European Commission

Research and Innovation

Living guidelines on the **responsible use of generative AI in research**

ERA Forum Stakeholders' document

Third version, May 2026

Living guidelines on the responsible use of generative AI in research

European Commission
Directorate-General for Research and Innovation
Directorate E-Prosperity
Unit E4 - AI in Science & Critical Technologies

To receive news and event alerts on AI in science, sign up here

Email eu-ai-in-science@ec.europa.eu
RTD-PUBLICATIONS@ec.europa.eu

European Commission
B-1049 Brussels

Manuscript completed in May 2026
Third edition

The European Commission shall not be liable for any consequence stemming from the reuse of this publication.

© European Union, 2026

The Commission's reuse policy is implemented under Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39, ELI: http://data.europa.eu/eli/dec/2011/833/oj).
Unless otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/). This means that reuse is allowed, provided appropriate credit is given and any changes are indicated.

For any use or reproduction of elements that are not owned by the European Union, permission may need to be sought directly from the respective rightholders. The European Union does not own the copyright in relation to the following elements:
Cover: © MicroOne #288703015, creativeteam #323412491, skypicsstudio #286372753, Viktoriia #345410470 #479686001, 2020. © Tartila #279519357, © naum #481091871, 2022. © AKrasov #473038224, 2024. Source: Stock.Adobe.com

EUROPEAN COMMISSION

Living guidelines on the **responsible use of generative AI in research**

ERA Forum Stakeholders' document

2026

Directorate-General for Research and Innovation

Table of contents

1. INTRODUCTION 3

2. GUIDELINES ON THE RESPONSIBLE USE OF GENERATIVE AI IN RESEARCH 7

2.1. RECOMMENDATIONS FOR RESEARCHERS 7

2.2. RECOMMENDATIONS FOR RESEARCH ORGANISATIONS 10

2.3. RECOMMENDATIONS FOR RESEARCH FUNDING ORGANISATIONS 11

BACKGROUND 13

RESEARCH INTEGRITY 13

TRUSTWORTHY AI 13

OTHER FRAMEWORKS OF PRINCIPLES 14

SYNTHESIS FOR THE DEVELOPMENT OF EU GUIDELINES 14

REFERENCES 15

ARTICLES AND PAPERS 15

GUIDELINES ANALYSED 16

1. INTRODUCTION

Artificial intelligence (AI) has become an integral part of our daily lives, transforming how we live and work. Over the past few years, there has been a rapid and disruptive acceleration of AI capabilities, driven by significant advances in widespread data availability, computing power and machine learning techniques. Remarkable strides were made in particular in the development of foundation models - AI models trained on extensive amounts of unlabelled data[1]. These advancements have given rise to what is known as 'General purpose AI', capable of performing a wide array of tasks. This includes "generative AI" which can generate various forms of new content (text, code, data, images, music, voice, videos, etc.), usually based on instructions (also known as prompts) provided by the user. The quality of the output produced by these models is such that it may be difficult to distinguish it from human-generated content.

The widespread and rapid uptake of generative AI has drawn a lot of attention and led to wide policy and institutional responses. The EU is taking the global lead with the AI Act, and other international governance efforts are taking place. These include the Hiroshima process[2] led by the G7, the Bletchley Declaration[3] signed after the first AI Safety Summit, the Paris Statement signed after the 2025 AI Action Summit[4] and the Council of Europe Framework Convention on AI[5].

Generative AI and research

Generative AI provides many opportunities for different sectors. However, it also harbours risks, such as the large-scale generation of misinformation, intellectual property and data protection issues and other unethical uses with significant societal and environmental consequences.

Research is one of the sectors that could be most significantly disrupted by generative AI. AI has great potential for accelerating scientific discovery, leading to new research breakthroughs and significant productivity gains, and improving the effectiveness and pace of research and verification processes. Positive examples of the use of these tools by researchers include supporting non-native speakers in producing texts in multiple languages, producing text summaries from different sources across extremely large corpuses quickly, automatically retrieving and contextualising a wide body of knowledge and helping to write code to analyse data[6]. However, the technology also entails the risk of abuse. Some risks are due to the tool's technical limitations, and others have to do with the (intentional or unintentional) use of the tool in ways that erode sound research practices. Other risks for research in Europe could stem from the proprietary nature of some of the tools (for example, lack of openness, fees to access the service, use of input data), the

---

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=COM:2024:28:FIN

[2] https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-ai-systems

[3] https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

[4] Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet

[5] The Framework Convention on Artificial Intelligence - Artificial Intelligence

[6] Results of a survey by Nature, https://doi.org/10.1038/d41586-023-02980-0 and follow up survey https://doi.org/10.1038/d41586-024-03940-y

concentration of ownership or the undesirable transfer of critical technology and intellectual property[7].

