of best practices, benchmarks, and guidelines for the use of AI in scientific research aimed at ensuring epistemic integrity, validity and open publication in accordance with law and conducted in an ethically appropriate manner
- developing education, training, and skills development for researchers, supplemented by the creation of attractive career options for early career AI researchers to facilitate retention and recruitment of talented AI researchers within public research institutions
- developing publicly-funded, transparent guidelines and metrics, using them as the basis for independent evaluation and ranking of scientific journals by reference to their adherence to

principles of scientific rigour and integrity. The publication of these evaluations and rankings would be intended to provide a more thorough, rigorous, informed, and transparent indication of the relative ranking of scientific journals in terms of their scientific rigour and integrity than existing market-based metrics devised by industry, helping to identify predatory and fraudulent journals

- establishing an EU 'AI for social protection' institute, which engages in information exchange and collaborates with other similar public institutes concerned with monitoring and addressing societal and systemic threats posed by AI in Europe and globally, proactively monitoring and providing periodic reports and making recommendations aimed at addressing threats to safety, security, sustainability, and democracy

We note, however, that formulating policy in response to technological innovation is a notoriously fraught endeavour due to the pervasive uncertainty that surrounds them (Brownsword et al, 2017) and we therefore avoid the employing the language of 'solutions' or the mindset of solutionism (Yeung, 2023). Moreover, the scale and complexity of many of these challenges are very high and cannot be easily nor quickly resolved. Accordingly, the policy options canvassed below are unlikely to be sufficient to 'solve' the challenges identified in this report but are better understood as potentially valuable starting points. Moreover, given the nature and magnitude of this uncertainty that surrounds technological innovation, we underline the importance of putting in place institutional measures to systematically monitor and report on the impact and effect of AI in science and across society as its takeup advances and as the technology itself matures. It is in this spirit of humility that these policy options are offered for consideration.

The challenges

Opacity, scientific validity and epistemic integrity

Although many scientists have long drawn attention to science's 'reproducibility crisis', the takeup of AI is exacerbating this problem. In Chapter 4, we identify a number of ways in which the takeup of AI tools has compounded and magnified problems of reproducibility. Many of these problems can be attributed to the opacity of AI tools, arising from multiple sources:

- Researchers themselves may fail to provide sufficiently complete or detailed information to allow others to replicate their findings. The reproducibility of machine-generated outputs requires information about the code, data and computing infrastructure employed to produce a given set of research findings (Gundersen et al, 2018; Henderson et al, 2018; Hutson, 2018a; Montgomery, 2019).
- The accuracy of AI models is a product of the underlying data upon which it is trained. Yet human researchers make important choices when identifying, collecting and curating the data upon which to train their model (Boyd & Crawford, 2012; Gitelman, 2013). Yet in real-world practice, the provenance of datasets and the 'journey' through which that data has proceeded may be highly obscure and practically impossible to trace (Leonelli, 2023a), preventing disclosure of its provenance and the contextual conditions from which it originated and was subsequently parsed.

- Poor outcomes may arise when AI tools are employed in scientific research due to failure to update the model, and inherent differences between training data and real world population. Reproducibility requires that all these factors are adequately disclosed to enable proper evaluation of the model.
- Some AI systems (including those based on deep learning neural networks) operate as 'black boxes', in that is difficult or even impossible to explain how their results were generated or the specific features in the data that produced the results identified and this, in turn, may make it practically impossible to identify spurious correlations. Moreover, without the ability to explain and interpret the results of an AI model, this may preclude revisions and refinements that would otherwise lead to performance improvement, while making it difficult to detect unfairly discriminatory outputs.
- The most recent very large AI models may generate special challenges, even for researchers developing their own AI models, when attempting to publish their work in an open and reproducible manner (concerning, for example, the need to respect IP and privacy rights). In particular, researchers who use commercial, closed source models are thereby precluded from providing the level of transparency in relation to the underlying models, data and computational conditions employed that would typically be required for the purposes of scientific publication (W. Wang et al, 2023).

Without more sustained, systematic attempts to address these concerns, there is a substantial risk that AI research, and the use of AI in scientific research more generally, will continue to be plagued by invalid and erroneous scientific outputs. This risks bringing the scientific endeavour increasingly into disrepute, and undermining public trust in science more generally. For example, in cases of misdiagnosis by AI in a clinic (Păvăloaia & Necula, 2023), stakeholders may begin to resist the technology due to distrust in their predictions (Bitkina et al, 2023; Mukhamediev et al, 2022; Păvăloaia & Necula, 2023).

Bias, respect for legal and fundamental rights and other ethical concerns

The importance of cross-disciplinary knowledge and expertise also extends to the need for AI researchers, and for researchers seeking to employ AI techniques, to be appropriately trained to ensure that they have adequate knowledge of the legal rights, duties and ethical concerns that may be implicated by their research. Although the matter has not been systematically studied, there is cogent evidence to indicate that those with technical skills necessary to undertake AI research and/or to use AI for research purposes lack an adequate understanding of the ethical risks posed by its use. The potential for AI systems to unfairly discriminate between persons and groups on the basis of race and gender is well documented (Benjamin, 2019; Buolamwini & Gebru, 2018; Noble, 2018; O'Neil, 2016; Sweeney, 2013), particularly when AI systems are used to inform decisions about an individual's access to housing, hiring, educational opportunities, and the court system (Mehrabi et al, 2019; Morse et al, 2022). Although bias mitigation techniques are now a burgeoning field of AI research, the extent to which researchers who employ AI in science reflect critically on the potential for unfair bias in the datasets on which they seek to train their model, or in the outcomes produced by their model, has not been systematically studied. Although principles of research ethics are well established and administered at the local level by research ethics committees, these principles were developed in a pre-digital age in response in response to a series of scandals involving the abuse and

exploitation of research participants which came to light following World War II. These norms are focused primarily on the importance of securing informed consent from research participants, rather than attending to unfair discrimination against historically marginalised groups. As a result, group-based discrimination and broader societal harms may be given due systematic attention within conventional research ethics review (CCA, 2022).

At the same time, research employing AI may make use of datasets that include personal data, including data garnered through the use of social media platforms, for which informed consent has not been provided (CCA, 2022). FAIR data principles aim to support the responsible collection, curation and reuse of scholarly data which enhance the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. There is the potential for conflict and confusion, however, in identifying what the lawful and ethical data handling for research purposes may require in any given setting, including respect for FAIR data principles. For example, the GDPR exempts the processing of personal data for scientific research from its purview under certain conditions, but this may be at odds with the research ethics conventions requiring informed consent of individual research participants. At the same time, the use of AI tools may enable the generation of new and unanticipated insights which individuals could not reasonably have foreseen and may be unwanted, unwelcome or rights-intrusive, particularly when it entails the aggregation of data from multiple sources (Metcalf & Crawford, 2016; Metcalf et al, 2021). According to Demos, this is an evolving problem; for example, risks relating to privacy breaches increase with the linking of datasets, and possibilities of introducing bias into decision making processes increase as we rely more on AI trained on datasets that themselves contain hidden biases (Procter et al, 2020). In addition, the use of machine learning techniques may create new forms of machine-generated bias (for example, visual perception bias), thereby prioritising some forms of knowledge over others in ways that implicitly devalues other forms of knowledge and insight that are not machine readable yet nonetheless valuable and important (Boyd & Crawford, 2012). Finally, the ready availability of generative AI may facilitate plagiarism and other forms of research misconduct and unethical scientific practices which may be difficult to identify, bringing scientific research into disrepute.

Resource inequality between public and private sector research

The funding landscape for AI research is characterised by significant resource disparities between academia and industry, particularly in more recent years (see Chapter 2). For example, a study by the EU's Joint Research Centre, TechWatch, estimated that 68% of AI investment in 2020 came from the private sector with 32% from the public sector, in which the former was growing at a faster rate. Industry is estimated to have spent $340 billion US dollars on AI in 2021, far outstripping the $1.5 billion and $1.2 billion allocated to AI respectively by the US non-defence government agencies and the European Commission in the same year (Ahmed et al, 2023). It is now industry rather than academia that is producing cutting-edge machine learning research and training large AI models. This inequality in resourcing between public and private sector research in AI is manifest in relation to all the key resources upon which AI development relies.

In relation to computing power, the number of model parameters is a key determinant of the computing power needed. In 2021, industry models were 29 times bigger, on average, than academic models, resulting

in a relative shortfall of computing available to academics (Ahmed et al, 2023). Ahmed cites data from Canada's National Advanced Research Computing Platform which indicates that demand on their platform for GPUs, the most common chips used in AI, increased 25-fold since 2013, but supply has only been able to meet 20% of this demand in recent years. Meanwhile, competition for AI talent has intensified. Data from North American universities indicates that while only 21% of AI PhD graduates went to industry in 2004, by 2020 almost 70% headed to industry following graduation. A similar trend can be observed in the rate at which computer science research faculty are being employed and retained (Ahmed et al, 2023).

The resulting inequality is a matter of scientific and public concern for several reasons. First, private firms do not typically make public the underlying code and the underlying datasets for their AI models, thereby precluding public scientists from evaluating their validity, robustness and vulnerabilities (Ahmed et al, 2023). This, in turn, means that public researchers who deploy these privately produced AI models in their research cannot subject the tools themselves to independent review, and are therefore unable to evaluate their robustness, reliability, accuracy and, in turn, the epistemic validity of the outputs produced from their deployment. This deepens the scientific 'reproducibility crisis' referred to under The challenges. Secondly, if public scientists lack the resources to develop their own AI models, their capacity to offer public interest alternatives is thereby limited. In particular, as Ahmed and colleagues have argued, some useful capabilities of AI systems seem to be 'emergent', meaning that they only acquire these characteristics once these systems are especially large. This includes negative characteristics, such as toxicity in AI-generated language and stereotyping (Ahmed et al, 2023). If it is left to industry alone to produce cutting-edge AI models, this may result in the neglect of public interest research which may be commercially unprofitable, including addressing the needs of those from lower-income countries. Others express concern that because private sector control of state-of-the art AI models is driven by commercial imperatives, it has focused on substituting human labour rather than on seeking to augment human capabilities through human-machine collaboration which may be more conducive to human wellbeing (Ahmed et al, 2023). Thirdly, the Big Tech firms that now dominate the AI innovation frontier have developed a number of strategies through which they exploit the labour of academic researchers through collaborations to produce research insights, yet with whom they do not share the patent ownership to which that research has contributed (Rikap, 2023a). They also use open-source development platforms to test and improve their software, enabling them to take advantage of the work of the developer community provided on a voluntary basis and for which they are uncompensated. Through these and other strategies, the resulting corporate innovation systems enable Big Tech firms to maintain secrecy over knowledge inflows while minimising outflows to profit from AI scientific knowledge (Rikap, 2023a).

Adverse social impacts: threats to safety, security and democracy

Scientific misinformation and misconduct

Although the EU's commitment to open science directly reflects a basic principle of universalism that underpins the scientific endeavour (see Chapter 1), the general purpose nature of AI systems, combined with the openness and public accessibility of publicly funded AI research, has enabled them to be used for

malicious purposes. As a result, a variety of forms of information-based harms and wrongs have proliferated. In the scientific domain, this includes the production of 'fake science' in which generative AI applications are enabling the growth of 'paper mills' – companies that produce fake scientific papers yet may be difficult to detect (Liverpool, 2023). In addition, the pressure to publish scientific papers, together with lucrative business models in scientific publishing, have fostered the emergence of predatory journals which make use of the popular open access model – charging fees to authors, rather than to readers – to publish poor quality or even fake scientific papers. *The Economist* reports that, according to one firm that monitors and blacklists English language predatory journals, some 1000 such journals existed in 2010 while today there are at least 13 000. Together with the growing number of scientific publications and proliferation of commercially produced journals for profit, this is placing the scientific peer-review system under ever-increasing strain (Hanson et al, 2023). Yet our review of evidence indicated that the capacity to use AI tools to automate the process of peer-review has not been demonstrated. At the same time, several scientists participating in our Policy Design workshop emphasised the vital importance of entrusting scientific peer-review to human scientists, based on the nature of the scientific endeavour as communal practice entailing deliberation and dialogue between scientific peers and which cannot be legitimately delegated to machines (see 'Policy design workshop report').

Malicious use: Threats to biosecurity, cybersecurity and democratic freedom

We have already noted that the general purpose character of contemporary AI systems means that they are a double-edged sword, capable of being used intentionally for both noble and malicious ends. This is especially vivid in life sciences, where AI has facilitated the development of new biological materials and engineering new living systems and organisms. Although these offer many very important benefits in the form of new vaccines, biotherapeutics, and carbon-capturing microbes, the use of AI could also accidentally or deliberately cause significant harm. These capabilities might also be harnessed by malicious actors, making the threat of biological catastrophe a possibility and for which appropriate safeguards are needed (Carter et al, 2023).

Similarly, we have also noted how generative AI tools such as ChatGPT may extend the range of cybersecurity threats, for example, by introducing malware that may be invulnerable to conventional cybersecurity protections, while enabling those without specialist expertise to create malware. Another major threat, both to security and democratic freedom and respect for fundamental rights, arises from the use of AI for manipulative, fraudulent or other malicious purposes, through image, text, and audio content generation. The use of AI for political microtargeting and misinformation is now considered to pose a significant and on-going threat to free and fair elections and thus to democratic freedom. However, more recent advances in machine vision techniques have enabled the ready availability of AI-generated deepfakes. Not only does this magnify the possibility of manipulative and fraudulent applications, but the proliferation of misogyny in the form of deepfake pornography is particularly troubling, serving to dehumanise women and girls who are rendered practically powerless to defend themselves or prevent these forms of abuse and bullying causing particularly invidious harms yet produced at scale. Although AI researchers are working towards various technological approaches to detect fake AI-created images and

audio, they have yet to demonstrate their capabilities to provide scalable, real-world protection against these kinds of attacks.

Broader social, collective and existential harms

Researchers have drawn attention to a variety of societal harms that the acceleration of AI uptake and its continued advances may generate. In particular, they have drawn attention to the very high carbon footprint produced by training state-of-the art foundation models, and the global political instability, insecurity and destructive potential of AI-enabled weapons development for military purposes. These serious social concerns highlight the need to attend more carefully to the social and moral responsibility of AI researchers, recognising that the powerful capabilities wrought by the development of ever-more powerful AI models may be used for good or for ill, and the need for more systematic and sustained attention to legitimate and effective governance of science in the service of humanity.

Policy design options

The preceding outline has outlined the nature and content of key challenges facing Europe that we have identified in seeking to accelerate the uptake of AI in science in a responsible and timely manner. We now offer a series of inter-related proposals which are intended to help address those challenges which we are worthy of consideration. The evidence review process for this report (see Annex 3) is based on a choice of methods of rapid literature review and expert workshops as an alternative to a systematic review of evaluation of the literature in order to adapt to the short timeframe of the request. We provide the following options for policy based on this process, which did not include a review of evaluation studies to provide guidance on their likely effectiveness and relative strengths and shortcomings of these proposals. Further critical research, examination, and discussion in conjunction with important stakeholders, including the scientific academies, AI researchers of various levels of seniority, researchers from disciplines who are already embracing the use of AI tools in their scientific research and with the editors of high-quality scientific journals would be necessary for a better grasp on the best way forward.

Research and development of best practices, guidelines and protocols

Lack of attention to established principles of scientific rigour in both AI research and research undertaken with the assistance of AI systems is at least in part due to the novelty of the tools, the research itself, and a lack of understanding and clarity about the specific norms applicable to the research, both in terms of scientific rigour and of legal and ethical obligations. This may be a product of lack of understanding and awareness of the norms themselves, but it is also, perhaps primarily, due to a lack of clarity about how those norms should be operationalised in specific, localised research settings. However, given the novelty of AI as a tool for knowledge discovery and task automation, we also identified evidence suggesting a lack of clear benchmarks, guidelines, protocols, and best practice conventions to which researchers can adhere. Accordingly, during our review, we identified a number of papers and initiatives in which researchers have

taken steps to address these deficiencies, particularly in relation to concerns which can be broadly understood as contributing to the 'reproducibility crisis'.

For example, several researchers have sought to identify methodological weaknesses in scientific papers, developing taxonomies to understand various forms of error, and positing a range of metrics, benchmarks, and frameworks that, if adopted, could be expected to help overcome these shortcomings in research practice. We saw less evidence indicating that there was a body of nascent scholarship produced by law and applied ethics researchers seeking to investigate the legal and ethical dimensions of AI research, or AI-enabled research, and to propose more concrete guidelines and principles to help address identified deficiencies in research practices. The emergence of LLMs that have novel and powerful automated text-generation capabilities also indicates a need to establish guidelines for the ethical use of LLMs in research that, among other things, address concerns related to data privacy, algorithmic fairness, replicability and the potential misuse of LLM-generated findings (Grossmann et al, 2023). Although, over time, the research community may develop appropriate norms, best practices and guidelines without sustained policy intervention, reliance on spontaneous, organic development by the research community is unlikely to occur in a timely nor systematic manner.