The impact of generative AI on research and various aspects of the scientific process calls for reflection, for example, when working with text (summarising papers, brainstorming or exploring ideas, drafting or translating). In many respects, these tools could harm research integrity and raise questions about the ability of current models to combat deceptive scientific practices and misinformation.

Why these guidelines

Different institutions, including universities, research organisations, funding bodies and publishers have issued guidance on how to use these tools appropriately to ensure that benefits of those tools are fully utilised. The proliferation of guidelines[8] and recommendations has created a complex landscape that makes it difficult to decide which guidelines should be followed in a particular context.

For this reason, the European Research Area Forum[9] (composed of European countries and research and innovation stakeholders[10]), decided to develop guidelines on the use of generative AI in research for: funding bodies, research organisations and researchers, both in the public and private research ecosystems.

These guidelines focus on one particular type of AI used in the research process: generative AI. This is an important step to prevent misuse and ensure that generative AI plays a positive role as part of research practices. One of the goals of these guidelines is that the scientific community uses this technology in a responsible manner. Yet, the development of a robust framework for generative AI in scientific research cannot be the sole responsibility of policymakers (at European and national levels). Universities, research organisations, funding bodies, research libraries, learned societies, publishers, research managers, research support staff and researchers at all stages of their careers are essential in shaping the discussion on AI and how it can serve the public interest in research. They should all actively engage in discussions about the responsible and effective deployment of AI applications, promoting awareness and cultivating a responsible use of AI as part of a research culture based on shared values. Rules and recommendations must go hand in hand with a broad engagement of those involved in public and private research, both organisations and individuals, to develop a culture of using generative AI in research appropriately and effectively.

These guidelines intend to set out common directions on the responsible use of generative AI. While non-binding, they should be considered as a supporting tool for researchers, research organisations and research funding bodies, including the ones applying to the European Framework Programme for Research and Innovation. They take into account key principles on research integrity as well as already existing frameworks for the use of AI in general and in research specifically. Users of these guidelines are encouraged to adapt them to their specific contexts and situations, keeping proportionality in mind.

---

[7] Council Recommendation of 23 May 2024 on enhancing research security (ST/9097/2024/INIT)

[8] Different guidelines that have been identified are listed in the References section of this document.

[9] https://european-research-area.ec.europa.eu/era-forum

[10] https://european-research-area.ec.europa.eu/era-history

These guidelines complement and build on the EU AI policy framework, including the AI Act[11]. They complement other policy activities on the impact of AI in science. These include the opinion of the Scientific Advice Mechanism[12] (SAM) on AI and a policy brief[13] published by the European Commission's Directorate-General for Research and Innovation, framing challenges and opportunities.

Key principles

The set of principles framing these guidelines are based on pre-existing relevant frameworks:

- the European Code of Conduct for Research Integrity[14];
- the work and guidelines on trustworthy AI, developed by the High-Level Expert Group on AI[15];

Building on the commonalities of the currently emerging guidelines from various stakeholders, the key principles behind these guidelines for the responsible use of generative AI in research are:

- **Reliability** in ensuring the quality of research, reflected in the design, methodology, analysis and use of resources. This includes aspects related to verifying and reproducing the information produced by the AI for research. It also involves being aware of possible equality and non-discrimination issues in relation to bias and inaccuracies.
- **Honesty** in developing, carrying out, reviewing, reporting and communicating on research transparently, fairly, thoroughly and impartially**.** This principle includes disclosing that generative AI has been used.
- **Respect** for colleagues, research participants, research subjects, society, ecosystems, cultural heritage and the environment. Responsible use of generative AI should consider the limitations of the technology, its environmental impact[16] and its societal effects (bias, diversity, non-discrimination, fairness and prevention of harm). This includes the proper management of information, respect for privacy, confidentiality and intellectual property rights, and proper citation.
- **Accountability** for the research from idea to publication, for its management and organisation, for training, supervision and mentoring, and for its wider societal impacts. This includes responsibility for all output that a researcher produces, underpinned by the notion of human agency and oversight.

---

[11] https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[12] https://scientificadvice.eu/advice/artificial-intelligence-in-science

[13] https://research-and-innovation.ec.europa.eu/document/download/1e2a4c9c-d3f1-43e9-9488-c8152aabf25f_en

[14] European Code of Conduct for Research Integrity Revised Edition 2023

[15] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[16] When assessing the environmental impact of AI, it is important to examine and monitor both its software and hardware life cycles. See *Artificial Intelligence (AI) end-to-end: The environmental impact of the full AI lifecycle needs to be comprehensively assessed – Issue note*.

Next steps

Generative AI may create possibilities and risks that can be hardly anticipated and fully comprehended today. These guidelines are based on the current state of the technology and the surrounding policy landscape. Given the dynamic nature of both, these guidelines will have to adapt and evolve continuously. They will be updated on a regular basis to ensure that they remain a useful resource for researchers and organisations. These guidelines were developed and will continue to be developed collaboratively as part of the ERA Forum activities. Users of these guidelines may provide feedback and suggestions for adaptations in the feedback form[17].

[17] Feedback form for the guidelines: https://ec.europa.eu/eusurvey/runner/feedback_GenAIResearch

2. GUIDELINES ON THE RESPONSIBLE USE OF GENERATIVE AI IN RESEARCH

Consider the use of generative AI by third parties that could be interacting with you/your organisation.

- Researchers and organisations remain mindful that third parties may use generative AI tools. For example, in the context of meetings, note-taking, summarising discussions or creating overviews of documents (that could be confidential). Therefore, they can face similar risks to the ones described in this guidance, namely on confidentiality, data protection or intellectual property rights related to protected or original information.
- Inform third parties about the of use of these kinds of tools and the potential risks, for example by sharing this guidance.
- In your contacts with third parties inform them if you are the one using those AI tools and how you are protecting collected information.

2.1. RECOMMENDATIONS FOR RESEARCHERS

For generative AI to be used in a responsible manner, researchers should:

1. **Remain ultimately responsible for scientific output**.
   - Researchers are accountable for the integrity of the content[18] they produce generated by or with the support of AI tools.
   - Researchers maintain a critical approach to using the output produced by generative AI and are aware of the tools' limitations, such as bias[19], hallucinations[20] and inaccuracies.
   - AI systems are neither authors nor co-authors. Authorship implies agency and responsibility, so it lies with human researchers.

Generative AI models can exhibit limitations that impact both their responses and usability. The most common types are:

- **Training Data Bias:** Biases in the data used to train generative AI models can lead to skewed responses, reflecting inaccuracies or systemic biases within the source material.
- **Prompt Bias[16]**: Models may align their answers with the perceived beliefs or preferences of the user, a phenomenon known as sycophantic behaviour, potentially leading to misleadingly agreeable or biased outputs.
- **Invented Citations and Incorrect Summaries**: Generative AI models may generate plausible sounding but incorrect citations, which can mislead users who rely on those sources for research or verification. Even when the

[18] Content should not be fabricated, falsified or plagiarised. For example, researchers do not use generative AI to falsify, alter or manipulate original research data. More details in Section 3.1 of *The European Code of Conduct for Research Integrity*, ALLEA 2023.

[19] Ethical prompting should be applied. See the Ethical Prompting for Generative AI guidelines of the ETH Zurich for more information. *https://www.research-collection.ethz.ch/handle/20.500.11850/672207,* ETH Zurich 2024. On sycophantic behaviour see for example https://doi.org/10.48550/arXiv.2310.13548

[20] "hallucination" refers to the generation of false, nonsensical, or inaccurate information by large language models (LLMs) or other generative AI systems.

citations are correct (the paper exists, has the proper title...), the AI-generated summary of those papers may be incorrect. Therefore, researchers are responsible for checking all the references and content of the citations and the summaries.

- **Interpretability**: generative AI models operate as "black boxes," making it difficult to understand how specific responses are generated. This opacity underscores the importance of cross validation, especially in automated data analysis, where model responses can significantly impact conclusions.