Accordingly, we suggest policy interventions to be considered that may help accelerate and foster the development of appropriate best practices, protocols, benchmarks, and guidelines for the use of AI in scientific research that promote and ensure that this is undertaken in a manner directed at securing epistemic integrity, validity, and open publication, with the aim of addressing concerns about the limited reproducibility, interpretability and transparency of research. They should also ensure that research undertaken with the assistance of AI tools and systems conforms with basic principles of research integrity, thereby reducing the risk of scientific misconduct.

Such guidance should also aim to ensure that AI is employed in scientific research in accordance with applicable legal rights and interests, particularly of those affected by the research (including the authors and owners of copyright-protected works), and that research is conducted in an ethically appropriate manner, following proper appraisal of its implications for individuals, groups, and society more generally. The need for diversity and inclusion, particularly of vulnerable and other under-represented groups, including those with disabilities and special needs, warrants special attention given the prevalence of algorithmic bias and unfair discrimination against historically marginalised groups.

The goal of formulating suitable guidance is aimed at ensuring epistemic integrity. This does not mean insisting on transparency in relation to every aspect of a research process or method: rather, it is to help identify which practices, and which forms of opacity are damaging to research and its role in society. In so doing, this should enable researchers to better develop protocols and systems for comprehensive quality assessment, adopting interdisciplinary collaboration and training, and seeking new machine learning classifications for known discriminators (Hutson, 2020; Mitchell, 2023; Rudin, 2019; Sejnowski, 2020). The formulation of such guidance can also help organise emerging practices from the AI field itself, such as Datasheets for Datasets (Gebru et al, 2021), data statements for natural language processing (Bender & Friedman, 2018), or advice on adopting software engineering practices for accountable AI systems

development (Hutchinson et al, 2021). Likewise, standards for high quality software can facilitate the development of trustworthy AI tools. Research software is a fundamental and vital part of research, yet significant challenges to discoverability, productivity, quality, reproducibility, and sustainability exist. The emergence of FAIR Principles for Research Software, for example, reflects a maturation of the research community, recognising that research software is a type of digital object to which FAIR should be applied (Barker et al, 2022). The need to establish a common framework of indicators and best practices for research software quality across domains is particularly important given the risks of poor quality but rapidly produced using generative AI.

At the same time, in formulating such guidance, interpretive flexibility is necessary in order to allow for the development of discipline-specific norms of appropriateness, even in how concepts such as 'replicability' are understood. To this end, there are several organisations and institutions throughout Europe that may have a valuable and significant role to play in light of their distinctive knowledge and high level of relevant expertise, including European research academies (such as those who participate in the SAPEA network), and the European Research Council. Given the evolving nature of the discipline and the pace at which it is developing, guidelines need to be sufficiently flexible to allow for advances in scientific discovery, while providing more useful, actionable guidance to researchers. For example, this may necessitate revision of research ethics guidelines, including guidance on what constitutes scientific misconduct.

Once developed, widespread publication and awareness raising activities are likely to be necessary to promote adherence to the guidelines by researchers across a variety of communities of scientific practice, including editors and publishers of scientific journals. Consideration may also be given to requiring adherence to any resulting guidelines as a condition of conditions of an award of EU research funding.

Researcher education and training

Suitable education, training and skills development for researchers is essential to address the lack of awareness and understanding among researchers of the scope, content and application of norms of scientific rigour, and of legal duties and ethical obligations. Our analysis of the evidence indicates the need to provide appropriate training to scientists from a wide range of disciplinary domains, prioritising researchers already engaged in AI research and those already using AI in their research. This may require the development of new education and training programmes, which include consideration of scientific rigour, adherence to legal and ethical norms, and the importance of sensitivity to domain-specific and discipline-specific norms of appropriateness. It may also require more investment in AI ethics research, to address concerns about the increasing capture of academic work in AI ethics that is sponsored by Big Tech firms, and to call for peer review editors to insist on the proper disclosures of funding sources, so that the ethical integrity of research is not undermined by potential conflicts of interest. To this end, we also suggest that in order to address existing gaps in the knowledge and skills of researchers, these training programmes should consider including a number of substantive issues including the capabilities and limits of AI, an understanding of the importance of data quality, cleaning, curation and provenance and its transparent and accountable handling, how the use of AI implications legal and ethical norms (including data governance,

privacy, copyright law and equality law), particularly concerning the collection, processing and sharing of data and the need for, and challenges associated with, cross-disciplinary research.

Academic publishing

Scientific publishing by commercial publishers is primarily driven by commercial imperatives rather than a commitment to scientific rigour and integrity. In recent years, the number of scientific journals produced by commercial publishers has significantly expanded (Hanson et al, 2023). There is some evidence to suggest that this has facilitated the publication of increasing volumes of low-quality papers, and that the emergence of generative AI tools has enabled the growth of paper mills and other forms of scientific fraud. A comparative analysis of the scientific peer review process of the most prestigious science journals, and those which are much less so, indicates that the most prestigious journals devote considerably more time and human resources to the process. This reminds us that high quality science takes time, yet the pressure on researchers to publish may create incentives for low quality research, enabled by the proliferation of poor-quality journals.

At present, relatively little rigorous and systematic information is available about the practices and quality of scientific journals. Although the publishing industry has developed 'journal impact factors', which are intended to provide a measure of the importance of a journal, they are based on crude quantitative calculations involving automatically counting the number of times selected articles are cited within a particular year. To help counter the publication of fraudulent papers and low quality scientific papers that AI tools may otherwise accelerate, consideration could be given to developing publicly funded, transparent guidelines and metrics, informed by principles of scientific rigour and integrity, which might also regularly publish an analysis and ranking of scientific journal quality that provides a more thorough, rigorous, informed and transparent indication of the relative ranking of scientific journals in terms of their scientific rigour and integrity. While it is beyond the remit of this working group to identify whether any specific institution or organisation might be best entrusted with this responsibility, it may be worth considering the role that the European academy networks could play. The functions of this organisation could also include systematically monitoring the scientific publishing field to identify predatory journals and fraudulent papers.

Coordinated EU effort: a state-of-the-art AI research facility

One of the most pressing and important challenges concerns the inability of public AI researchers to access computational resources and high quality datasets to undertake cutting-edge AI research. This is now overwhelmingly dominated by private scientific laboratories hosted by Big Tech, who are not required to pursue research that aligns with the principles of the scientific endeavour, particularly those of open communication and common heritage. The resulting inequality substantially limits the ability of public AI researchers to offer public interest alternatives, to test and evaluate the AI models produced by Big Tech (because they do not have access to the underlying code or source data), or to undertake valuable research that serves the public interest and that would otherwise be neglected because it is not sufficiently lucrative to attract private sector interest. In addition, overdependence by researchers on commercial AI models may

exacerbate the reproducibility crisis, because the underlying models themselves are not open for public scrutiny and thus not have not been subjected to open peer-review evaluation to identify their validity and limits. This places public AI research in a position of comparative disadvantage.

Several countries have announced initiatives to increase the compute available for research and academia, including the US National AI Research Resource, Canada's Digital Research Infrastructure Strategy, and the Swiss AI Initiative. These are in addition to initiatives to take stock of compute capacity and needs, including those of researchers, such as the Canadian Digital Research Infrastructure Needs Assessment (Pérez-Jvostov et al, 2021) and the UK's 2022 Future of Compute review. These initiatives are in keeping with the OECD observation that, without increased access to high-performance computing and software to support the development of AI in science, less well-funded research groups are at a disadvantage because state-of-the-art computing resources are prohibitively expensive for many researchers (OCDE, 2023).

Accordingly, one possible way forward for the European Commission to consider is investing in, establishing, and provide ongoing funding for a state-of-the art facility for academic research in Europe, that would provide the level of resource needed to enable public resources to engage in cutting-edge AI research while making these facilities available to public scientists seeking to use AI for scientific research. This European AI super-centre would provide public scientists and researchers (that is, those employed by publicly-funded universities and research institutes that operate on a not-for-profit basis) access to infrastructure and inputs needed to undertake cutting-edge AI research. This facility would be comprised of the following:

- massive computational power
- sustainable cloud infrastructure
- repository of high quality, clean, responsibly collected and curated datasets
- an AI scientific advisory and skills unit engaged in developing best practice research standards for AI and developing and delivering appropriate training and skills development programmes to address existing lack of awareness and skills concerning the matters discussed under Best practices, guidelines and protocols and education.

The core mission of this proposed European AI super-centre would be a commitment to cutting-edge responsible AI research that has scientific integrity in the service of the public good. So understood, it is not intended to compete with or replicate AI research undertaken by Big Tech, which is driven by commercial imperatives reflected in its 'move fast and break things' mentality. If this domain of research is characterised as a race, then public science is already disadvantaged, in that it is bound by more rigorous norms of integrity and responsibility to public values which are not binding on commercial, for-profit institutions. At the same time, there are systematic risks associated with increasing dependence on Big Tech and their appropriation and exploitation of scientific knowledge, through leveraging the knowledge and labour of others who do not receive a proportionate share of the resulting profits.

It is also important to stress the transnational collaboration that this initiative should have, especially in the light of the advanced state of research on AI in non-EU countries such as the UK and Switzerland. Moreover, the role of this centre could also be to link to similar experiences, with the same founding values, from non-

European countries, strengthening the research connections on responsible AI in the academic community at a global scale.

In exploring this option, the European Commission needs to address 'brain drain', in which talented AI researchers are lured into private labs that offer state-of-the art computing resources, reducing the availability of skilled scientists engaged in public research. The European Commission has also identified brain drain and barriers to building academic careers as challenges in the development of AI research in the EU (European Commission, Petkova, & Roman, 2023), suggesting that the right talent, such as research engineers, needs to be attracted and maintained, and that existing researchers need to be trained in AI to successfully adopt AI in science. The Commission report also points out that EU universities and other research institutions struggle to compete with the job offers and research opportunities in the private sector. Introducing more career incentives for researchers to motivate them to choose the academic career path over the private sector is suggested as a way to strengthen the EU's visibility and potential in AI research (European Commission, Petkova, & Roman, 2023). Accordingly, attractive, sustainable alternative career pathways for talented AI early career researchers are an urgent priority. A European AI super-centre could make opportunities and incentives available, creating career pathways to attract talented early-career researchers engaged in AI research, seeking to cultivate an open, autonomy-respecting research environment that allows them to engage more freely in curiosity-driven, ambitious research projects motivated by intellectual inquiry for the public good, protected from the demands and pressures of teaching and short-term publication.

In addition, to support public interest AI research that would otherwise be neglected by commercial labs, we suggest that the Commission consider targeted funding research programmes to meet the needs of low-income countries, perhaps involving collaborations with public universities from low-income countries.

In evaluating this option, we are mindful that there are a number of significant challenges involved in setting up such a super-centre. In particular, there have been a number of ambitious European science initiatives, but these have had mixed success. Accordingly, we suggest that the Commission may wish to engage in further research and analysis to help identify, from past experience, under what conditions are these initiatives more likely to succeed. For example, we recognise that there is a tension between the value of a single centralised facility and a network of more localised facilities, including the needs and interests of member states and their communities in relation to its geographical location.

AI for social and environmental protection

As the power and capabilities of AI models have advanced, it has become even more important to seek to understand these models and prevent them from being used for malicious purposes, or in ways that generate unintended social, group and individual harm. For this reason, a number of countries, including the US and the UK, are now establishing AI safety institutes to undertake sustained research into the range of risks to safety and security. While it is already evident that these systemic and social risks may arise in relation to human health and safety (due, for example, to threats concerning AI-enabled bioweapon development), as the power of AI models has grown, so too have the risks to democracy (including respect

for the fundamental rights and dignity of persons) and to sustainable development. As already indicated, the training of LLMs consume vast amounts of energy, and much more needs to be done to address their adverse environmental impacts. This could include systematically gathering insight on progress towards standardised metrics for carbon measurement, identifying and evaluating the most promising paths for development and adoption, and advocating for mandatory disclosure (in addition to disclosure requirements for general purpose AIs introduced by the AI Act), possibly in collaboration with EU AI Office initiatives to assess and minimise the impact of AI systems on environmental sustainability.

Accordingly, it may be valuable in Europe to consider creating an EU institute with regular access to skilled expert advisors and suitably competent and trained staff to undertake research, engage in routine monitoring and foresight to identify and to collate information concerning emerging risks produced by the use of AI models that may threaten safety, security, democracy and sustainable development. The Commission may wish to consider the relationship of such an institute with the EU Office for AI that will be established under the AI Act. The institute's functions should include duties to:

- proactively monitor potential vulnerabilities and misuse for AI in ways that pose societal risks, particularly to safety, security democracy and sustainability
- provide the European Office for AI with regular systematic reports of identified vulnerabilities and emerging risks
- engage in information exchange and collaboration with other similar public institutes in Europe and around the world, proactively monitoring and providing periodic reports and making recommendations to address these threats
- formulate concrete policy proposals that will help to mitigate AI-generated threats, including threats to sustainable development (including but not limited to those suggested above) and to biosecurity (for example, by recommending restrictions on publication and transparency requirements that would otherwise conventionally apply to AI research, while restricting access to trusted researchers due to potential for abuse and misuse)
- support the Commission and EU member states in seeking agreement at the international level to put in place binding legal limits on use of AI for military purposes, including warfare

Similarly to a Europe-wide AI research facility, international coordination of these activities would be essential. The recent UN interim report on *Governing AI for Humanity* proposes the guiding principles to foster such collaboration – inclusiveness, public interest, international governance of AI, and so on – and it also identifies specific institutional functions to implement the above principles (United Nations, 2023).

Conclusion

The policy options outlined above, which we offer for consideration, are intended to help address the challenges that this report has identified in seeking to accelerate the uptake of AI in science in a timely and responsible manner. We believe that more evidence and investigation of these proposals is required in order to provide a more informed appraisal. They may also serve to address larger concerns about the

concentration of power in the AI sector by a handful of private global tech giants (Verdegem, 2022). Nevertheless, our hope is that they offer valuable starting points.

In conclusion, we acknowledge that in establishing policy priorities in the field of AI research, there are often tensions among different principles and values. So, for example, the goal of openness in science that would require full transparency in the publication of AI research may be in tension with the need to mitigate threats and risks associated with the malicious use of that research. While resolving these conflicts and tensions inevitably require the making of normative trade-offs, democratic political communities should aspire to making these trade-offs intentionally, openly and in a manner that is respectful of fundamental rights and in consultation with affected publics.

References

Abdelnabi, S., & Fritz, M. (2020). Adversarial Watermarking Transformer: Towards Tracing Text Provenance with Data Hiding. arXiv:2009.03015. Retrieved September 01, 2020, from https://ui.adsabs.harvard.edu/abs/2020arXiv200903015A

Ada Lovelace Institute. (2023a). Inclusive AI governance: civil society participation in standards development. https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/

Ada Lovelace Institute. (2023b). Safe before sale: Learnings from the FDA's model of life sciences oversight for foundation models. https://www.adalovelaceinstitute.org/report/safe-before-sale/

Ahmed, N., Wahed, M., & Thompson, N. C. (2023). The growing influence of industry in AI research. Science, 379(6635), 884–886. https://doi.org/10.1126/science.ade2420

Ahn, M., Brohan, A., Brown, N., Chebotar, Y., Cortes, O., David, B., Finn, C., Fu, C., Gopalakrishnan, K., Hausman, K., Herzog, A., Ho, D., Hsu, J., Ibarz, J., Ichter, B., Irpan, A., Jang, E., Ruano, R. J., Jeffrey, K., … Zeng, A. (2022). Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. arXiv preprint arXiv:2204.01691. https://doi.org/10.48550/arXiv.2204.01691

Alharbi, W. S., & Rashid, M. (2022). A review of deep learning applications in human genomics using next-generation sequencing data. Human Genomics, 16(1), 26. https://doi.org/10.1186/s40246–022–00396-x

ALLEA. (2023). The European Code of Conduct for Research Integrity – Revised Edition 2023. https://allea.org/portfolio-item/european-code-of-conduct-2023/

ANEC. (2021). The role of standards in meeting consumer needs and expectations of AI in the European Commission proposal for an Artificial Intelligence Act (Position Paper ANEC-DIGITAL-2021-G-141, Issue. https://www.anec.eu/images/Publications/position-papers/Digital/ANEC-DIGITAL-2021-G-141.pdf

Ares, N. (2021). Machine learning as an enabler of qubit scalability. Nature Reviews Materials, 6(10), 870–871. https://doi.org/10.1038/s41578–021–00321-z

Assael, Y., Sommerschield, T., Shillingford, B., Bordbar, M., Pavlopoulos, J., Chatzipanagiotou, M., Androutsopoulos, I., Prag, J., & de Freitas, N. (2022). Restoring and attributing ancient texts using deep neural networks. Nature, 603(7900), 280–283. https://doi.org/10.1038/s41586–022–04448-z

Babl, F. E., & Babl, M. P. (2023). Generative artificial intelligence: Can ChatGPT write a quality abstract? Emergency Medicine Australasia, 35(5), 809–811. https://doi.org/10.1111/1742–6723.14233

Ball, P. (2023). Is AI leading to a reproducibility crisis in science? Nature, 624, 22–25. https://doi.org/10.1038/d41586–023–03817–6