2. **Use generative AI transparently.**
   - Researchers, to be transparent, detail which generative AI tools have been used substantially[21] in their research processes. When generative AI meaningfully shapes results, researchers transparently note its use according to the guidelines of their journal or standards in their discipline in the methods section (or equivalent) responsibly evaluating the extent of the contribution[22]. This transparency principle also applies to the use of AI to detect or screen bad practices (i.e. detect hallucinations or plagiarism) while keeping human supervision.
   - Researchers take into account the stochastic (random) nature of generative AI tools, which is the tendency to produce different output from the same input. Researchers aim for reproducibility and robustness in their results and conclusions. They disclose or discuss the limitations of generative AI tools used, including possible biases in the generated content, as well as possible mitigation measures.
3. **Pay particular attention to issues related to privacy, confidentiality and intellectual property rights when sharing sensitive or protected information with AI tools.**
   - Researchers remain mindful that generated or uploaded input (text, data, prompts, images, etc.) could be used for other purposes, such as the training of AI models. Therefore, they protect unpublished or sensitive work (such as their own or others' unpublished work) by taking care not to upload it into an external AI system unless there are assurances that the data will not be re-used, e.g., to train future language models or to the untraceable and unverifiable reuse of data.
   - Researchers take care not to provide third parties' personal data to external generative AI systems unless the data subject (individual) has given them their consent and researchers have a clear goal for which the personal data are to be used so compliance with EU data protection rules[23] is ensured[24].
   - Researchers understand the technical, ethical and security implications regarding privacy, confidentiality and intellectual property rights. They check, for example, their institutional guidelines, the privacy options of the tools, who is managing the tool (public or private institutions, companies, etc.), where the tool is running and

[21] In this case, for example, using generative AI as a basic editorial support tool is not a substantial use. However, using it as a method or source, interpreting data analysis, carrying out a literature review, identifying research gaps, formulating research aims, developing hypotheses, etc. could have a substantial impact.

[22] References to the tool could include the name, version, date, etc. and how it was used and affected the research process. If relevant, researchers make the input (prompts) and output available, in line with open science principles. Specifically, wherever possible, final output shall follow the FAIR principles (Findable, Accessible, Interoperable, Reusable) and be itself machine actionable.

[23] https://commission.europa.eu/law/law-topic/data-protection/eu-data-protection-rules_en. See also the First Orientations for ensuring data protection compliance when using Generative AI systems by the European Data Protection Supervisor.

[24] Apart from personal consent there could be other lawful basis that could justify the processing of personal data.

implications for any information uploaded. This could range from closed environments, hosting on a third-party infrastructure with guaranteed privacy, to open internet-accessible platforms.

4. **Respect applicable national, EU and international legislation**. In particular, the output produced by generative AI can be especially sensitive in relation to the protection of intellectual property rights and personal data.
   - Researchers pay attention to the potential for plagiarism (text, code, images, etc.) when using outputs from generative AI. Researchers respect others' authorship and cite their work where appropriate. The output of a generative AI (such as a large language model) may be based on someone else's results and require proper recognition and citation[25].
   - The output produced by generative AI can contain personal data. If this becomes apparent, researchers are responsible for handling any personal data output responsibly and appropriately, and EU data protection rules are to be followed.
5. **Continuously learn how to use generative AI tools properly to maximise their benefits, including by undertaking training.**
   - Generative AI tools are evolving quickly, and new ways to use them are regularly discovered[26]. Researchers stay up to date on the best practices and share them with colleagues and other stakeholders.
   - Researchers aim at minimising the environmental impact[27] of generative AI, by evaluating whether and which AI tool is best suited for the intended task and by using the most effective prompting techniques.
6. **Refrain from using generative AI tools substantially[28] in sensitive activities that could impact other researchers or organisations (for example peer review[29], evaluation of research proposals, etc).**
   - Avoiding the use of generative AI tools eliminates the potential risks of unfair treatment or assessment that may arise from these tools' limitations (such as hallucinations and bias).
   - Moreover, this will safeguard the original unpublished work of fellow researchers from potential exposure or inclusion in an AI model (under the conditions detailed above in the recommendation for researchers #3).

---

[25] The technology is evolving, and generative AI tools will be more capable of providing correct citations. Tools providing citations could be more practical and could be favoured, but the final responsibility on the citation and its correctness remains with the researcher.