Bansal, R., Samanta, B., Dalmia, S., Gupta, N., Vashishth, S., Ganapathy, S., Bapna, A., Jain, P., & Talukdar, P. (2024). LLM Augmented LLMs: Expanding Capabilities through Composition. arXiv preprint arXiv:2401.02412. https://doi.org/10.48550/arXiv.2401.02412

Barker, M., Chue Hong, N. P., Katz, D. S., Lamprecht, A.-L., Martinez-Ortiz, C., Psomopoulos, F., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., & Honeyman, T. (2022). Introducing the FAIR Principles for research software. Scientific Data, 9(1), 622. https://doi.org/10.1038/s41597–022–01710-x

Begou, N., Vinoy, J., Duda, A., & Korczynski, M. (2023). Exploring the Dark Side of AI: Advanced Phishing Attack Design and Deployment Using ChatGPT. arXiv:2309.10463. Retrieved September 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230910463B

Bender, E. M., & Friedman, B. (2018). Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. Transactions of the Association for Computational Linguistics, 6, 587–604. https://doi.org/10.1162/tacl_a_00041

References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? ? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada. https://doi.org/10.1145/3442188.3445922

Benjamin, R. (2019). Race After Technology: Abolitionist Tools for the New Jim Code. https://www.wiley.com/en-us/Race+After+Technology:+Abolitionist+Tools+for+the+New+Jim+Code-p-9781509526437

Bernstein, S., Diamond, R., Jiranaphawiboon, A., McQuade, T., & Pousada, B. (2022). The Contribution of High-Skilled Immigrants to Innovation in the United States. National Bureau of Economic Research Working Paper 30797. https://doi.org/10.3386/w30797

BEUC. (2022). Regulating AI to Protect the Consumer – Position Paper on the AI Act. https://www.beuc.eu/sites/default/files/publications/beuc-x-2021–088_regulating_ai_to_protect_the_consumer.pdf

Bied, G., Solal, N., Perennes, E., Hoffmann, M., Caillou, P., Crépon, B., Gaillac, C., & Sebag, M. (2023). Toward Job Recommendation for All. IJCAI 2023, 5906–5914. https://doi.org/10.24963/ijcai.2023/655

Bircan, T., & Salah, A. A. A. (2022). A Bibliometric Analysis of the Use of Artificial Intelligence Technologies for Social Sciences. Mathematics, 10(23), 4398. https://www.mdpi.com/2227–7390/10/23/4398

Bird, S. J. (2014). Socially Responsible Science Is More than "Good Science". Journal of Microbiology & Biology Education, 15(2), 169–172. https://doi.org/doi:10.1128/jmbe.v15i2.870

Bitkina, O. V., Park, J., & Kim, H. K. (2023). Application of artificial intelligence in medical technologies: A systematic review of main trends. Digital Health, 9. https://doi.org/doi:10.1177/20552076231189331

Boiko, D. A., MacKnight, R., Kline, B., & Gomes, G. (2023). Autonomous chemical research with large language models. Nature, 624(7992), 570–578. https://doi.org/10.1038/s41586–023–06792–0

Boyd, D., & Crawford, K. (2012). Critical questions for Big Data. Information, Communication & Society, 15(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Braunerhjelm, P. (2010). Entrepreneurship, Innovation and Economic Growth-past experience, current knowledge and policy implications. https://www.diva-portal.org/smash/record.jsf?dswid=-855&pid=diva2%3A484894

Brown, R. (2023). The AI generation: how universities can prepare students for the changing world. Demos. https://demos.co.uk/research/the-ai-generation-how-universities-can-prepare-students-for-the-changing-world/

Brownsword, R., Scotford, E., & Yeung, K. E. (2017). The Oxford Handbook of Law, Regulation and Technology (E. S. e. Roger Brownsword (ed.), Karen Yeung (ed.), Ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199680832.001.0001

Buchanan, B., & Imbrie, A. (2024). The New Fire: War, Peace, and Democracy in the Age of AI. The MIT Press. https://mitpress.mit.edu/9780262548489/the-new-fire/

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research. https://proceedings.mlr.press/v81/buolamwini18a.html

Burger, B., Kanbach, D., 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

Calafiura, P., Rousseau, D., & Terao, K. (2022). Artificial Intelligence for High Energy Physics. https://doi.org/10.1142/12200

Carlsson, B., Acs, Z. J., Audretsch, D. B., & Braunerhjelm, P. (2009). Knowledge creation, entrepreneurship, and economic growth: a historical review. Industrial and Corporate Change, 18(6), 1193–1229. https://doi.org/10.1093/icc/dtp043

Carter, S. R., Wheeler, N. E., Chwalek, S., Isaac, C. R., & Jaime Yassif. (2023). The Convergence of Artificial Intelligence and the Life Sciences: Safeguarding Technology, Rethinking Governance, and Preventing Catastrophe. https://www.nti.org/analysis/articles/the-convergence-of-artificial-intelligence-and-the-life-sciences/

References

Casper, S., Davies, X., Shi, C., Gilbert, T. K., Scheurer, J., Rando, J., Freedman, R., Korbak, T., Lindner, D., Freire, P., Wang, T., Marks, S., Segerie, C.-R., Carroll, M., Peng, A., Christoffersen, P., Damani, M., Slocum, S., Anwar, U., … Hadfield-Menell, D. (2023). Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback. https://doi.org/10.48550/arXiv.2307.15217

CCA. (2022). Leaps and Boundaries (The Expert Panel on Artificial Intelligence for Science and Engineering, Council of Canadian Academies., Issue. https://www.cca-reports.ca/reports/ai-for-science-and-engineering/

CEDEFOP. (2023). Going digital means skilling for digital: Using big data to track emerging digital skill needs. Policy brief. Publications Office of the European Union. https://doi.org/10.2801/772175

CEP. (2021). Paving the digital path in Central and Eastern Europe: Regional perspectives on advancing digital transformation and cooperation. https://www.cep.si/wp-content/uploads/2021/11/Paving-the-Digital-Path-in-CEE-publication.pdf

Checco, A., Bracciale, L., Loreti, P., Pinfield, S., & Bianchi, G. (2021). AI-assisted peer review. Humanities and Social Sciences Communications, 8(1), 1–11. https://doi.org/10.1057/s41599–020–00703–8

Chen, C., & Shu, K. (2023). Combating Misinformation in the Age of LLMs: Opportunities and Challenges. arXiv:2311.05656. Retrieved November 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv231105656C

Chiang, T. (2000). Catching crumbs from the table. Nature, 405(6786), 517–517. https://doi.org/10.1038/35014679

Chin, Z.-Y., Jiang, C.-M., Huang, C.-C., Chen, P.-Y., & Chiu, W.-C. (2023). Prompting4Debugging: Red-Teaming Text-to-Image Diffusion Models by Finding Problematic Prompts. arXiv:2309.06135. Retrieved September 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230906135C

Choudhary, K., DeCost, B., Chen, C., Jain, A., Tavazza, F., Cohn, R., Park, C. W., Choudhary, A., Agrawal, A., Billinge, S. J. L., Holm, E., Ong, S. P., & Wolverton, C. (2022). Recent advances and applications of deep learning methods in materials science. npj Computational Materials, 8(1), 59. https://doi.org/10.1038/s41524–022–00734–6

Chubb, J., Cowling, P., & Reed, D. (2022). Speeding up to keep up: exploring the use of AI in the research process. AI & SOCIETY, 37(4), 1439–1457. https://doi.org/10.1007/s00146–021–01259–0

Chui, M. (2023). The state of AI in 2023: Generative AI's breakout year. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year#/

Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., Adi, Y., & Défossez, A. (2023). Simple and Controllable Music Generation. arXiv:2306.05284. Retrieved June 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230605284C

Cranmer, M., Sanchez Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., & Ho, S. (2020). Discovering Symbolic Models from Deep Learning with Inductive Biases Advances in Neural Information Processing Systems 33 (NeurIPS 2020), https://proceedings.neurips.cc/paper/2020/hash/c9f2f917078bd2db12f23c3b413d9cba-Abstract.html

Cuoco, E., Powell, J., Cavaglià, M., Ackley, K., Bejger, M., Chatterjee, C., Coughlin, M., Coughlin, S., Easter, P., Essick, R., Gabbard, H., Gebhard, T., Ghosh, S., Haegel, L., Iess, A., Keitel, D., Márka, Z., Márka, S., Morawski, F., … Williams, D. (2021). Enhancing gravitational-wave science with machine learning. Machine Learning: Science and Technology, 2(1), 011002. https://doi.org/10.1088/2632–2153/abb93a

Dalgaard, M., Motzoi, F., Sørensen, J. J., & Sherson, J. (2020). Global optimization of quantum dynamics with AlphaZero deep exploration. npj Quantum Information, 6(1), 6. https://doi.org/10.1038/s41534–019–0241–0

Dalla Benetta, A., Sobolewski, M., & Nepelski, D. (2021). AI Watch: 2020 EU AI investments. EUR 30826 EN, Publications Office of the European Union, Luxembourg, JRC126477. https://doi.org/doi:10.2760/017514

Daston, L., & Galison, P. (2007). Objectivity. Cambridge, Mass.: Zone Books. https://philpapers.org/rec/DASO-2

de Melo-Martín, I., & Intemann, K. (2023). Socially responsible science: Exploring the complexities. European Journal for Philosophy of Science, 13(3), 33. https://doi.org/10.1007/s13194–023–00537–6

References

de Oliveira, R. C., & de Souza e Silva, R. D. (2023). Artificial Intelligence in Agriculture: Benefits, Challenges, and Trends. Applied Sciences, 13(13), 7405. https://www.mdpi.com/2076–3417/13/13/7405

Degrave, J., Felici, F., Buchli, J., Neunert, M., Tracey, B., Carpanese, F., Ewalds, T., Hafner, R., Abdolmaleki, A., de las Casas, D., Donner, C., Fritz, L., Galperti, C., Huber, A., Keeling, J., Tsimpoukelli, M., Kay, J., Merle, A., Moret, J.-M., … Riedmiller, M. (2022). Magnetic control of tokamak plasmas through deep reinforcement learning. Nature, 602(7897), 414–419. https://doi.org/10.1038/s41586–021–04301–9

del Rio-Chanona, M., Laurentsyeva, N., & Wachs, J. (2023). Are Large Language Models a Threat to Digital Public Goods? Evidence from Activity on Stack Overflow. arXiv:2307.07367. https://doi.org/10.48550/arXiv.2307.07367

Deng, G., Liu, Y., Mayoral-Vilches, V., Liu, P., Li, Y., Xu, Y., Zhang, T., Liu, Y., Pinzger, M., & Rass, S. (2023). PentestGPT: An LLM-empowered Automatic Penetration Testing Tool. arXiv:2308.06782. Retrieved August 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230806782D

Dignum, V. (2020). AI is multidisciplinary. AI Matters, 5(4), 18–21. https://doi.org/10.1145/3375637.3375644

Donovan, M. (2023). How AI is helping historians better understand our past. MIT Technology Review. https://www.technologyreview.com/2023/04/11/1071104/ai-helping-historians-analyze-past/

Drew, L. (2023). The rise of brain-reading technology: what you need to know. Nature, 623(7986), 241–243. https://doi.org/10.1038/d41586–023–03423–6

Duan, C., Nandy, A., & Kulik, H. J. (2022). Machine Learning for the Discovery, Design, and Engineering of Materials. Annual Review of Chemical and Biomolecular Engineering, 13(1), 405–429. https://doi.org/10.1146/annurev-chembioeng-092320–120230

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT, 71, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Dybul, M., Basu, P., Cameron, E., Cicero, A., Diggans, J., Esvelt, K., Feruglio, S. L., Flyangolts, K., Inglesby, T., Lipsitch, M., Maruta, T., Mattison, J., Nelson, C., Palmer, M., Qureshi, C., Relman, D., Simmonds-Isler, J., Weber, A., Yassif, J. M., & Ofili, D. (2023). Biosecurity in the Age of AI. In. Available: https://www.helenabiosecurity.org/.

Ekpenyong, A. (2021, 2021). Digital Humanities Scholarship: A Model for Reimagining Knowledge Work in the 21st Century. Diversity, Divergence, Dialogue, Cham, Switzerland.

Elali, F. R., & Rachid, L. N. (2023). AI-generated research paper fabrication and plagiarism in the scientific community. Patterns, 4(3). https://doi.org/10.1016/j.patter.2023.100706

Elrod, P. D., & Tippett, D. D. (2002). The "death valley" of change. Journal of Organizational Change Management, 15(3), 273–291. https://doi.org/10.1108/09534810210429309

EMPL. (2022). Digitalisation and changes in the world of work: Literature review. https://data.europa.eu/doi/10.2861/291260

EPRS. (2022). Digital transformation: Cost of non-Europe. https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)699475

ESIR. (2023). Industry 5.0 and the future of work: Making Europe the centre of gravity for future good-quality jobs. Publications Office of the European Union. https://doi.org/10.2777/685878

Esplugas, M. (2023). The use of artificial intelligence (AI) to enhance academic communication, education and research: a balanced approach. JOURNAL OF HAND SURGERY-EUROPEAN VOLUME. https://doi.org/10.1177/17531934231185746

European Commission, D.-G. f. R. I., Arranz, D., Bianchini, S., Di Girolamo, V., & Ravet, J. (2023). Trends in the use of AI in science – A bibliometric analysis. Publications Office of the European Union. https://doi.org/10.2777/418191

References

European Commission, D.-G. f. R. I., Mendez, E., & Lawrence, R. (2020). Progress on open science: towards a shared research knowledge system : final report of the open science policy platform. Publications Office. https://data.europa.eu/doi/10.2777/00139

European Commission, D.-G. f. R. I., Petkova, D., & Roman, L. (2023). AI in science – Harnessing the power of AI to accelerate discovery and foster innovation – Policy brief. Publications Office of the European Union. https://doi.org/10.2777/401605

European Commission, D. C., Hartmann, C., Allan, J., Hugenholtz, P., Quintais, J., & Gervais, D. (2020). Trends and developments in artificial intelligence – Challenges to the intellectual property rights framework – Final report. Publications Office of the European Union. https://data.europa.eu/doi/10.2759/683128

European Commission, E. R. C. E. A. (2023). Use and impact of artificial intelligence in the scientific process – Foresight. Publications Office of the European Union. https://doi.org/10.2828/10694

Fabbrizzi, S., Papadopoulos, S., Ntoutsi, E., & Kompatsiaris, I. (2022). A survey on bias in visual datasets. Computer Vision and Image Understanding, 223, 103552. https://doi.org/10.1016/j.cviu.2022.103552

Ferrara, E. (2023). Social bot detection in the age of ChatGPT: Challenges and opportunities. First Monday, 28(6). https://doi.org/10.5210/fm.v28i6.13185

Filippi, E., Bannò, M., & Trento, S. (2023). Automation technologies and their impact on employment: A review, synthesis and future research agenda. TECHNOLOGICAL FORECASTING AND SOCIAL CHANGE, 191. https://doi.org/10.1016/j.techfore.2023.122448

Firat, M., & Kuleli, S. (2023). What if GPT4 Became Autonomous: The Auto-GPT Project and Use Cases. Journal of Emerging Computer Technologies, 3(1), 1–6. https://doi.org/10.57020/ject.1297961

Flanagin, A., Bibbins-Domingo, K., Berkwits, M., & Christiansen, S. L. (2023). Nonhuman "Authors" and Implications for the Integrity of Scientific Publication and Medical Knowledge. JAMA, 329(8), 637–639. https://doi.org/10.1001/jama.2023.1344

Frank, J., Herbert, F., Ricker, J., Schönherr, L., Eisenhofer, T., Fischer, A., Dürmuth, M., & Holz, T. (2023). A Representative Study on Human Detection of Artificially Generated Media Across Countries. arXiv:2312.05976. Retrieved December 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv231205976F

Gallegos, I. O., Rossi, R. A., Barrow, J., Mehrab Tanjim, M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2023). Bias and Fairness in Large Language Models: A Survey. arXiv:2309.00770. Retrieved September 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230900770G

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. npj Digital Medicine, 6(1), 75. https://doi.org/10.1038/s41746–023–00819–6

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997. https://doi.org/10.48550/arXiv.2312.10997

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., III, H. D., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86–92. https://doi.org/10.1145/3458723

Gefen, A., Saint-Raymond, L., & Venturini, T. (2020). AI for Digital Humanities and Computational Social Sciences. In B. Bertrand & G. Malik (Eds.), Reflections on AI for Humanity. https://hal.science/hal-03043393

Gemini Team, Anil, R., Borgeaud, S., Wu, Y., Alayrac, J.-B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A. M., Hauth, A., Millican, K., Silver, D., Petrov, S., Johnson, M., Antonoglou, I., Schrittwieser, J., Glaese, A., Chen, J., Pitler, E., … Vinyals, O. (2023). Gemini: A Family of Highly Capable Multimodal Models. arXiv:2312.11805. Retrieved December 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv231211805G

Gill, S. S., Xu, M., Ottaviani, C., Patros, P., Bahsoon, R., Shaghaghi, A., Golec, M., Stankovski, V., Wu, H., Abraham, A., Singh, M., Mehta, H., Ghosh, S. K., Baker, T., Parlikad, A. K., Lutfiyya, H., Kanhere, S. S., Sakellariou, R., Dustdar, S., … Uhlig, S. (2022). AI for next generation computing: Emerging trends and future directions. Internet of Things, 19, 100514. https://doi.org/10.1016/j.iot.2022.100514

References

Girdhar, R., Singh, M., Brown, A., Duval, Q., Azadi, S., Saketh Rambhatla, S., Shah, A., Yin, X., Parikh, D., & Misra, I. (2023). Emu Video: Factorizing Text-to-Video Generation by Explicit Image Conditioning. arXiv:2311.10709. Retrieved November 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv231110709G

Gitelman, L. e. (2013). 'Raw Data' Is an Oxymoron. The MIT Press. https://doi.org/10.7551/mitpress/9302.001.0001

Glausiusz, J. (2019). Tenure denial, and how early-career researchers can survive it. Nature, 565(7740), 525–527. https://doi.org/10.1038/d41586–019–00219–5

Goggin, G., & Soldatić, K. (2022). Automated decision-making, digital inclusion and intersectional disabilities. New Media & Society, 24(2), 384–400. https://doi.org/10.1177/14614448211063173

Gomez-Herrera, E., & Koeszegi, S. (2022). A gender perspective on artificial intelligence and jobs: The vicious cycle of digital inequality. Working Paper 15/2022.