[26] See as an example of considerations to follow Nogueira *et al,* 2025 https://doi.org/10.1177/03400352241304121

[27] This can be seen as part of the broader initiatives to reduce the carbon footprint of research, following the principles for sustainable coding and sustainable research (see for example https://doi.org/10.1371/journal.pcbi.1009324 and https://doi.org/10.1371/journal.pcbi.1008148). Preliminary data on the environmental cost of generative AI can be found in the AI Index Report 2023.

[28] In this case, for example, using generative AI to search background info for a review using it to better understand punctually some ideas or concepts is not a substantial use, while delegating the evaluation or the assessment of a paper, drafting peer review report, helping to understand the main elements of a paper, or evaluating a funding proposal is a substantial use.

[29] ChatGPT is transforming peer review — how can we use it responsibly?

2.2. RECOMMENDATIONS FOR RESEARCH ORGANISATIONS

For generative AI to be used in a responsible manner, research organisations should:

1. **Promote, guide and support the responsible use of generative AI in research activities.**
   - Research organisations provide and/or facilitate training for all career levels and disciplines, including for research managers and research support staff, on using generative AI, especially (but not exclusively) on verifying output, maintaining privacy, addressing biases and protecting intellectual property rights and sensitive knowledge.
   - Research organisations promote an atmosphere of trust where researchers are encouraged to transparently disclose the use of generative AI without concerns for adverse effects.[30]
   - Research organisations provide support and guidelines to ensure compliance with ethical and legal requirements (EU data protection rules, protection of intellectual property rights, etc.).
2. **Actively keep track of the evolution and use of generative AI systems within their organisations.**
   - Research organisations remain mindful of the research activities and processes for which they use generative AI to better support its future use[26]. This knowledge can:
     - be used to provide further guidance on using generative AI, help identify training needs and understand what kind of support could be most beneficial;
     - help anticipate and guard against possible misuse and abuse of AI tools;
     - be published and shared with the scientific community.
   - Research organisations analyse the limitations of the technology and tools and provide feedback and recommendations to their researchers.
   - Research organisations keep track of the environmental impact of generative AI within their organisations and promote awareness raising initiatives to help their staff pick the most sustainable option.
3. **Reference or integrate these generative AI guidelines into their general research guidelines for good research practices and ethics.**
   - Using these guidelines as a basis for discussion, research organisations openly consult their research staff and stakeholders on the use of generative AI and related policies.
   - Research organisations apply these guidelines whenever possible. If needed, they could be complemented with specific additional recommendations and/or exceptions that should be published for transparency.
4. **Whenever possible and necessary, implement locally hosted or cloud-based generative AI tools that they govern themselves[31]. This enables their employees to feed their scientific data into a tool that ensures data protection and confidentiality.**
   - Organisations ensure the appropriate level of cybersecurity of these systems, especially those connected to the internet.

---

[30] Disclosure of the use of generative AI for assistance when writing may lead to a lower quality score during assessment, see for example Li *et al*, 2024, DOI: 10.18653/v1/2024.emnlp-main.279 or Suchikova, Y., & Tsybuliak, N. (2025). The Purity Myth: Why Stigmatizing GAI in Academic Writing Is Harmful. Science Communication, 47(5), 753-761. https://doi.org/10.1177/10755470241313233. Research organisations should try to avoid situations like these ones.

[31] Or governed by trustworthy third parties e.g. partner research organizations, the EU or trusted countries.

2.3. RECOMMENDATIONS FOR RESEARCH FUNDING ORGANISATIONS

Research funding organisations operate in different contexts and follow different mandates and regulations, which may not align with a single set of guidelines. The recommendations below outline a set of measures and good practices for organisations to implement in ways that best fit their specific circumstances and objectives.