Grossmann, I., Feinberg, M., Parker, D. C., Christakis, N. A., Tetlock, P. E., & Cunningham, W. A. (2023). AI and the transformation of social science research. Science, 380(6650), 1108–1109. https://doi.org/10.1126/science.adi1778

Gundersen, O. E., Gil, Y., & Aha, D. W. (2018). On Reproducible AI: Towards Reproducible Research, Open Science, and Digital Scholarship in AI Publications. AI Magazine, 39(3), 56–68. https://doi.org/10.1609/aimag.v39i3.2816

Guthrie, S., Lichten, C. A., van Belle, J., Ball, S., Knack, A., & Hofman, J. (2017). Understanding mental health in the research environment: A Rapid Evidence Assessment. RAND Corporation. https://doi.org/10.7249/RR2022

Hajkowicz, S., Naughtin, C., Sanderson, C., Schleiger, E., Karimi, S., Bratanova, A., & Bednarz, T. (2022). Artificial intelligence for science – Adoption trends and future development pathways. CSIRO Data61, Brisbane, Australia. https://www.csiro.au/-/media/D61/AI4Science-report/AI-for-Science-report-2022.pdf

Hanson, M. A., Gómez Barreiro, P., Crosetto, P., & Brockington, D. (2023). The strain on scientific publishing. arXiv:2309.15884. Retrieved September 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230915884H

Hassoun, S., Jefferson, F., Shi, X. H., Stucky, B., Wang, J., & Rosa, E. (2022). Artificial Intelligence for Biology. INTEGRATIVE AND COMPARATIVE BIOLOGY, 61(6), 2267–2275. https://doi.org/10.1093/icb/icab188

He, J., & Vechev, M. (2023). Large Language Models for Code: Security Hardening and Adversarial Testing. arXiv:2302.05319. Retrieved February 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230205319H

Heaven, W. (2020). AI is wrestling with a replication crisis. MIT Technology Review. https://www.technologyreview.com/2020/11/12/1011944/artificial-intelligence-replication-crisis-science-big-tech-google-deepmind-facebook-openai/

Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., & Meger, D. (2018). Deep reinforcement learning that matters Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, New Orleans, Louisiana, USA.

Henry, S., & McInnes, B. T. (2017). Literature Based Discovery: Models, methods, and trends. Journal of Biomedical Informatics, 74, 20–32. https://doi.org/10.1016/j.jbi.2017.08.011

Ho, J., Chan, W., Saharia, C., Whang, J., Gao, R., Gritsenko, A., Kingma, D. P., Poole, B., Norouzi, M., Fleet, D. J., & Salimans, T. (2022). Imagen Video: High Definition Video Generation with Diffusion Models. arXiv:2210.02303. Retrieved October 01, 2022, from https://ui.adsabs.harvard.edu/abs/2022arXiv221002303H

Hutchinson, B., Smart, A., Hanna, A., Denton, E., Greer, C., Kjartansson, O., Barnes, P., & Mitchell, M. (2021). Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada. https://doi.org/10.1145/3442188.3445918

References

Hutson, M. (2018a). Artificial intelligence faces reproducibility crisis. Science, 359(6377), 725–726. https://doi.org/10.1126/science.359.6377.725

Hutson, M. (2018b). Missing data hinder replication of artificial intelligence studies. Science. https://doi.org/10.1126/science.aat3298

Hutson, M. (2020). Core progress in AI has stalled in some fields. Science, 368(6494), 927–927. https://doi.org/10.1126/science.368.6494.927

Ignat, O., Jin, Z., Abzaliev, A., Biester, L., Castro, S., Deng, N., Gao, X., Gunal, A., He, J., Kazemi, A., Khalifa, M., Koh, N., Lee, A., Liu, S., Min, D. J., Mori, S., Nwatu, J., Perez-Rosas, V., Shen, S., … Mihalcea, R. (2023). A PhD Student's Perspective on Research in NLP in the Era of Very Large Language Models. arXiv:2305.12544. Retrieved May 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230512544I

Iten, R., Metger, T., Wilming, H., del Rio, L., & Renner, R. (2020). Discovering Physical Concepts with Neural Networks. Physical Review Letters, 124(1), 010508. https://doi.org/10.1103/PhysRevLett.124.010508

Jagannadharao, A., Beckage, N., Nafus, D., & Chamberlin, S. (2023). Timeshifting strategies for carbon-efficient long-running large language model training. Innovations in Systems and Software Engineering. https://doi.org/10.1007/s11334–023–00546-x

Jardim, P., Rose, C., Ames, H., Echavez, J., Van de Velde, S., & Muller, A. (2022). Automating risk of bias assessment in systematic reviews: a real-time mixed methods comparison of human researchers to a machine learning system. BMC MEDICAL RESEARCH METHODOLOGY, 22(1). https://doi.org/10.1186/s12874–022–01649-y

Jin, Z., Liu, J., Lyu, Z., Poff, S., Sachan, M., Mihalcea, R., Diab, M. T., & Schölkopf, B. (2024). Can Large Language Models Infer Causation from Correlation? Under review as a conference paper at ICLR 2024, Vienna, Austria. https://openreview.net/forum?id=vqIH0ObdqL&noteId=CyjFILUwms

Johnson, P., Laurell, C., Ots, M., & Sandström, C. (2022). Digital innovation and the effects of artificial intelligence on firms' research and development-Automation or augmentation, exploration or exploitation? TECHNOLOGICAL FORECASTING AND SOCIAL CHANGE, 179. https://doi.org/10.1016/j.techfore.2022.121636

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589. https://doi.org/10.1038/s41586–021–03819–2

Jurowetzki, R., Hain, D., Mateos-Garcia, J., & Stathoulopoulos, K. (2021). The Privatization of AI Research(-ers): Causes and Potential Consequences – From university-industry interaction to public research brain-drain? , arXiv:2102.01648. Retrieved February 01, 2021, from https://ui.adsabs.harvard.edu/abs/2021arXiv210201648J

Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. Nature Climate Change, 12(6), 518–527. https://doi.org/10.1038/s41558–022–01377–7

Kabashkin, I., Misnevs, B., & Puptsau, A. (2023). Transformation of the University in the Age of Artificial Intelligence: Models and Competences. TRANSPORT AND TELECOMMUNICATION JOURNAL, 24(3), 209–216. https://doi.org/10.2478/ttj-2023–0017

Kak, A., Myers, S., & Whittaker, M. (2023). Make no mistake—AI is owned by Big Tech. MIT Technology Review. https://www.technologyreview.com/2023/12/05/1084393/make-no-mistake-ai-is-owned-by-big-tech/

Kaplan, F., & di Lenardo, I. (2017). Big Data of the Past. Frontiers in Digital Humanities, 4. https://doi.org/10.3389/fdigh.2017.00012

Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. Patterns, 4(9). https://doi.org/10.1016/j.patter.2023.100804

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3(6), 422–440. https://doi.org/10.1038/s42254–021–00314–5

References

Karpas, E., Abend, O., Belinkov, Y., Lenz, B., Lieber, O., Ratner, N., Shoham, Y., Bata, H., Levine, Y., & Leyton-Brown, K. (2022). MRKL Systems: A modular, neuro-symbolic architecture that combines large language models, external knowledge sources and discrete reasoning. arXiv preprint arXiv:2205.00445. https://doi.org/10.48550/arXiv.2205.00445

Khan, J. (2021). European academic brain drain: A meta-synthesis. European Journal of Education, 56(2), 265–278. https://doi.org/10.1111/ejed.12449

Khurana, R. (2022). Artificial Intelligence as a Vehicle for Innovation: Literature Review and Bibliometric Study. Asia Pacific Journal of Information Systems, 32(4), 916–944. https://doi.org/10.14329/APJIS.2022.32.4.916

Kıcıman, E., Ness, R., Sharma, A., & Tan, C. (2023). Causal Reasoning and Large Language Models: Opening a New Frontier for Causality. arXiv:2305.00050. Retrieved April 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230500050K

Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., & Goldstein, T. (2023). A Watermark for Large Language Models. arXiv:2301.10226. Retrieved January 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230110226K

Kismihók, G. (2021). Mentális egészséghez köthető problémák a magyar kutatásban - ReMO workshop Zenodo. https://doi.org/10.5281/ZENODO.5554725

Kismihók, G., Cardells, F., Güner, P. B., Kersten, F., Björnmalm, M., Stroobants, K., Mol, S. T., Huber, F., Seipelt, J., Kretzschmar, W. W., Bajanca, F., Shawrav, M. M., Dahle, S., Carbajal, G. V., Harrison, S., Trusilewicz, L. N., Hnatkova, E., Cophignon, A., Keszler, Á., … Parada, F. (2019). Declaration on Sustainable Researcher Careers. Brussels: Marie Curie Alumni Association and European Council of Doctoral Candidates and Junior Researchers. https://doi.org/10.5281/ZENODO.3082244

Kong, S., Cheung, W., & Zhang, G. (2022). Evaluating artificial intelligence literacy courses for fostering conceptual learning, literacy and empowerment in university students: Refocusing to conceptual building. COMPUTERS IN HUMAN BEHAVIOR REPORTS, 7. https://doi.org/10.1016/j.chbr.2022.100223

Kong, S., Cheung, W., & Zhang, G. (2023). Evaluating an Artificial Intelligence Literacy Programme for Developing University Students? Conceptual Understanding, Literacy, Empowerment and Ethical Awareness. EDUCATIONAL TECHNOLOGY & SOCIETY, 26(1), 16–30. https://doi.org/10.30191/ETS.202301_26(1).0002

Kousha, K., & Thelwall, M. (2023). Artificial intelligence to support publishing and peer review: A summary and review. Learned Publishing, 37(1), 4–12. https://doi.org/10.1002/leap.1570

Krenn, M., Kottmann, J. S., Tischler, N., & Aspuru-Guzik, A. (2021). Conceptual Understanding through Efficient Automated Design of Quantum Optical Experiments. Physical Review X, 11(3), 031044. https://doi.org/10.1103/PhysRevX.11.031044

Krenn, M., & Zeilinger, A. (2020). Predicting research trends with semantic and neural networks with an application in quantum physics. Proceedings of the National Academy of Sciences, 117(4), 1910–1916. https://doi.org/10.1073/pnas.1914370116

Kreuk, F., Synnaeve, G., Polyak, A., Singer, U., Défossez, A., Copet, J., Parikh, D., Taigman, Y., & Adi, Y. (2022). AudioGen: Textually Guided Audio Generation. arXiv:2209.15352. Retrieved September 01, 2022, from https://ui.adsabs.harvard.edu/abs/2022arXiv220915352K

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. https://doi.org/10.1145/3065386

Kumar, P., Chauhan, S., & Awasthi, L. K. (2023). Artificial Intelligence in Healthcare: Review, Ethics, Trust Challenges & Future Research Directions. Engineering Applications of Artificial Intelligence, 120, 105894. https://doi.org/10.1016/j.engappai.2023.105894

Lane, M., & Saint-Martin, A. (2021). The impact of Artificial Intelligence on the labour market: What do we know so far? OECD Social, Employment and Migration Working Papers No. 256. OECD Publishing. https://doi.org/10.1787/7c895724-en

Lazard, G. A. G. (2023). Geopolitics of Artificial Intelligence. https://www.lazard.com/research-insights/the-geopolitics-of-artificial-intelligence/

References

Lazzeretti, L., Innocenti, N., Nannelli, M., & Oliva, S. (2023). The emergence of artificial intelligence in the regional sciences: a literature review. European Planning Studies, 31(7), 1304–1324. https://doi.org/10.1080/09654313.2022.2101880

Lee, J.-U., Puerto, H., van Aken, B., Arase, Y., Forde, J. Z., Derczynski, L., Rücklé, A., Gurevych, I., Schwartz, R., Strubell, E., & Dodge, J. (2023). Surveying (Dis)Parities and Concerns of Compute Hungry NLP Research. https://doi.org/10.48550/arXiv.2306.16900

Leonelli, S. (2018). Rethinking Reproducibility as a Criterion for Research Quality. In Including a Symposium on Mary Morgan: Curiosity, Imagination, and Surprise (Research in the History of Economic Thought and Methodology,, Vol. 36B) (pp. 129–146). Emerald Publishing Limited, Leeds. https://doi.org/10.1108/S0743–415420180000036B009

Leonelli, S. (2020). Learning from Data Journeys. In S. Leonelli & N. Tempini (Eds.), Data Journeys in the Sciences (pp. 1–24). Springer International Publishing. https://doi.org/10.1007/978–3–030–37177–7_1

Leonelli, S. (2023a). Opacity and reproducibility in data processing: Reflections on the dependence of AI on the data ecosystem. In S. Andreas, E. Anna, R. Markus, R. Fabian, S. Jens, & W. Alexander (Eds.), Beyond Quantity (pp. 313–324). transcript Verlag. https://doi.org/10.1515/9783839467664–017

Leonelli, S. (2023b). Philosophy of Open Science. Cambridge University Press. https://doi.org/10.1017/9781009416368

Levecque, K., Anseel, F., De Beuckelaer, A., Van der Heyden, J., & Gisle, L. (2017). Work organization and mental health problems in PhD students. Research Policy, 46(4), 868–879. https://doi.org/10.1016/j.respol.2017.02.008

Lewis, D. (2023). Brain-spine interface allows paralysed man to walk using his thoughts. Nature, 618(7963), 18. https://doi.org/10.1038/d41586–023–01728–0

Lewis, G., Millett, P., Sandberg, A., Snyder-Beattie, A., & Gronvall, G. (2019). Information Hazards in Biotechnology. Risk Analysis, 39(5), 975–981. https://doi.org/10.1111/risa.13235

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Advances in Neural Information Processing Systems 33 (NeurIPS 2020),

Li, H., Su, Y., Cai, D., Wang, Y., & Liu, L. (2022). A survey on retrieval-augmented text generation. arXiv preprint arXiv:2202.01110. https://doi.org/10.48550/arXiv.2202.01110

Li, S., Han, C., Yu, P., Edwards, C., Li, M., Wang, X., Fung, Y. R., Yu, C., Tetreault, J. R., Hovy, E. H., & Ji, H. (2023). Defining a New NLP Playground. https://doi.org/10.48550/arXiv.2310.20633

Ligozat, A.-L., Lefevre, J., Bugeau, A., & Combaz, J. (2022). Unraveling the Hidden Environmental Impacts of AI Solutions for Environment Life Cycle Assessment of AI Solutions. Sustainability, 14(9), 5172. https://www.mdpi.com/2071–1050/14/9/5172

Lindgren, H., & Heintz, F. (2023, March 6–8, 2023). The wasp-ed AI curriculum : A holistic curriculum for artificial intelligence INTED 2023, 17th International Technology, Education and Development Conference, Valencia, Spain. http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-206835

Liu, Z., & Tegmark, M. (2022). Machine Learning Hidden Symmetries. Physical Review Letters, 128(18), 180201. https://doi.org/10.1103/PhysRevLett.128.180201

Liverpool, L. (2023). AI intensifies fight against 'paper mills' that churn out fake research. Nature, 618(7964), 222–223. https://doi.org/10.1038/d41586–023–01780-w

López Cobo, M., De Prato, G., Alaveras, G., Righi, R., Samoili, S., Hradec, J., Ziemba, L. W., Pogorzelska, K., & Cardona, M. (2019). Artificial intelligence, high performance computing and cybersecurity. P. O. o. t. E. Union. https://data.europa.eu/doi/10.2760/016541

Lorenz, P., Perset, K., & Berryhill, J. (2023). Initial policy considerations for generative artificial intelligence. https://doi.org/10.1787/fae2d1e6-en

References

Luitse, D., & Denkena, W. (2021). The great Transformer: Examining the role of large language models in the political economy of AI. Big Data & Society, 8, 205395172110477. https://doi.org/10.1177/20539517211047734

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. Journal of the Association for Information Science and Technology, 74(5), 570–581. https://doi.org/10.1002/asi.24750