For generative AI to be used in a responsible manner, research funding organisations should:

1. **Promote and support the responsible use of generative AI in research.**
   - Research funding organisations design funding instruments that are open, receptive and supportive of the responsible and ethical use of generative AI technologies in research activities.
   - Research funding organizations require funded research and grantees to be in line with existing national, EU and international legislation (where applicable) and good practices for the use of generative AI.
   - Research funding organisations encourage researchers, research managers, research support staff and research organisations to use generative AI ethically and responsibly, including respecting legal and research standards requirements.
2. **Review the use of generative AI in their internal processes. They will lead the way by ensuring they use it transparently and responsibly.**
   - Research funding organisations remain fully responsible for the use of generative AI in their activities, in line with the accountability principle that emphasises responsibility and human oversight.
   - Research funding organisations use generative AI transparently, in particular for activities related with the management of assessment and evaluation, and without compromising their fairness and the confidentiality of the content.[32]. Generative AI could be used to improve internal processes for funders, but it cannot take a role in the assessment or evaluation of the scientific content of projects.
   - When choosing and using generative AI tools, research funding organisations will carefully consider the tool's adherence to standards and challenges in its use, when sharing sensitive or protected information with AI tools. Organisations take into account elements concerning quality, transparency, integrity, data protection, confidentiality and respect for intellectual property rights, as well as its environmental impact.
3. **Request transparency from applicants on their use of generative AI facilitating ways to report it.**
   - Applicants declare if they substantially[33] used generative AI tool(s) to prepare their application.
   - Applicants provide information on the role of generative AI in the research activities proposed and carried out.

---

[32] In activities related with the assessment and evaluation of the scientific content in itself, generative AI should be avoided as described in the recommendation for researchers number 6.

[33] As mentioned in the recommendations for researchers, using generative AI as a basic support tool for authors is not a substantial use. However, interpreting data analysis, carrying out a literature review, identifying research gaps, formulating research aims, developing hypotheses, etc. could have a substantial impact.

- Research funding organisations promote an atmosphere where researchers can disclose the use of generative AI transparently and without concerns of being disadvantaged[24].

4. **Monitor and get actively involved in the fast-evolving generative AI landscape.**
   - Research funding organisations promote and fund training and educational programmes for an ethical and responsible use of AI in scientific research.
5. **When using generative AI as part of the internal processes, take into account the risks related to hidden prompts[34] in content received by third parties, in particular in applications and proposals.**
   - Research funding organisations remain aware that content may contain hidden prompts or invisible instructions intended to influence generative AI tools.
   - Where relevant, they set clear rules prohibiting such manipulation and introduce appropriate safeguards, in particular in the IT systems involved in the processing of that information.

[34] A hidden prompt is an instruction secretly embedded in a document or input for an AI system, without being visible to the human reader. Its purpose is to influence how the AI responds, for example by steering it toward a more favourable assessment. In practice, this can be done through invisible text, formatting tricks, or metadata that a person may not notice but an AI tool may still process.

BACKGROUND

RESEARCH INTEGRITY

The European Code of Conduct for Research Integrity by All European Academies (ALLEA) lays down a set of principles to produce sound research, including ethical aspects. These principles include:

- **reliability** in ensuring the quality of research, reflected in the design, methodology, analysis and use of resources;
- **honesty** in developing, carrying out, reviewing, reporting and communicating on research transparently, fairly, thoroughly and impartially;
- **respect** for colleagues, research participants, research subjects, society, ecosystems, cultural heritage and the environment;
- **accountability** for the research from idea to publication, for its management and organisation, for training, supervision, and mentoring, and for its wider societal impacts.

TRUSTWORTHY AI

The *Ethics Guidelines for Trustworthy AI* by the EU High-Level Expert Group on AI focuses on developing responsible AI systems (rather than using them). This work on trustworthy AI contains a set of ethical principles that were translated into operational requirements for developing AI systems. The requirements could be useful inspiration for drawing up guidelines on the use of generative AI.

The four ethical principles for AI systems are:

1. respect for human autonomy;
2. prevention of harm;
3. fairness;
4. explicability.

These ethical principles were used to develop these seven operational key requirements:

1. human agency and oversight[35];
2. technical robustness and safety;
3. privacy and data governance;
4. transparency;
5. diversity, non-discrimination and fairness;
6. environmental and societal well-being;
7. accountability.

[35] Human agency and oversight includes human-in-the-loop (HITL), human-on-the-loop (HOTL) and human-in-command (HIC) approaches. HITL involves human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable. HOTL involves human intervention during the design cycle of the system and monitoring how the system operates. HIC involves overseeing the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation (Ethics Guidelines for Trustworthy AI).