Mao, Y., Rafner, J., Wang, Y., & Sherson, J. (2023). A Hybrid Intelligence Approach to Training Generative Design Assistants: Partnership Between Human Experts and AI Enhanced Co-Creative Tools. In P. Lukowicz, S. Mayer, J. Koch, J. Shawe-Taylor, & I. Tiddi (Eds.), HHAI 2023: Augmenting Human Intellect (pp. 108 - 123). IOS Press. https://ebooks.iospress.nl/doi/10.3233/FAIA230078

Margoni, T., & Kretschmer, M. (2022). A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology. GRUR International, 71(8), 685–701. https://doi.org/10.1093/grurint/ikac054

Mariani, M. M., Machado, I., Magrelli, V., & Dwivedi, Y. K. (2023). Artificial intelligence in innovation research: A systematic review, conceptual framework, and future research directions. Technovation, 122, 102623. https://doi.org/10.1016/j.technovation.2022.102623

Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Terah Lyons, Manyika, J., Ngo, H., Niebles, J. C., Parli, V., Shoham, Y., Russell Wald, J. C., & Perrault, R. (2023). The AI Index 2023 Annual Report. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2023. https://aiindex.stanford.edu/report/

Mattijssen, L. M. S., Bergmans, J. E., Van Der Weijden, I. C. M., & Teelken, J. C. (2020). In the eye of the storm: the mental health situation of PhD candidates. Perspectives on Medical Education, 10(2), 71–72. https://doi.org/10.1007/S40037–020–00639–4

Mayer, H. M. (2021). Revolutionary NLP Model GPT-3 Poised To Redefine AI And Next Generation Of Startups. https://www.forbes.com/sites/hannahmayer/2021/01/02/revolutionary-nlp-model-gpt-3-poised-to-redefine-ai-and-next-generation-of-startups/?sh=296feac877b3

McDermott, M. B. A., Wang, S., Marinsek, N., Ranganath, R., Foschini, L., & Ghassemi, M. (2021). Reproducibility in machine learning for health research: Still a ways to go. SCIENCE TRANSLATIONAL MEDICINE, 13(586), eabb1655. https://doi.org/10.1126/scitranslmed.abb1655

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). A Survey on Bias and Fairness in Machine Learning. https://doi.org/10.48550/ARXIV.1908.09635

Merchant, A., Batzner, S., Schoenholz, S. S., Aykol, M., Cheon, G., & Cubuk, E. D. (2023). Scaling deep learning for materials discovery. Nature, 624(7990), 80–85. https://doi.org/10.1038/s41586–023–06735–9

Merow, C., Serra-Diaz, J., Enquist, B., & Wilson, A. (2023). AI chatbots can boost scientific coding. NATURE ECOLOGY & EVOLUTION, 7(7), 960–962. https://doi.org/10.1038/s41559–023–02063–3

Merton, R., & Sztomka, P. (1996). Per Merton RK. 1942. The Ethos of Science, J. Legal and Political Sociology. 1: 115-126. Reprinted. In Social Structure and Science. University of Chicago Press, Chicago.

Metcalf, J., & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. Big Data & Society, 3(1), 2053951716650211. https://doi.org/10.1177/2053951716650211

Metcalf, J., Moss, E., Watkins, E., Watkins, E., Singh, R., & Elish, M. C. (2021). Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts ACM Conference on Fairness, Accountability,and Transparency (FAccT '21), Virtual Event, Canada.ACM.

Miailhe, N. (2018). The geopolitics of artificial intelligence: The return of empires? Politique étrangère, Autumn Issue(3), 105–117. https://www.cairn-int.info/article-E_PE_183_0105–.htm

Micklitz, H. W. (2023). The Role of Standards in Future EU Digital Policy Legislation. https://www.beuc.eu/sites/default/files/publications/BEUC-X-2023–096_The_Role_of_Standards_in_Future_EU_Digital_Policy_Legislation.pdf

References

Mihai, F., Aleca, O. E., & Gheorghe, M. (2023). Digital Transformation Based on AI Technologies in European Union Organizations. Electronics, 12(11), 2386. https://www.mdpi.com/2079–9292/12/11/2386

Milano, S., McGrane, J. A., & Leonelli, S. (2023). Large language models challenge the future of higher education. Nature Machine Intelligence, 5(4), 333–334. https://doi.org/10.1038/s42256–023–00644–2

Mitchell, M. (2023). AI's challenge of understanding the world. Science, 382(6671), eadm8175. https://doi.org/doi:10.1126/science.adm8175

Montgomery, J. (2019). The AI revolution in science: applications and new research directions. https://royalsociety.org/blog/2019/08/the-ai-revolution-in-science/

Morgan, F. E., Boudreaux, B., Lohn, A. J., Ashby, M., Curriden, C., Klima, K., & Grossman, D. (2020). Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World. RAND Corporation. https://doi.org/10.7249/RR3139–1

Morse, L., Teodorescu, M. H. M., Awwad, Y., & Kane, G. C. (2022). Do the Ends Justify the Means? Variation in the Distributive and Procedural Fairness of Machine Learning Algorithms. Journal of Business Ethics, 181(4), 1083–1095. https://doi.org/10.1007/s10551–021–04939–5

Mukhamediev, R. I., Popova, Y., Kuchin, Y., Zaitseva, E., Kalimoldayev, A., Symagulov, A., Levashenko, V., Abdoldina, F., Gopejenko, V., Yakunin, K., Muhamedijeva, E., & Yelis, M. (2022). Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges. Mathematics, 10(15), 2552. https://www.mdpi.com/2227–7390/10/15/2552

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press. https://nyupress.org/9781479837243/algorithms-of-oppression/

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group.

OCDE. (2021). Reducing the precarity of academic research careers. https://doi.org/10.1787/0f8bd468-en

OCDE. (2022). Measuring the environmental impacts of artificial intelligence compute and applications. https://doi.org/10.1787/7babf571-en

OCDE. (2023). Artificial Intelligence in Science. https://doi.org/10.1787/a8d820bd-en

OECD. (2007). Best Practices for Ensuring Scientific Integrity and Preventing Misconduct. https://web-archive.oecd.org/2012–06–15/129568–40188303.pdf

OECD. (2020). Building digital workforce capacity and skills for data intensive science. OECD Publishing. https://doi.org/10.1787/e08aa3bb-en

OECD. (2023). OECD Employment Outlook 2023 : Artificial Intelligence and the Labour Market. https://doi.org/10.1787/08785bba-en

Offert, F., & Bell, P. (2021). Perceptual bias and technical metapictures: critical machine vision as a humanities challenge. AI & SOCIETY, 36(4), 1133–1144. https://doi.org/10.1007/s00146–020–01058-z

Oliveira, A. L., Domingos, T., Figueiredo, M., & Lima, P. U. (2023). DeepThought: An Architecture for Autonomous Self-motivated Systems. arXiv preprint arXiv:2311.08547. https://doi.org/10.48550/arXiv.2311.08547

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Leoni Aleman, F., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2023). GPT-4 Technical Report. arXiv:2303.08774. Retrieved March 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230308774O

Packin, N. G. (2021). Disability Discrimination Using AI Systems, Social Media and Digital Platforms: Can We Disable Digital Bias? Journal of International and Comparative Law, 8(2), 487. https://doi.org/10.2139/ssrn.3724556

Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. Nature, 613(7942), 138–144. https://doi.org/10.1038/s41586–022–05543-x

References

Pasquetto, I. V., Jahani, E., Atreja, S., & Baum, M. (2022). Social Debunking of Misinformation on WhatsApp: The Case for Strong and In-group Ties. Proc. ACM Hum.-Comput. Interact., 6(CSCW1), Article 117. https://doi.org/10.1145/3512964

Pasquetto, I. V., Olivieri, A. F., Tacchetti, L., Riotta, G., & Spada, A. (2022). Disinformation as Infrastructure: Making and Maintaining the QAnon Conspiracy on Italian Digital Media. Proc. ACM Hum.-Comput. Interact., 6(CSCW1), Article 84. https://doi.org/10.1145/3512931

Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its (dis)contents: A survey of dataset development and use in machine learning research. Patterns, 2(11), 100336. https://doi.org/10.1016/j.patter.2021.100336

Păvăloaia, V. D., & Necula, S. C. (2023). Artificial Intelligence as a Disruptive Technology—A Systematic Literature Review. Electronics (Switzerland), 12(5). https://doi.org/10.3390/electronics12051102

Pearson, K. A., Palafox, L., & Griffith, C. A. (2017). Searching for exoplanets using artificial intelligence. Monthly Notices of the Royal Astronomical Society, 474(1), 478–491. https://doi.org/10.1093/mnras/stx2761

Pérez-Jvostov, F., Iron, K., Khair, S., Sahrakorpi, S., & Zhang, Q. (2021). Researcher Needs Assessment: summary of what we heard. D. R. A. o. Canada. https://alliancecan.ca/sites/default/files/2022–03/needsassessment_alliance_20220126.pdf

Pournaras, E. (2023). Science in the Era of ChatGPT, Large Language Models and Generative AI: Challenges for Research Ethics and How to Respond. arXiv:2305.15299. Retrieved May 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230515299P

Pretorius, L. (2023). Fostering AI literacy: A teaching practice reflection. JOURNAL OF ACADEMIC LANGUAGE AND LEARNING, 17(1), T1-T8.

Procter, R., Glover, B., & Jones, E. (2020). Research 4.0: Research in the Age of Automation. https://demos.co.uk/wp-content/uploads/2020/09/Research-4.0-Report.pdf

Qazi, S., Khawaja, B. A., & Farooq, Q. U. (2022). IoT-Equipped and AI-Enabled Next Generation Smart Agriculture: A Critical Review, Current Challenges and Future Trends. IEEE Access, 10, 21219–21235. https://doi.org/10.1109/ACCESS.2022.3152544

Quinn, P. (2021). Research under the GDPR – a level playing field for public and private sector research? Life Sciences, Society and Policy, 17(1), 4. https://doi.org/10.1186/s40504–021–00111-z

Rafner, J., Bantle, C., Dellermann, D., Söllner, M., Zaggl, M. A., & Sherson, J. (2022). Towards hybrid intelligence workflows: integrating interface design and scalable deployment. The 1st International Conference on Hybrid Human-Artificial Intelligence (HHAI2022), Amsterdam, Netherlands.

Rafner, J., Beaty, R. E., Kaufman, J. C., Lubart, T., & Sherson, J. (2023). Creativity in the age of generative AI. Nature Human Behaviour, 7(11), 1836–1838. https://doi.org/10.1038/s41562–023–01751–1

Rafner, J., Dellermann, D., Hjorth, A., Verasztó, D., Kampf, C., Mackay, W., & Sherson, J. (2021). Deskilling, Upskilling, and Reskilling: a Case for Hybrid Intelligence. Morals & Machines, 1, 24–39. https://doi.org/10.5771/2747–5174–2021–2–24

Rafner, J., Gajdacz, M., Kragh, G., Hjorth, A., Gander, A., Palfi, B., Berditchevskiaia, A., Grey, F., Gal, K., Segal, A., Wamsley, M., Miller, J., Dellermann, D., Haklay, M., Michelucci, P., & Sherson, J. (2022). Mapping Citizen Science through the Lens of Human-Centered AI. Human Computation, 9(1), 66–95. https://doi.org/10.15346/hc.v9i1.133

Rahman, M., & Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. APPLIED SCIENCES-BASEL, 13(9). https://doi.org/10.3390/app13095783

Rallabhandi, K. (2023). The Copyright Authorship Conundrum for Works Generated by Artificial Intelligence: A Proposal for Standardized International Guidelines in the WIPO Copyright Treaty. Geo. Wash. Int'l L. Rev, 55(2), 311–347. https://www.proquest.com/openview/fdfb424b3c88e9b516cdb5c7d2a50026/1?pq-origsite=gscholar&cbl=44595

Rességuier, A., & Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. Big Data & Society, 7(2), 2053951720942541. https://doi.org/10.1177/2053951720942541

References

Reuer, K., Landgraf, J., Fösel, T., O'Sullivan, J., Beltrán, L., Akin, A., Norris, G. J., Remm, A., Kerschbaum, M., Besse, J.-C., Marquardt, F., Wallraff, A., & Eichler, C. (2023). Realizing a deep reinforcement learning agent for real-time quantum feedback. Nature Communications, 14(1), 7138. https://doi.org/10.1038/s41467–023–42901–3

Rikap, C. (2023a). The expansionary strategies of intellectual monopolies: Google and the digitalization of healthcare. Economy and Society, 52(1), 110–136. https://doi.org/10.1080/03085147.2022.2131271

Rikap, C. (2023b). Intellectual monopolies as a new pattern of innovation and technological regime. Industrial and Corporate Change. https://doi.org/10.1093/icc/dtad077

Rikap, C. (2023c). Same End by Different Means: Google, Amazon, Microsoft and Facebook's Strategies to Dominate Artificial Intelligence. Available at SSRN: https://ssrn.com/abstract=4472222. https://doi.org/10.2139/ssrn.4472222

Rikap, C. (2023d). Working Paper - Mapping the cloud: Big Tech taking the sky by storm. CITYPERC Working Paper, No. 2023–05, City, University of London, City Political Economy Research Centre (CITYPERC), London. https://www.econstor.eu/bitstream/10419/280831/1/1850871124.pdf

Rikap, C., & Lundvall, B.-Å. (2022). The Digital Innovation Race: Conceptualizing the Emerging New World Order. Palgrave Macmillan Cham. https://doi.org/10.1007/978–3–030–89443–6

Ringel Morris, M., Sohl-dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., & Legg, S. (2023). Levels of AGI: Operationalizing Progress on the Path to AGI. arXiv:2311.02462. Retrieved November 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv231102462R

Rogers, A., Balasubramanian, N., Derczynski, L., Dodge, J., Koller, A., Luccioni, S., Sap, M., Schwartz, R., Smith, N. A., & Strubell, E. (2023). Closed AI Models Make Bad Baselines. Hacking Semantics. https://hackingsemantics.xyz/2023/closed-baselines/

Rosenberg, N. (1985). The commercial exploitation of science by American industry.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256–019–0048-x

Ruiz-Gonzalez, C., Arlt, S., Petermann, J., Sayyad, S., Jaouni, T., Karimi, E., Tischler, N., Gu, X., & Krenn, M. (2023). Digital discovery of 100 diverse quantum experiments with PyTheus. Quantum, 7, 1204. https://doi.org/10.22331/q-2023–12–12–1204

Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021). "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan. https://doi.org/10.1145/3411764.3445518

Sankar Sadasivan, V., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). Can AI-Generated Text be Reliably Detected? , arXiv:2303.11156. Retrieved March 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230311156S

Saphra, N., Fleisig, E., Cho, K., & Lopez, A. (2023). First Tragedy, then Parse: History Repeats Itself in the New Era of Large Language Models. https://doi.org/10.48550/arXiv.2311.05020

Schleiss, J., Laupichler, M., Raupach, T., & Stober, S. (2023). AI Course Design Planning Framework: Developing Domain-Specific AI Education Courses. EDUCATION SCIENCES, 13(9). https://doi.org/10.3390/educsci13090954

Schölkopf, B., Locatello, F., Bauer, S., Ke, N., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward Causal Representation Learning. Proceedings of the IEEE, PP, 1–23. https://doi.org/10.1109/JPROC.2021.3058954

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. Communications of the ACM, 63(12), 54–63. https://doi.org/10.1145/3381831

Sejnowski, T. J. (2020). The unreasonable effectiveness of deep learning in artificial intelligence. Proceedings of the National Academy of Sciences, 117(48), 30033–30038. https://doi.org/10.1073/pnas.1907373117

References

Sherson, J., Rabecq, B., Dellermann, D., & Rafner, J. (2023). A Multi-Dimensional Development and Deployment Framework for Hybrid Intelligence. In P. Lukowicz, S. Mayer, J. Koch, J. Shawe-Taylor, & I. Tiddi (Eds.), HHAI 2023: Augmenting Human Intellect (Vol. 368, pp. 429 - 432). IOS Press. https://doi.org/10.3233/FAIA230119

Soliman, M., Fatnassi, T., Elgammal, I., & Figueiredo, R. (2023). Exploring the Major Trends and Emerging Themes of Artificial Intelligence in the Scientific Leading Journals amidst the COVID-19 Era. BIG DATA AND COGNITIVE COMPUTING, 7(1), 12. https://www.mdpi.com/2504–2289/7/1/12

Srnicek, N. (2016). Platform Capitalism. https://www.wiley.com/en-us/Platform+Capitalism-p-9781509504862

Su, M., Peng, H., & Li, S. (2022). A visualized bibliometric analysis of mapping research trends of machine learning in engineering (MLE). Expert Syst. Appl., 186, 11. https://doi.org/10.1016/j.eswa.2021.115728

Sweeney, L. (2013). Discrimination in Online Ad Delivery. Available at SSRN: https://ssrn.com/abstract=2208240. https://doi.org/10.2139/ssrn.2208240

Szymanski, N. J., Rendy, B., Fei, Y., Kumar, R. E., He, T., Milsted, D., McDermott, M. J., Gallant, M., Cubuk, E. D., Merchant, A., Kim, H., Jain, A., Bartel, C. J., Persson, K., Zeng, Y., & Ceder, G. (2023). An autonomous laboratory for the accelerated synthesis of novel materials. Nature, 624(7990), 86–91. https://doi.org/10.1038/s41586–023–06734-w