OTHER FRAMEWORKS OF PRINCIPLES

KU Leuven principles were used in similar exercises and for the development of these guidelines:

- **transparency** on the use of generative AI, depending on the type of use;
- **verification** of the correctness of the generated output with attention to correctly attributing the source;
- **respect** for personal data and confidential information by not entering these on platforms that are not managed on proprietary servers;
- **responsibility** for the correct use of generative AI (primarily help and support) and the published output.

Specialists in AI, generative AI, computer science, psychology and sociology hosted by the Institute for Advanced Study at the University of Amsterdam identified the principles of accountability, transparency and independent oversight (in relation to developers of generative AI). Based on these principles, they proposed living guidelines for the responsible use of generative AI[36].

These principles could be complemented with those from other sources. These include the principles put forward by the Association for Computer Machinery on generative AI Technologies that have a broad scope for those developing, deploying and using the technology. Other sources are the *Values Framework for the Organisation of Research* developed by Science Europe or the *STM[37] principles for ethical and trustworthy AI*.

SYNTHESIS FOR THE DEVELOPMENT OF EU GUIDELINES

The ALLEA principles in the European Code of Conduct for Research Integrity serve as guidance for ethical and responsible research. The same value system applies when using generative AI in research. The ALLEA principles can therefore be used as a basis for these guidelines and be enriched with parts from the other frameworks described in the previous sections.

[36] https://www.nature.com/articles/d41586-023-03266-1

[37] The International Association of Scientific, Technical and Medical Publishers

REFERENCES

ARTICLES AND PAPERS

Birhane, A., Kasirzadeh, A., Leslie, D. & Wachter, S., 'Science in the Age of Large Language Models', *Nature Reviews Physics*, Vol. 5, 2023, pp. 277-280. https://doi.org/10.1038/s42254-023-00581-4

Conroy, G., 'Scientific Sleuths Spot Dishonest ChatGPT Use in Papers', *Nature*, 2023. *https://doi.org/10.1038/d41586-023-02477-w*

Conroy, G., 'How ChatGPT and Other AI Tools Could Disrupt Scientific Publishing', *Nature*, Vol. 622, 2023, pp. 234-236. https://doi.org/10.1038/d41586-023-03144-w

Ganjavi, C., Eppler, M., Pekcan, A., Biedermann, B., Abreu, A., Collins, G., Gill, I. Cacciamani, G., 'Bibliometric Analysis of Publisher and Journal Instructions to Authors on Generative-AI in Academic and Scientific Publishing', *arXiv*, 2023. https://doi.org/10.48550/arXiv.2307.11918

Lenharo, M., 'ChatGPT turns two: how the AI chatbot has changed scientists' lives', *Nature,* Vol. 636, pp. 281-282 (2024). https://doi.org/10.1038/d41586-024-03940-y

Li, Z., Liang, C., Peng, J., Yin, M., 'How Does the Disclosure of AI Assistance Affect the Perceptions of Writing?' *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 4849–4868 (2024). 10.18653/v1/2024.emnlp-main.279

Lorenz, P., Perset, K. & Berryhill, J., 'Initial Policy Considerations for Generative Artificial Intelligence', *OECD Artificial Intelligence Papers*, No. 1, OECD Publishing, Paris, 2023. https://doi.org/10.1787/fae2d1e6-en

Lund, B.D., Wang, T., Mannuru, N.R., Nie, B., Shimray, S. & Wang, Z., '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*, Vol. 74, No. 5, 2023, pp. 570-581. http://dx.doi.org/10.1002/asi.24750

Nogueira, L. A., Moltubakk, S. T., Fagervik, A., & Langfeldt, I. B. 'Cutting through the noise: Assessing tools that employ artificial intelligence.', IFLA Journal, 0(0), 2025 https://doi.org/10.1177/03400352241304121

Parrilla, J.M., 'ChatGPT Use Shows that the Grant-Application System is Broken', *Nature*, 2023. https://doi.org/10.1038/d41586-023-03238-5

International Association of Scientific, Technical and Medical Publishers (STM), 'AI Ethics in Scholarly Communication: STM Best Practice Principles for Ethical, Trustworthy and Human-centric AI', 2021 . https://coilink.org/20.500.12592/hqbzq79

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., ... & Perez, E. (2023). 'Towards understanding sycophancy in language models.' *arXiv* https://doi.org/10.48550/arXiv.2310.13548