Tamburrini, G. (2022). The AI Carbon Footprint and Responsibilities of AI Scientists. Philosophies, 7(1), 4. https://www.mdpi.com/2409–9287/7/1/4

Tapeh, A. T. G., & Naser, M. Z. (2023). Artificial Intelligence, Machine Learning, and Deep Learning in Structural Engineering: A Scientometrics Review of Trends and Best Practices. Archives of Computational Methods in Engineering, 30(1), 115–159. https://doi.org/10.1007/s11831–022–09793-w

Thelwall, M., Kousha, K., Wilson, P., Makita, M., Abdoli, M., Stuart, E., Levitt, J., Knoth, P., & Cancellieri, M. (2023). Predicting article quality scores with machine learning: The U.K. Research Excellence Framework. Quantitative Science Studies, 4(2), 547–573. https://doi.org/10.1162/qss_a_00258

Thomas, R., Bhosale, U., Shukla, K., & Kapadia, A. (2023). Impact and perceived value of the revolutionary advent of artificial intelligence in research and publishing among researchers: a survey-based descriptive study. Sci Ed, 10(1), 27–34. https://doi.org/10.6087/kcse.294

Togelius, J., & Yannakakis, G. N. (2023). Choose Your Weapon: Survival Strategies for Depressed AI Academics. arXiv:2304.06035. Retrieved March 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230406035T

Trinh, T. H., Wu, Y., Le, Q. V., He, H., & Luong, T. (2024). Solving olympiad geometry without human demonstrations. Nature, 625(7995), 476–482. https://doi.org/10.1038/s41586–023–06747–5

Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K. A., Ceder, G., & Jain, A. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. Nature, 571(7763), 95–98. https://doi.org/10.1038/s41586–019–1335–8

Tuomi, I., Cachia, R., & Villar-Onrubia, D. (2023). On the futures of technology in education: Emerging trends and policy implications. Publications Office of the European Union. https://doi.org/10.2760/079734

UNESCO. (2019). Planning education in the AI era: Lead the leap International Conference on Artificial Intelligence and Education: Final report, https://unesdoc.unesco.org/ark:/48223/pf0000370967

United Nations, A. A. B. (2023). Interim Report: Governing AI for Humanity. https://www.un.org/en/ai-advisory-body

Upshall, M. (2022). An AI toolkit for libraries. INSIGHTS-THE UKSG JOURNAL, 35. https://doi.org/10.1629/uksg.592

Valavi, E., Hestness, J., Ardalani, N., & Iansiti, M. (2022). Time and the Value of Data. arXiv:2203.09118. Retrieved March 01, 2022, from https://ui.adsabs.harvard.edu/abs/2022arXiv220309118V

References

van Erp, M., Tullett, W., Christlein, V., Ehrhart, T., Hürriyetoğlu, A., Leemans, I., Lisena, P., Menini, S., Schwabe, D., Tonelli, S., Troncy, R., & Zinnen, M. (2023). More than the Name of the Rose: How to Make Computers Read, See, and Organize Smells. The American Historical Review, 128(1), 335–369. https://doi.org/10.1093/ahr/rhad141

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, Vol 30, arXiv:1706.03762. https://doi.org/10.48550/arXiv.1706.03762

Verdegem, P. (2022). Dismantling AI capitalism: the commons as an alternative to the power concentration of Big Tech. AI Soc, 1–11. https://doi.org/10.1007/s00146–022–01437–8

Veselovsky, V., Horta Ribeiro, M., & West, R. (2023). Artificial Artificial Artificial Intelligence: Crowd Workers Widely Use Large Language Models for Text Production Tasks. arXiv:2306.07899. Retrieved June 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230607899V

Villalobos, P., Sevilla, J., Heim, L., Besiroglu, T., Hobbhahn, M., & Ho, A. (2022). Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning. https://doi.org/10.48550/arXiv.2211.04325

Voosen, P. (2023). AI churns out lightning-fast forecasts as good as the weather agencies'. Science, 382(6672). https://doi.org/10.1126/science.adm9275

Vuorikari Rina, R., Kluzer, S., & Punie, Y. (2022). DigComp 2.2: The Digital Competence Framework for Citizens - With new examples of knowledge, skills and attitudes. https://EconPapers.repec.org/RePEc:ipt:iptwpa:jrc128415

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., Anandkumar, A., Bergen, K., Gomes, C. P., Ho, S., Kohli, P., Lasenby, J., Leskovec, J., Liu, T. Y., Manrai, A., … Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. Nature, 620(7972), 47–60. https://doi.org/10.1038/s41586–023–06221–2

Wang, Q., Downey, D., Ji, H., & Hope, T. (2023). Learning to Generate Novel Scientific Directions with Contextualized Literature-based Discovery. arXiv preprint arXiv:2305.14259. https://doi.org/10.48550/arXiv.2305.14259

Wang, W., Wang, G., Marivate, V., & Hufton, A. L. (2023). On the transparency of large AI models. Patterns (N Y), 4(7), 100797. https://doi.org/10.1016/j.patter.2023.100797

Welker, Y. (2023a). For disabilities and designated groups, the Digital Services and Market Acts may complement AI and data policies to ensure algorithmic safety and accountability. OECD.AI Policy Observatory. https://oecd.ai/en/wonk/disabilities-designated-groups-digital-services-market-acts

Welker, Y. (2023b). Generative AI holds great potential for those with disabilities - but it needs policy to shape it. World Economic Forum. https://www.weforum.org/agenda/2023/11/generative-ai-holds-potential-disabilities/

Whittaker, M., Alper, M., Bennett, C. L., Hendren, S., Kaziunas, L., Mills, M., Morris, M. R., Rankin, J., Rogers, E., & West, S. M. (2019). Disability, Bias, and AI (Workshop Report from the AI Now Institute at New York University (NYU), the NYU Center for Disability Studies, and Microsoft, Issue. https://ainowinstitute.org/wp-content/uploads/2023/04/disabilitybiasai-2019.pdf

Wong, F., Zheng, E. J., Valeri, J. A., Donghia, N. M., Anahtar, M. N., Omori, S., Li, A., Cubillos-Ruiz, A., Krishnan, A., Jin, W., Manson, A. L., Friedrichs, J., Helbig, R., Hajian, B., Fiejtek, D. K., Wagner, F. F., Soutter, H. H., Earl, A. M., Stokes, J. M., … Collins, J. J. (2023). Discovery of a structural class of antibiotics with explainable deep learning. Nature. https://doi.org/10.1038/s41586–023–06887–8

Woolston, C. (2018). Science PhDs lead to enjoyable jobs. Nature, 555, 277. https://doi.org/10.1038/d41586–018–02696–6

Xue, M., Cao, X., Feng, X., Gu, B., & Zhang, Y. (2022). Is College Education Less Necessary with AI? Evidence from Firm-Level Labor Structure Changes. Journal of Management Information Systems, 39(3), 865–905. https://doi.org/10.1080/07421222.2022.2096542

Yeung, K. (2023). Dispelling the Digital Enchantment: how can we move beyond its destructive influence and reclaim our right to an open future? Prometheus, 39(1), 8–27. https://doi.org/10.13169/prometheus.39.1.0008

References

Yeung, K., Howes, A., & Pogrebna, G. (2019). AI Governance by Human Rights-Centred Design, Deliberation and Oversight: An End to Ethics Washing. In I. M. D. a. F. P. (eds.) (Ed.), The Oxford Handbook of AI Ethics, Oxford University Press (2019). https://doi.org/10.2139/ssrn.3435011

You, H., Zhang, H., Gan, Z., Du, X., Zhang, B., Wang, Z., Cao, L., Chang, S.-F., & Yang, Y. (2023). Ferret: Refer and Ground Anything Anywhere at Any Granularity. arXiv:2310.07704. Retrieved October 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv231007704Y

Yu, N., Davis, L., & Fritz, M. (2018). Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints. arXiv:1811.08180. Retrieved November 01, 2018, from https://ui.adsabs.harvard.edu/abs/2018arXiv181108180Y

Yu, N., Skripniuk, V., Chen, D., Davis, L., & Fritz, M. (2020). Responsible Disclosure of Generative Models Using Scalable Fingerprinting. arXiv:2012.08726. Retrieved December 01, 2020, from https://ui.adsabs.harvard.edu/abs/2020arXiv201208726Y

Zahlan, A., Ranjan, R. P., & Hayes, D. (2023). Artificial intelligence innovation in healthcare: Literature review, exploratory analysis, and future research. Technology in Society, 74. https://doi.org/10.1016/j.techsoc.2023.102321

Zhang, Y., Luo, M. Q., Wu, P., Wu, S., Lee, T. Y., & Bai, C. (2022). Application of Computational Biology and Artificial Intelligence in Drug Design. INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES, 23(21), Article 13568. https://doi.org/10.3390/ijms232113568

Zou, A., Wang, Z., Carlini, N., Nasr, M., Zico Kolter, J., & Fredrikson, M. (2023). Universal and Transferable Adversarial Attacks on Aligned Language Models. arXiv:2307.15043. Retrieved July 01, 2023, from https://ui.adsabs.harvard.edu/abs/2023arXiv230715043Z

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs. https://www.amazon.com/Age-Surveillance-Capitalism-Future-Frontier/dp/1610395697

Annex 1. Responsibilities and working structure within the Scientific Advice Mechanism

- The Group of Chief Scientific Advisors was responsible for developing the Scientific Opinion, which contains evidence-based policy recommendations. Four members of the Group were involved with the project, namely the Chair of the Group, Nicole Grobert, as well as Maarja Kruusmaa and Alberto Melloni.
- The Science Policy, Advice and Ethics Unit at DG RTD (the Secretariat) assisted the Advisors in the development of their Scientific Opinion. Ingrid Zegers, Jean-François Dechamp, Daniela Melandri and Gintarė Juškaitė coordinated the project.
- SAPEA was responsible for independently producing the rapid Evidence Review Report that informs the Scientific Opinion. Within SAPEA, Euro-CASE served as lead Academy Network for the topic. Marie Franquin, Euro-CASE Scientific Policy Officer, coordinated the report's development, with the support of the SAPEA team of scientific policy officers: Louise Edwards (Academia Europaea), Stephany Mazon (YASAS), Celine Tschirhart (ALLEA), Rafael Carrascosa Marzo (Academia Europaea), and Rúben Castro (FEAM).

To jointly coordinate the project within the SAM, regular SAM coordination team meetings took place, chaired by Nicole Grobert. The participants from SAPEA were the co-chairs of the SAPEA working group, the Board member of the network leading on the topic, and members of staff supporting the project.

This rapid Evidence Review Report addresses key areas 2, 3 and 4 of the scoping paper. Key area 1 of the Scoping Paper was addressed separately by SAPEA through a foresight workshop and report that was handed over to the Commission at the end of 2023. The foresight workshop report is published and available on the SAM website.

Annex 2. Selection of experts

In line with the SAPEA quality assurance guidelines, SAPEA set up an interdisciplinary working group with six members from six European countries, chaired by Anna Fabijańska, Andrea Emilio Rizzoli (from 19 September 2023), and Virginia Dignum (until 6 September 2023).

The co-chairs of the working group were proposed by the lead Academy Network, Euro-CASE, and approved by the SAPEA board after their declarations of interests were assessed.

SAPEA issued a call for nominations describing the scope, timeline and expertise required. The areas of expertise were previously discussed and coordinated with the Advisors and the Secretariat. The call for nominations was sent via the Academy Networks to their member academies, which were invited to nominate experts. Experts were also identified through desk research by the Academy Networks.

The selection committee for the working group met on 28 August 2023. In line with the SAPEA quality assurance guidelines, the selection committee comprised:

- the working group co-chairs (Virginia Dignum and Anna Fabijańska)
- the secretary-general of the lead Academy Network, Euro-CASE (Patrick Maestro)
- the president of another SAPEA Network, Academia Europaea (Marja Makarow)

SAPEA received a total of 172 nominations for the working group. The experts were selected on the basis of scientific excellence and disciplinary requirements as a priority, taking into account commitment and time availability, the criteria set out in our strategy of diversity and inclusiveness, and other requirements communicated to the committee in advance:

- inter- and multidisciplinarity
- involvement in the wider scientific community, i.e. not Fellows of academies
- inclusion of early- and mid-career researchers
- gender balance
- wide geographical coverage, including from Widening countries

In the final working group, 50% of selected experts were female and 67% were mid-career researchers. 6 European countries are represented in the group, with 2 members from central/eastern Europe, 1 from southern Europe, and 3 from Western Europe. 3 experts came from Widening Countries.[7]

[7] These calculations reflect the final working group composition, i.e. the working group members who developed the content of the evidence review report.

Annex 2. Selection of experts

The composition of the working group was approved by the SAPEA Board. All working group members were required to complete the Standard Declaration of Interests form of the European Commission, in accordance with SAPEA's quality guidelines. In the assessment, no conflicts of interests were detected.

Annex 3. Evidence review process

We compiled this rapid Evidence Review Report based on input from the experts and their in-depth knowledge of the field, together with literature reviews conducted on specific key areas of the Scoping Paper (see Annex 5), and 3 evidence-gathering workshops. In terms of data management, SAPEA commits to Open Science and FAIR principles.

The evidence necessary to respond to the question in the Scoping Paper was discussed, debated and assessed by the Working Group members at Working Group meetings, and was written up in iterative drafts of the Report. The literature reviewed for this report was not systematically checked for sponsorship or authors' conflict of interest statements.

The final draft underwent a double-blind peer review.

Timeline

- **September 2023:** Final formation of working group
- **October 2023:** Working group meeting
- **November 2023:** Evidence-gathering expert workshop (key area 2)
- **December 2023:** Evidence-gathering expert workshop (key area 3); working group meeting
- **January 2024:** Evidence-gathering expert workshop (key area 4); working group meeting; production of first draft
- **February 2024:** Peer review; working group addresses peer reviewers' comments; production of final draft
- **March 2024:** SAPEA endorsement
- **April 2024:** Publication of Scientific Opinion and rapid evidence review report

Requested literature reviews

A literature review team was formed, comprising information specialists and methodologists at Cardiff University, who are responsible for conducting systematic literature reviews. The European Information Hub at Cardiff University was also responsible for developing an EU policy mapping to support the work.

To complement their knowledge, the working group made use of literature searches on:

- Area 1, Deep Dive 1 and Area 2, Deep Dive 2 (rapid review and synthesis of results), including a bibliometric analysis
- request on industry collaboration on published papers

- Area 3 (rapid review and synthesis of results)
- Area 4 (rapid review and synthesis of results, abstracts only)
- Area 4, additional focused questions:
    - How significant is the research in AI undertaken by private sector tech firms within Europe and across the globe?
    - To what extent do the principles of research ethics (based on the Helsinki Declaration) which apply to university research involving human subjects apply to research undertaken outside university laboratories?
    - What is the current state of the art in understanding the content, scope and application of the (a) GDPR research exception, (b) text and data mining exception, (c) progress on the EU plant o t develop an EU Copyright and Data Legislative framework for research?
    - How does the European Commission's Open Science policy and the current EU copyright law framework (including the Information Society Directive 2001) affect access to, and re-use of copyright protected works for the purposes of scientific research?
- sources of funding from the private and public sector into AI research
- evaluation of large, publicly funded research structures and their outcomes
- current measures to help identify and reduce environmental footprint of ICT use including AI
- international political negotiations about weapons treaties for AI, especially autonomous weapons

The rapid reviews were conducted systematically, and protocols were recorded and submitted alongside the screened results, and EndNote files were retained with all the extracted results. Bibliographic databases such as Scopus and Web of Science (and others) were used in the literature searches, alongside further screening of grey literature (using Overton) and using EUR-LEX, the EU Publications Office catalogue and other databases European Sources Online. The inclusion/exclusion criteria were discussed with appropriate members of the Working Group (when necessary), as well as other members of the Literature Review Team. Full details and search strategies are provided in Annex 5.

Evidence-gathering expert workshops

In line with our quality assurance guidelines, evidence-gathering expert workshops are a vital part of the rapid evidence review process. Together with the literature reviews, they constitute the main avenue for evidence gathering from the wider scientific community.

Three evidence-gathering expert workshops were organised to support the evidence review for this topic:

- **Key Area 2: The impact of AI on the scientific process:** This workshop was held on 15–16 November 2023 as an online meeting. Participants included 18 invited experts, all members of the SAPEA Working Group, SAPEA representatives and staff, members of the Group of Chief Scientific Advisors, staff members from the Secretariat, and staff of the European Commission. The workshop aimed to gather evidence to answer the questions set out in the scoping paper under this key area:

What is the impact of AI on the scientific process in your area of expertise, and its potential to re-shape science and its governance practices? What is the impact (positive and negative) of AI on everyday scientific practice and workflow in your area of expertise? Explore gaps, potential risks, workflows and checks that could be put in practice in your area of expertise.

- **Key Area 3: People:** This workshop was held on 7 December 2023 as an online meeting. Participants included 9 invited experts, members of the SAPEA Working Group, SAPEA representatives and staff, staff members from the Secretariat, and staff of the European Commission. The workshop aimed to gather evidence to answer the questions set out in the Scoping Paper, under this key area:

  How can the EU best prepare for the impact and requirements of AI on the education and careers of the scientists and researchers of today and tomorrow, and what skills and competencies should education policies prioritise in this context? What are ways to ensure that researchers (at all stages of their education and professional development) and organisations have sufficient knowledge on using AI in science (and on related skills such as IT and computing, statistics, data analytics) and affordable access to infrastructure, data, computing capacity and AI tools and technologies? Which scientific jobs carry a high risk of being outsourced to AI-based technology; and the impact of AI (taking over some of researchers' tasks) on scientific workforce and researchers' careers?'

- **Key Area 4: Policy design:** This workshop was held on 10 January 2024 as a hybrid meeting online and in Brussels. Participants included 9 invited experts, all members of the SAPEA Working Group, SAPEA representatives and staff, members of the Group of Chief Scientific Advisors, staff members from the Secretariat and staff of the European Commission. The workshop aimed to gather evidence to answer the question set out in the Scoping Paper, under this key area:

  How can the European Commission accelerate a responsible uptake of AI in science (including providing access to high quality AI, respecting European values) in order to boost the EU's innovation and prosperity, strengthen the EU's position in science and ultimately contribute to solving Europe's societal challenges?

Workshop format

For all 3 evidence-gathering workshops, the workshop format was as follows:

- At the beginning of each workshop day, SAPEA and the Advisors provided an introduction to the Scientific Advice Mechanism, the topic and the background to the request.
- One or two working group members were in charge of moderating the talks and discussions for each day. The invited experts presented evidence about the impact of AI on scientific areas and the scientific process.
- Each talk was followed by questions and the day ended with a general discussion between all participants.

The content of each presentation, the opinions shared by the experts and the discussions are summarised in workshop reports, in which the names of all participants can be found, along with the workshop programme. These reports are published along the evidence review report on our website.

Selection of experts

The experts were selected by SAPEA and the member of the SAPEA working group in charge of each workshop on the basis of scientific excellence and disciplinary requirements as a priority, taking into account commitment and time availability, and the criteria set out in our strategy of diversity and inclusiveness:

- inter- and multidisciplinarity
- involvement in the wider scientific community;
- inclusion of early- and mid-career researchers
- gender balance
- wide geographical coverage, including from Widening countries

The list of areas of expertise that should be covered in the workshop was established in coordination with the SAM Secretariat and the member of the SAPEA working group in charge of each workshop. Experts involved in the workshop were selected from the list of nominees for the call for nominations for the topic (see Annex 2). Additional experts were also identified through desk research by the Academy Networks and working group members.

Workshop process

Experts received the scoping paper, along with the questions posed in relevant key areas of the scoping paper, in advance of the workshop. They were asked to present on the scientific topic of interest, related to their area of expertise. In line with the principle of transparency, workshop expert participants were asked to declare any conflict of interests and any interest that might be perceived by SAPEA as giving rise to a conflict of interests in relation to this scientific topic at the beginning of their presentations. Four experts across all three workshops informed participants about a potential conflict of interests; the existence of the potential conflict of interest was acknowledged by the participants during the presentation and the discussions.

Experts also received specific instructions about the format of presentations:

- listing the scientific publications cited in the presentations
- preparing the content of the presentation by drawing on their own research but also on their broad knowledge of the field
- keeping the confidentiality of participants until the publication of the report

The report summarising each workshop was prepared by SAPEA and sent to all experts present for review before publication. To encourage openness and the sharing of information, the Chatham House rule applied, and the public summary report was prepared in an anonymous, non-attributed style.

Peer review

In line with our quality assurance guidelines, we followed a double-blind peer review process. Euro-CASE, the lead Academy Network for this report, established the areas of expertise needed for peer reviewers based on the key areas described in the scoping paper, namely scientific process, people, and policy design.

The partner network YASAS compiled a list of experts based on academy and network nominations. YASAS suggested a list of experts to the SAPEA board based on the areas of expertise defined by Euro-CASE, complementarity of expertise, expertise that included a broad overview of the field rather than in-depth knowledge in a narrow field, taking into account gender and geographical balance, and inclusion of early and mid-career experts. The SAPEA board, excluding Euro-CASE, gave the final approval for the list of peer reviewers to be invited.

Following these directions, four reviewers accepted the invitation. Of these reviewers, two were female, and all were mid-career researchers. One was from a Widening country, two from Southern Europe, and one from Northern Europe and one from Western Europe. Peer reviewers were asked to declare any conflict of interests and any interest that might be perceived by SAPEA as giving rise to a conflict of interests in relation to this scientific topic, using a form which was assessed by Euro-CASE and YASAS. No conflict of interest was detected for any of the peer reviewers.

Responses were received in February 2024, anonymised by YASAS and then shared with Euro-CASE and the working group. Members of the working group reviewed the responses and agreed on the actions that should be taken to address them. The draft rapid evidence review report was then revised.

Revisions following peer review

Peer review comments were positive overall. Three of four peer reviewers found that the report satisfactorily addressed the questions posed in the scoping paper, that the literature cited was up-to-date (some additional literature sources suggested by the peer reviewers were incorporated into the text by the working group), that arguments advanced in the report showed the requisite degree of analytical rigour, that conclusions and policy options were well supported by the scientific evidence, and that there were no signs of biases or undue influence from individuals or interest groups. One peer reviewer reported a lack of acknowledgement of the gaps in evidence of AI in arts and culture. However, the working group found this to be beyond the scope of the report.

In response to comments from the peer reviewers, the working group provided additional clarifications by:

- further clarifying AI as a general-purpose technology, so the general concerns and policies about AI will also impact AI in science. However, the report focuses on specific concerns for science and not the impact of AI on society in general
- further contextualising the policy options by acknowledging the scale and complexity of many of the challenges tackled in the report cannot be easily or quickly resolved

- clarifying any ambiguity between research in AI and AI in research. The report covers AI in research and some aspects of research in AI (the aspects that are relevant for AI in research)

The working group also provided additional evidence and emphasis about:

- additional information on LLMs as well as justification for the strong emphasis of the report on generative AI and LLMs, which are the current state of the art technology in AI
- the importance of cross-border European collaboration (including non-EU countries which are strong in AI, such as UK and Switzerland)
- the impact of AI on the humanities, including further examples and potential for developing new areas of research. Additional evidence revealed that social sciences and humanities are differently affected by the uptake of AI: for example, they are less subject to the 'brain drain'
- the reproducibility 'crisis' and the need for standardisation and transparency in reporting
- the consequences of incorporating private solutions into public research (e.g. data cascade)

A few additional references were also added regarding the need for developing soft skills, environmental protection, FAIR in AI in science, and social inequalities.

After the reviewers' comments were addressed by the working group, the peer reviewers' comments, the working group's responses and actions were sent to the SAPEA Board, which approved the outcome of the peer review process.

Plagiarism check

In accordance with the quality assurance guidelines, a plagiarism check on the main report was run by Cardiff University using Turnitin software.

Publication

This evidence review report is to be handed over to the Group of Chief Scientific Advisors on 25 March 2024. At the time of writing, it is planned to publish in April 2023, along with the Advisors' scientific opinion.

The main report will be accompanied by four parallel documents: three expert workshop reports, and one policy landscape mapping. All documents can be accessed on the SAM website.

Annex 4. Policy landscape summary

The EU policy landscape document provides an overview of legal acts and preparatory documents relevant to understanding the approach developed over the years by the EU on AI. Particular focus was given to preparatory documents made available by the European Commission, as the sole institution with powers of legislative initiative.

- The first section analyses the texts relevant to the EU policy on AI, from the Digital Single Market strategy in 2017 to the EU AI Act, as well as the most recent developments on AI liability, Web 4.0 and security.
- The second section focuses on documents relevant to research and innovation as regards AI, notably the developments in infrastructure sharing and coordination within the European Research Area, boosting private sector innovation, addressing security and intellectual property concerns, and the establishment of an AI Office within the European Commission.
- The third section summarises recent European initiatives and texts to support the development of a pan-European talent pool, high-quality and inclusive digital education and training, and strategies for universities.
- The fourth section addresses other relevant legislation and policy instruments relevant to AI, such as data governance, digital services governance, and developing sustainable digital infrastructure.

The narrative has been produced by Frederico Rocha, on behalf of SAPEA's literature review team. The full policy landscape is available as a separate document, published on the SAM website.

Annex 5. Literature search strategies

The SAPEA consortium supports open science practices. The following search strategies were designed in response to requests for literature reviews made by members of the working group. The strategies show the date of the search, sources searched, keywords and date limits (if applicable). 'N' shows the number of potentially relevant results that were scanned. Where multiple sources have been searched, a deduplication process has taken place.

Key area 1, deep dive 1; key area 2, deep dive 2 (rapid review and synthesis of results)

- **Key area 1, Vision and foresight:** What impetus could AI give to scientific productivity and what benefits, challenges and risks would AI-enabled research bring to the European innovation ecosystem and the society as a whole?
- **Deep dive 1, AI's disruptive potential:** Which scientific domains are experiencing (or could experience in the near future) the most positive impact of AI-enabled research, and in what areas does one expect major breakthroughs? Conversely, in which R&I fields is AI not sufficiently developed yet, also in comparison to other countries?
- **Key area 2, Scientific process:** What is the impact of AI on the scientific process, and its potential to re-shape science and its governance practices?
- **Deep dive 2, AI's impact on scientific practice:** What is the impact (positive and negative) of AI on everyday scientific practice and workflow (such as on hypothesis generation, experiment design, monitoring and simulation, scientific publication of research results, intellectual property rights, etc.)?

Scopus, 02/08/2023, searched by MK

```
TITLE ( ( "artificial* intelligen*" OR ai OR "machine learning" OR "deep learning" OR "neural network*" OR "convolutional network*" ) AND ( trend* OR forecast* OR foresight* OR vision* OR strateg* OR breakthrough* OR impact* OR emerg* OR innovat* OR novelt* OR disrupt* OR understand* OR discover* OR advances OR advancement* OR paradigm* OR productiv* OR challeng* OR opportunit* OR benefit* OR risk* ) AND ( scien* OR research* OR academi* OR scholar* OR studies OR technolog* OR biotechnolog* OR medic* OR health* OR sociolog* OR humanit* OR economics OR physics OR chemistry OR nanotechnology OR "climate change" OR robotics ) AND ( review* OR overview* OR survey* OR reflection* OR analys* OR outline* ) )
```

Retrieved all results for screening. Papers retrieved: 763. Automatically deduplicated in EndNote by author, year, title. 2 duplicates discarded. Imported the WoS and ACM searches into the same library (see rows

below). 1169 records after automatic deduplication. An additional 282 records removed during manual deduplication, resulting in 887 records. The deduplicated library was exported to Rayyan for screening.

Web of Science core collection, 02/08/2023, searched by MK

```
TI=(("artificial* intelligen*" OR ai OR "machine learning" OR "deep learning" OR "neural network*" OR "convolutional network*") AND (trend* OR forecast* OR foresight* OR vision* OR strateg* OR breakthrough* OR impact* OR emerg* OR innovat* OR novelt* OR disrupt* OR understand* OR discover* OR advances OR advancement* OR paradigm* OR productiv* OR challeng* OR opportunit* OR benefit* OR risk*) AND (scien* OR research* OR academi* OR scholar* OR studies OR technolog* OR biotechnolog* OR medic* OR health* OR sociolog* OR humanit* OR economics OR physics OR chemistry OR nanotechnolog* OR "climate change" OR robotics) AND (review* OR overview* OR survey* OR reflection* OR analys* OR outline*))
```

Retrieved all results for screening. Papers retrieved: 596. Imported into the same library as the Scopus search. 331 references imported, 265 duplicates automatically discarded.

The ACM Guide to Computing Literature, 02/08/2023, searched by MK

```
Title:(("artificial* intelligen*" OR ai OR "machine learning" OR "deep learning" OR "neural network*" OR "convolutional network*") AND (trend* OR forecast* OR foresight* OR vision* OR strateg* OR breakthrough* OR impact* OR emerg* OR innovat* OR novelt* OR disrupt* OR understand* OR discover* OR advances OR advancement* OR paradigm* OR productiv* OR challeng* OR opportunit* OR benefit* OR risk*) AND (scien* OR research* OR academi* OR scholar* OR studies OR technolog* OR biotechnolog* OR medic* OR health* OR sociolog* OR humanit* OR economics OR physics OR chemistry OR nanotechnolog* OR "climate change" OR robotics) AND (review* OR overview* OR survey* OR reflection* OR analys* OR outline*))
```

Retrieved all results for screening. Papers retrieved: 112. Imported into the same library as the Scopus and WoS searches. 77 references imported, 35 duplicates automatically discarded.

Web of Science, 19/07/2023, searched by AW

```
TITLE only:
Artificial intelligence OR AI OR machine learning OR deep learning OR neural
AND Scien* OR Technolog* OR Biotechnolog* OR medic* OR material OR social
AND Emerging OR innovat* OR novelty OR disruptive OR understanding OR discovery OR
advances OR paradigm* OR productiv* OR challenges OR impact
AND Review
```

Papers retrieved: 88. Scoping search only (search strategy under development).

Scopus, 25/07/2023, searched by FR

```
TITLE only: "artificial intelligence" OR ai OR neural OR "machine learning" OR "deep learning") AND scien* OR technolog* OR research OR academi* OR scholar* AND emerg* OR innovat* OR discover* OR challeng* OR impact* OR understand* OR disrupt* OR novelty AND Review AND 2022 AND 2023
```

Papers retrieved: 78

7 reviews deemed relevant. Full list of retrieved papers and those deemed relevant are available in EndNote file.

The ACM Guide to Computing Literature, 01/08/2023, searched by MK

```
Title:(("artificial intelligence" OR ai OR "machine learning" OR "deep learning" OR "neural network*" OR"convolutional network*") AND (scien* OR research OR academi* OR scholar* OR technolog* OR biotechnolog* OR medic* OR health* OR "social science*" OR physics OR chemistry OR nanotechnology OR economics OR "climate change" OR robotics) AND (emerg* OR innovat* OR novelty OR disrupt* OR understand* OR discover* OR advances OR advancement* OR paradigm* OR productiv* OR challeng* OR impact* OR trend* OR opportunit* OR risk* OR foresight OR vision OR strateg* OR breakthrough* OR benefit*) AND (review OR overview))
```

Papers retrieved: 33 (+1 correction). 16 (+1 correction) potentially relevant papers. The majority are healthcare-related.

Overton, 27/07/2023, searched by LE

```
(title: AI OR "artificial intelligence" OR "deep learning" OR "machine learning") AND (title: research OR scien*) AND (trends OR opportunities OR challenges OR impact)
```

Papers retrieved: n/a. 16 useful reports downloaded on a range of relevant topics. Most are strategic overviews, trends etc.

Overton, 01/08/2023, searched by LE

```
title: AI OR "artificial intelligence" AND (foresight OR vision OR strateg*)
```

Papers retrieved: 213. 10 useful reports downloaded.

Overton, 09/08/2023, searched by LE

```
title: AI OR "artificial intelligence" AND title: society
```

Papers retrieved: 75. 2 useful reports downloaded.

Google, 15/08/2023, searched by AW

Specific search for blogs etc. To explore Deep Dive 1 sub-question on most likely AI breakthroughs:

```
"AI development trend* 2023" OR "AI breakthrough* 2023"
```

Results retrieved: 4.

Scopus, 29/08/2023, searched by MK

```
TITLE ( ( "artificial* intelligen*" OR ai OR "machine learning" OR "deep learning" OR "neural network*" OR "convolutional network*" ) AND ( trend* OR foresight* OR breakthrough* OR impact* OR innovat* OR novelt* OR discover* OR advances OR advancement* ) ) AND PUBYEAR > 2020 AND ( LIMIT-TO ( SUBJAREA , "MEDI" ) ) AND ( LIMIT-TO ( DOCTYPE , "re" ) OR LIMIT-TO ( DOCTYPE , "ed" ) OR LIMIT-TO ( DOCTYPE , "ch" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) )
```

Papers retrieved: 287. 3 included papers on drug discovery.

Web of Science, 30/08/2023, searched by MK

```
(TI=("artificial intelligence" or "automation")) AND TI=("systematic reviews"). Limited to >2020 and Review Article, Article, Early Access
```

Papers retrieved: 9. 3 relevant records downloaded.

Web of Science, 30/08/2023, searched by MK

```
(TI=("artificial intelligence" or "automat*")) AND TI=("systematic review*"). Limited to >2020 and Article
```

Papers retrieved: 29. 2 relevant records downloaded.

Web of Science, 31/08/2023, searched by MK

```
(TI=(chatgpt)) AND TI=(research or academic or scientific or writing or publishing). Limited to Review Article, Article, Early Access
```

Papers retrieved: 92. 7 relevant records downloaded.

Web of Science, 31/08/2023, searched by MK

```
(TI=(alphafold)). Limited to >2020 and Review Article
```

Papers retrieved: 12. 4 relevant records downloaded.