Van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL., 'ChatGPT: Five Priorities for Research', *Nature*, Vol. 614, pp. 224-226 (2023) https://doi.org/10.1038/d41586-023-00288-7

Van Noorden R. & Perkel J., 'AI and Science: What 1,600 Researchers Think', *Nature*, Vol. 621, pp. 672-675 (2023) https://doi.org/10.1038/d41586-023-02980-0

United Nations Environment Programme, 'Artificial Intelligence (AI) end-to-end: The Environmental Impact of the Full AI Lifecycle Needs to be Comprehensively Assessed - Issue Note' (2024) https://wedocs.unep.org/handle/20.500.11822/46288

Zou, J., 'ChatGPT is transforming peer review — how can we use it responsibly?', *Nature* Vol. 635, p.10 (2024) https://doi.org/10.1038/d41586-024-03588-8

GUIDELINES ANALYSED

Aalto University:
https://www.aalto.fi/en/services/responsible-use-of-artificial-intelligence-in-the-research-process

All European Academies (ALLEA): https://allea.org/wp-content/uploads/2023/06/European-Code-of-Conduct-Revised-Edition-2023.pdf (DOI: 10.26356/ECOC)

Arizona State University:
https://researchintegrity.asu.edu/export-controls-and-security/artificial-intelligence

Association for Computing Machinery (ACM):
https://www.acm.org/binaries/content/assets/public-policy/ustpc-approved-generative-ai-principles

Australian Government Tertiary Education Quality and Standards Agency (TEQSA): https://www.teqsa.gov.au/guides-resources/higher-education-good-practice-hub/artificial-intelligence

Bockting CL, van Dis EAM, van Rooij R, Zuidema W, Bollen J. Living guidelines for generative AI - why scientists must oversee its use. Nature. 2023 Oct;622(7984):693-696.
https://doi.org/10.1038/d41586-023-03266-1

Committee on Publication Ethics (COPE):
https://publicationethics.org/cope-position-statements/ai-author

Deakin University: https://deakin.libguides.com/generative-AI/ethics-evaluation

Deutsche Forschungsgemeinschaft (DFG):
https://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/stellungnahmen_papiere/2023/230921_statement_executive_committee_ki_ai.pdf

Elsevier: https://www.elsevier.com/about/policies/publishing-ethics

ETH Zurich: https://www.research-collection.ethz.ch/handle/20.500.11850/672207

Harvard University:
https://provost.harvard.edu/guidelines-using-chatgpt-and-other-generative-ai-tools-harvard

International Association of Scientific, Technical and Medical Publishers (STM):
Global Principles for AI

International Association of Scientific, Technical and Medical Publishers (STM): Ethical and practical guidelines for the use of generative AI in the publishing process

International Society for Computational Biology (ISCB): https://www.iscb.org/iscb-policy-statements/iscb-policy-for-acceptable-use-of-large-language-models

KU Leuven: https://research.kuleuven.be/en/integrity-ethics/integrity/practices/genai/genAI

Maastricht University: https://www.maastrichtuniversity.nl/news/guidelines-use-generative-ai

Science Europe: https://www.scienceeurope.org/our-priorities/research-culture/research-values-framework

United Nations Educational, Scientific and Cultural Organization (UNESCO): https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

University of Ljubljana: https://www.uni-lj.si/mma/uien/2023092013475160/?m=1695210471

University of Toronto: https://www.sgs.utoronto.ca/about/guidance-on-the-use-of-generative-artificial-intelligence/

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union.
You can contact this service:

– by freephone: **00 800 6 7 8 9 10 11** (certain operators may charge for these calls),

– at the following standard number: **+32 22999696**,

– via the following form: european-union.europa.eu/contact-eu/write-us_en.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU Publications

You can view or order EU publications at op.europa.eu/en/publications.
Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

EU open data

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

These guidelines are the result of a joint effort between the European Commission and the countries and stakeholders represented in the ERA Forum. They provide simple and actionable recommendations to key actors in the research community on the use of generative AI. These recommendations will bring clarity and reassurance to those using generative AI in their research activities.
These are 'living' guidelines, they will be updated regularly to keep up with the very fast technological development in this area.

Research and Innovation policy

Publications Office
of the European Union