Web of Science, 05/09/2023, searched by MK

```
(TI=(biology) AND TI=("artificial* intelligen*" OR ai OR "machine learning" OR "deep learning" OR "neural network*" OR "convolutional network*")). Limited to >2020 and Review Article, Article, Early Access
```

Papers retrieved: 84. 11 relevant records downloaded.

Proquest, 30/08/2023, searched by LE

```
Title (AI OR "artificial intelligence" AND title (research* OR scholar* OR science* OR humanities) AND title (trend* OR challenge* OR development* OR foresight OR future OR innovation*). Last 3 years
```

Papers retrieved: 225. 12 selected.

LibrarySearch, 30/08/2023, searched by LE

```
Title: AI OR "artificial intelligence" AND title: humanities. Since 1/1/2021
```

Papers retrieved: 150. 7 selected.

Proquest, 06/09/2023, searched by LE

```
title(AI OR "artificial intelligence" OR "deep learning" OR "machine learning" OR "neural networks") AND title("social science*")
```

5 selected.

Proquest, 06/09/2023, searched by LE

```
title(AI OR "artificial intelligence" OR "deep learning" OR "machine learning" OR "neural networks") AND title(humanities)
```

7 selected.

Proquest, 06/09/2023, searched by LE

```
title(AI OR "artificial intelligence" OR "deep learning" OR "machine learning" OR "neural networks") AND title("peer review*" OR "research assess*" OR "research eval*" OR "data manag*")
```

6 selected.

Annex 5. Literature search strategies

Proquest, 06/09/2023, searched by LE

```
title(AI OR "artificial intelligence" OR "deep learning" OR "machine learning" OR "neural networks") AND title(author* OR "IP" OR "intellectual property" OR copyright OR plagiar*). Last 12 months.
```

Proquest, 12/09/2023, searched by LE

```
title(data AND manag*) AND title(AI OR "artificial intelligence" OR "deep learning" OR "machine learning" OR "neural networks")
```

4 selected.

Scopus, 18/09/2023, searched by LE

```
(TITLE(AI OR "artificial intelligence" OR "deep learning" OR "machine learning" OR "neural network*")) AND (TITLE(publish* OR writ* OR author*)) AND PUBYEAR > 2020 AND PUBYEAR < 2024 AND NOT (correction OR erratum). Limit to articles.
```

Papers retrieved: 251. 34 selected.

Scopus, 18/09/2023, searched by LE

```
( TITLE ( research OR scien* AND ( evaluat* OR assess* ) ) ) AND ( TITLE ( ai OR "artificial intelligence" OR "machine learning" OR "deep learning" OR "neutral network*" ) ) AND PUBYEAR > 2020 AND PUBYEAR < 2024
```

Google, 10/10/2023, searched by AW

```
allintitle:(("artificial intelligence" OR AI) AND (guideline OR guidelines OR guidance OR recommendations) AND "research"). Limit to 2016-2023
```

Papers retrieved: not stated. 3 selected.

Google Scholar, 10/10/2023, searched by AW

```
allintitle:(("artificial intelligence" OR AI) AND (guideline OR guidelines OR guidance OR recommendations) AND "research"). Limit to 2016-2023
```

Papers retrieved: 36. 3 selected.

European Tools for Innovations Monitoring TIM, 10/10/2023, searched by AW

```
ti:(("artificial intelligence" OR AI) AND (guideline* OR guidance OR recommendations) AND research). Limit to 2016-2023
```

Papers retrieved: 20. 0 selected.

Scopus, 10/10/2023, searched by AW

```
TITLE ( ( ( "artificial intelligence" OR AI ) AND (guideline* OR guidance OR recommendations ) AND research)). Limit to 2016-2023
```

Papers retrieved: 38. 1 selected.

Personal communication and hand-searching of UNESCO website, OECD website, SIENNA codes and guidelines, 10/10/2023, searched by AW

4 selected.

Key area 3 (rapid review and synthesis of results)

Web of Science, 02/11/2023

```
1 TS=((AI OR "artificial intelligence"))
2 TS=((skill* OR competenc* OR literacy))
3 TS=((teach* OR instruct* OR train* OR educat*))
4 TS=((research* or scien*))
5 #1 AND #2 AND #3 AND #4
Limited to 2022 and 2023
```

Number of records: 635.

ERIC via Proquest, 03/11/2023

```
(AI OR "artificial intelligence")
AND (skill* OR competenc* OR literacy)
AND (teach* OR instruct* OR train* OR educat*)
AND (research* or scien*)
Limited to 2022 and 2023
```

Number of records: 210.

Overton, 03/11/2023

```
AI OR "artificial intelligence" AND (training OR skill* OR reskill* OR job* OR work* OR career* OR competenc* OR labour OR labor OR profession* OR talent OR litera*) AND (research* OR scien* OR universit* OR education). Limited to 2020 to 2023. First 20 pages of hits.
```

Over 100 pages of hits. Went through first 20 pages.

Web of Science, 08/11/2023

```
1 TS=(artificial* intelligen* OR ai OR automation)
2 TS=(((job OR labo$r) NEAR/3 (loss* OR market* OR risk OR force)) OR staff cut*
OR redundanc* OR "laid off" OR "lay off*")
3 TS=(scien* OR research* OR academi* OR graduate* OR R&D OR R+D OR RTD)
4 #1 AND #2 AND #3
Limited to 2022 and 2023
```

Number of records: 208.

Google Advanced Search, 14/11/2023

```
'Research*development framework' OR 'research* training' limited to Germany,
Finland, France, Italy or Netherlands and the past 12 months.
```

Top 10 hits for each search string were browsed.

Web browse for AI courses, 07/11/2023 and 14/11/2023:

- ellis/elise
- Coursera
- FutureLearn AI courses
- edX
- Springboard
- MyMOOC

Browsed courses for a sample of relevance as an AI-introductory course for researchers. 200 viewed, 13 selected.

Key area 4 (rapid review and synthesis of results and additional focused questions)

Scopus & Web of Science #1, 20/11/2023, searched by FR

```
"artificial intelligence" AND ( scien* OR innovation OR research ) AND (
responsible* OR uptake OR integrity )
Title only.
```

25 papers retrieved after de-duplication. 11 papers selected after screening.

Scopus & Web of Science #2, 28/11/2023, searched by FR

```
"research exception"
Title & Abstract. 2016-2023
```

20 papers retrieved after de-duplication. 5 papers selected after screening.

Scopus & Web of Science #3, 29/11/2023, searched by FR

```
"responsible science"
Title only. 2016-2023
```

27 papers retrieved after de-duplication.

Google #3, 11/12/2023, searched by FR

```
"artificial intelligence" AND ("responsible science" OR "research exception")
```

Overton #3, 11/12/2023, searched by FR

```
"artificial intelligence" AND "responsible science"
2020-2023
```

Scopus & Web of Science #4, 02/01/2024, searched by FR

```
"research exception" AND ( gdpr OR cdsm OR "data protection" OR "data mining" OR
"text mining" )
Text & Abstract
```

8 papers retrieved after de-duplication. 5 papers selected after screening.

Google #4, 02/01/2024, searched by FR

```
"research exception" gdpr data protection
```

4 papers retrieved after de-duplication. 4 papers selected after screening.

Scopus & Web of Science #5, 03/01/2024, searched by FR

```
( "copyright in the digital single market" OR cdsm ) AND "mining"
```

14 papers retrieved after de-duplication. 14 papers selected after screening.

Google #5, 03/01/2024, searched by FR

```
data text mining exception copyright europe*
```

4 papers retrieved after de-duplication. 4 papers selected after screening.

Scopus & Web of Science #6, 05/01/2024, searched by FR

```
( "open science" OR "open access" ) AND ( access OR reuse OR re-use ) AND
copyright AND ( research OR science ) AND europe*
```

165 papers retrieved after de-duplication. 13 papers selected after screening.

Cardiff University library search, 02/01/2024, searched by AW

```
(AI OR artificial intelligence) AND (global industr* OR tech* giant* OR big tech*
OR large tech*) In 2021 to 2024
```

165 papers retrieved after de-duplication. 1 paper selected after screening.

ABI/INFORM via Proquest, 02/01/2024, searched by AW

```
(AI OR artificial intelligence) AND (global industr* OR tech* giant* OR big tech*
OR large tech*) In 2021 to 2024. Limited to scholarly journals. Noted: Masses in
newspaper articles and magazines (>100)
```

5 papers retrieved after de-duplication. 2 papers selected after screening.

Business Source Premier via EBSCO, 02/01/024, searched by AW

```
(AI OR artificial intelligence) AND (global industr* OR tech* giant* OR big tech*
OR large tech*) In 2021 to 2024. Limited to scholarly journals. Noted: Masses in
newspaper articles and magazines (>100)
```

7 papers retrieved after de-duplication. 1 paper selected after screening.

Cardiff University library search, 02/01/2024, searched by AW

```
(AI OR artificial intelligence) AND (global industr* OR tech* giant* OR big tech*
OR large tech*) AND (significan* OR impact OR reach OR extent OR dominan*) TI/ABS
2021-2024
```

9 papers retrieved after de-duplication. 1 paper selected after screening.

ABI/INFORM via Proquest, 02/01/2024, searched by AW

```
(AI OR artificial intelligence) AND (global industr* OR tech* giant* OR big tech*
OR large tech*) AND (significan* OR impact OR reach OR extent OR dominan*) TI/ABS
2021-2024
```

185 papers retrieved after de-duplication. 5 papers selected after screening.

Business Source Premier via EBSCO, 02/01/024, searched by AW

```
(AI OR artificial intelligence) AND (global industr* OR tech* giant* OR big tech* OR large tech*) AND (significan* OR impact OR reach OR extent OR dominan*) TI/ABS 2021-2024
```

238 papers retrieved after de-duplication. 7 papers selected after screening.

Google Scholar, 04/01/2024, searched by AW

```
((ethic*AND (principles OR standard* OR advice OR rules)) OR code of conduct OR helsinki) AND (human AND (dignity OR rights OR subject*)) AND (EU OR Europe*) 2021-2024
Human rights AND biomedicine AND (EU OR Europe*) 2021-2024
Research AND ethic* AND (EU or Europe*) 2022-2024
```

Browsed 50 most relevant from each search. 2 selected after screening.

Dimensions, 05/01/2024, searched by AW

```
((ethic*AND (principles OR standard* OR advice OR rules)) OR code of conduct OR helsinki) AND (human AND (dignity OR rights OR subject*)) AND (EU OR Europe*) 2021-2024
Human rights AND biomedicine AND (EU OR Europe*) 2021-2024
Research AND ethic* AND (EU or Europe*) 2022-2024
```

Browsed 50 most relevant from each search. 1 selected after screening.

European Sources Online, 04/01/2024, searched by AW

Browsed with terms from above. 1 selected after screening.

Eurlex, 04/01/2024, searched by AW

Searches with terms from above (eg "research ethics" AND human). Complex searches not permitted. 6 selected after screening.

Website of European Group on Ethics in Science and New Technologies, 04/01/2024, searched by AW

Browsed. 3 selected.

Website of European Research Council, 05/01/2024, searched by AW

Browsed. 1 selected.

Links from other relevant publications

References from relevant publications. 2 selected.

LibSearch, 03/01/2024, searched by LE

```
Big tech AI
```

18 selected after screening.

Sources of funding from the private and public sector into AI research and evaluation of large, publicly funded research structures and their outcomes

Research funding

Overton, 18/01/2024, searched by MK

```
("academic research" or "academie" or "university" or "universities" or "public sector") and (ai or "artificial intelligence") and (funding or investment) ≥ 2020
```

44 768 hits. Downloaded relevant results from the first 3 pages.

Overton, 24/01/2024, searched by LE

```
AI OR "artificial intelligence" AND (invest* OR fund*)
```

Known to research team

European Commission (2023). AI in Science Harnessing the power of AI to accelerate discovery and foster innovation.

Research infrastructure

Known to research team

- Characteristics and regional coverage of the European Digital Innovation Hubs network
- Sectorial AI Testing and Experimentation Facilities under the Digital Europe Programme
- Commission Implementing Decision (EU) 2023/1534 of 24 July 2023 selecting the entities forming the initial network of European Digital Innovation Hubs in accordance with Regulation (EU) 2021/694 of the European Parliament and of the Council

Overton, 24/01/2024, searched by LE

```
AI OR "artificial intelligence" AND infrastructure. Last 3 years.
```

Rapid literature searches

In addition, rapid literature searches were conducted on the following:

- current measures to help identify and reduce environmental footprint of ICT use including AI
- international political negotiations about weapons treaties for AI, especially autonomous weapons
- industry collaboration on published papers

Annex 6. Acknowledgements

SAPEA wishes to thank the following people for their valued contributions and support in the production of this report.

Working group

The working group members who wrote this report are listed at the start of this report.

Peer reviewers

- Cristina Blanco Sío, Universidade da Coruña, Spain
- Maja Fjaestad, Karolinska Institutet, Sweden
- Fotis Psomopoulos, Centre for Research and Technology Hellas, Greece
- Héctor Zenil, The Alan Turing Institute; King's College London; Oxford Immune Algorithmics, UK

Experts participating in the evidence-gathering workshops

- Sören Auer, Leibniz University Hannover, Germany
- Peter Bell, Philipps-Universität Marburg, Germany
- Stephane Berghmans, European University Association, Belgium
- Verónica Bolón Canedo, Universidade da Coruña, Spain
- Loup Cellard, Datactivist, France
- Georgia Chalvatzaki, Technical University Darmstadt, Germany
- Fernando Ferreira, Federal University of Rio de Janeiro, Brasil
- Eva Fialová, Czech Academy of Sciences, Czechia
- Lucie Flek, University of Bonn, Germany
- Adina Magda Florea, University POLITEHNICA of Bucharest, Romania
- Mario Fritz, CISPA Helmholtz Center for Information Security, Germany
- Fatemeh Golpayegani, University College Dublin, Ireland
- Mihály Héder, Budapest University of Technology and Economics, Hungary
- Fredrik Heintz, Linköping University, Sweden
- Sam Illingworth, Edinburgh Napier University, UK
- Yannis Ioannidis, National and Kapodistrian University of Athens, Greece
- Gábor Kismihók, TIB Leibniz Information Centre for Science and Technology, Germany
- Tomáš Kozubek, Technical University of Ostrava, Czechia
- Mario Krenn, Max Planck Institute for the Science of Light, Germany
- Sabina Leonelli, University of Exeter, UK
- Victor Maojo, Universidad Politécnica de Madrid, Spain
- Nicolas Miailhe, The Future Society, France

- Nicolas Moës, The Future Society, Belgium
- Gianfranco Pacchioni, University of Milano Bicocca, Italy
- Janet Rafner, Aarhus University, Denmark
- Cecilia Rikap, University College London, UK
- Federica Russo, Utrecht University, The Netherlands
- Ute Schmid, University of Bamberg, Germany
- Thomas Schön, Uppsala University, Sweden
- Michèle Sebag, French National Centre for Scientific Research, France
- Jacob Sherson, Aarhus University, Denmark
- Olga Štěpánková, Czech Technical University in Prague, Czechia
- Simone Stumpf, University of Glasgow, UK
- Mike Teodorescu, University of Washington, USA
- Mike Thelwall, University of Sheffield, UK
- Yonah Welker, Yonah.org /.ai, Switzerland
- Nicole Wheeler, University of Birmingham, UK

Members of the selection committee

- Anna Fabijańska, Lodz University of Technology
- Virginia Dignum, Umeå University, Sweden
- Patrick Maestro, Secretary General of Euro-CASE, France
- Marja Makarow, President of Academia Europaea, member of the SAPEA Board, Finland

SAPEA staff members

- Marie Franquin, Scientific Policy Officer, Euro-CASE
- Rúben Castro, Scientific Policy Officer, FEAM
- Rafael Carrascosa Marzo, Scientific Policy Officer, AE
- Louise Edwards, Scientific Policy Officer, AE
- Rudolf Hielscher, Coordinator, acatech
- Stephany Mazon, Scientific Policy Officer, YASAS
- Céline Tschirhart, Scientific Policy Officer, ALLEA
- Toby Wardman, Head of Communications

Literature reviews and policy mapping

- Kate Bradbury, Cardiff University Library Services, Cardiff University
- Meg Kiseleva, Specialist Unit for Review Evidence, Cardiff University
- Frederico Rocha, European Information Service, Cardiff University
- Alison Weightman, Specialist Unit for Review Evidence, Cardiff University

Annex 6. Acknowledgements

Science writers

- Hubert Brychczyński, Poland
- Sejal Davla, consultant and science writer, Canada

Group of Chief Scientific Advisors to the European Commission

- Nicole Grobert, Chair
- Maarja Kruusmaa, Member
- Alberto Melloni, Member

Science Policy, Advice and Ethics Unit at DG RTD, European Commission

- Ingrid Zegers, team lead
- Jean-François Dechamp, Policy Officer
- Daniela Melandri, Policy Officer
- Gintarė Juškaitė, Blue Book Trainee

scientificadvice.eu
@EUScienceAdvice

Contact us
EC-SAM@ec.europa.eu

Within the Scientific Advice Mechanism, SAPEA is funded by the European Union.
The activities of associated partners Academia Europaea and Cardiff University
are funded by UKRI (grant number 10033786).