



Verantwortungsvolle Erforschung und Entwicklung von Künstlicher Intelligenz

Hinweise für Wissenschaftlerinnen und
Wissenschaftler der Max-Planck-Gesellschaft

Senatsbeschluss vom 22. November 2024

**Responsible Research and Development of
Artificial Intelligence:** Guidance for Scientists
of the Max Planck Society



Verantwortungsvolle Erforschung und Entwicklung von Künstlicher Intelligenz

HINWEISE FÜR WISSENSCHAFTLERINNEN UND WISSENSCHAFTLER DER MAX-PLANCK-GESELLSCHAFT ¹

Künstliche Intelligenz (KI) eröffnet transformative Möglichkeiten nicht nur für die Grundlagenforschung, sondern auch, um gesellschaftliche und wirtschaftliche Herausforderungen zu bewältigen und Wohlstand zu fördern. Als führende Forschungseinrichtung ist die Max-Planck-Gesellschaft (MPG) in einer hervorragenden Position, menschenzentrierte KI-Systeme durch interdisziplinäre Forschung zu fördern und dadurch das enorme Potenzial der KI zum Wohle der Menschheit zu entwickeln. Die bestehenden Leitlinien der MPG bilden den Rahmen für die Erforschung und Entwicklung vertrauenswürdiger KI. Die hier vorgelegten Hinweise ergänzen bestehende Leitlinien, um eine verantwortungsvolle Erforschung und Entwicklung von Künstlicher Intelligenz in der MPG sicherzustellen.

Die rasanten Fortschritte im Bereich der Künstlichen Intelligenz und ihr transformatives Potenzial haben Debatten **darüber** ausgelöst, welche Eigenschaften diese Systeme aufweisen sollten, um gesellschaftlichen Werten zu entsprechen. Internationale Gremien und globale Foren haben wiederholt zentrale Prinzipien betont, die die verantwortungsvolle Entwicklung von KI leiten sollten.² Mit der Einführung des *European AI Act* wurden dazu Normen rechtlich verankert. Zwar sind KI-Systeme, die ausschließlich für wissenschaftliche Zwecke entwickelt werden, von dieser Regulierung ausgenommen, jedoch wird erwartet, dass dabei ethische und professionelle Standards eingehalten werden.

Die Max-Planck-Gesellschaft leistet wegweisende Beiträge zur internationalen Spitzenforschung auf dem Gebiet der KI. Daher ist eine Reflexion darüber, inwiefern die bestehenden Leitlinien und Regeln der MPG verantwortungsvolle KI-Forschung fördern, essenziell. Die Freiheit der Wissenschaft ist ein fundamentales Gut, das eng mit der Verantwortung verbunden ist, Forschung mit ethischer Integrität zu betreiben. Dies gilt auch für die Erforschung und Entwicklung von KI-Systemen. Im Folgenden wird ein Überblick über den internationalen Diskurs zu verantwortungsvoller KI bereitgestellt. Zudem werden Prinzipien dargelegt, die für die KI-Forschung besonders bedeutsam sind. Es wird erläutert, wie etablierte Leitlinien und Regeln der MPG diese Prinzipien bereits in die Forschungspraxis integrieren.

¹ Der Senat der MPG hat beschlossen, dass das Papier „Verantwortungsvolle Erforschung und Entwicklung von Künstlicher Intelligenz“ alle fünf Jahre auf Aktualität überprüft und gegebenenfalls angepasst wird

² Die wichtigsten Policy-Paper, die diesen Diskurs widerspiegeln, sind im Anhang aufgeführt.



Zum Wohle der Menschheit

Der internationale Dialog über verantwortungsvolle KI wird von einem übergeordneten Ziel geleitet: KI-Systeme sollen dem Wohle der Menschheit dienen. Dieses Leitprinzip, welches das internationale (Soft) Law zu KI prägt, steht in direktem Einklang mit der Mission der MPG. Der erste Grundsatz des Code of Conduct der MPG lautet: „Es ist unser Auftrag, Grundlagenforschung auf höchstem Niveau zum Wohle der gesamten Menschheit zu betreiben“. Das Bestreben, Wissen zum Wohle der Allgemeinheit zu schaffen, spiegelt sich in allen MPG-Leitlinien wider.³

Schutz vor Diskriminierung

In der Diskussion um vertrauenswürdige KI spielt das Potenzial der Systeme, Diskriminierung zu fördern und bestehende Ungleichheiten zu verstärken, eine zentrale Rolle. Da KI-Systeme oft Vorhersagen auf Basis von Trainingsdaten treffen, können sie in den Daten vorhandene Verzerrungen (Biases) oder Unterrepräsentation reproduzieren. Daher sollten Trainingsdaten sorgfältig geprüft werden, um sicherzustellen, dass sie so vielfältig, repräsentativ und frei von Biases sind wie möglich. Damit Forschung dem Wohle der Menschheit dient, ist es von zentraler Bedeutung, dass durch Forschungsaktivitäten keine Gruppen benachteiligt werden. Besonderes Augenmerk gilt marginalisierten und unterrepräsentierten Gruppen, um jegliche Form der Diskriminierung zu vermeiden.⁴

Sicherheit

Der Grundsatz, Forschung zum Wohle der Menschheit zu betreiben, erfordert, dass entwickelte KI-Systeme sicher gestaltet sind. Dieser Anspruch wird auch in internationalen Policy-Dokumenten zu KI hervorgehoben. Sicherheit zu gewährleisten bedeutet Schäden zu verhindern, was im Fall von KI auch Robustheit und Absicherung umfasst. Robustheit bezieht sich auf die Fähigkeit eines Systems, unter sich ändernden Bedingungen, wie z.B. Datenschwankungen, eine konstante Leistung und Genauigkeit aufrechtzuerhalten. Absicherung fokussiert sich auf den Schutz von Systemen und Daten vor unbefugtem Zugriff oder böswilligen Angriffen.

Um möglichen Forschungsrisiken verantwortungsvoll zu begegnen und die Sicherheit von Forschung zu fördern, hat die MPG im Jahr 2010 „*Hinweise und Regeln der Max-Planck-Gesellschaft zum verantwortungsvollen Umgang mit Forschungsfreiheit und Forschungsrisiken*“ veröffentlicht.⁵ Diese schaffen einen Rahmen für Forschende, um die Freiheit ihrer wissenschaftlichen Arbeit bewusst mit etwaigen Forschungsrisiken abzuwägen – sowohl bei der Planung als auch im Verlauf eines Projektes. Dazu ist es nötig, über relevante und wahrscheinliche Risiken der Forschung zu reflektieren. KI-Systeme zeichnen sich durch autonome Fähigkeiten, Lernfähigkeit und flexible Anwendbarkeit in unterschiedlichen Bereichen aus. Deshalb sollten Forschende mögliche negative Auswirkungen der eigenen KI-Forschung kontinuierlich bewerten und entsprechende Maßnahmen zur Risikominimierung ergreifen.

Insgesamt bieten die bestehenden Leitlinien einen Rahmen für den Umgang mit den spezifischen Risiken in der KI-Forschung und stellen Forschenden der MPG eine geeignete Grundlage zur Bewertung ihrer Forschungsprojekte bereit. Die Hauptverantwortung für die Festlegung der ethischen Grenzen ihrer Arbeit liegt aufgrund ihrer Expertise bei den Forschenden. Sie können jedoch Unterstützung von der Institutsleitung, dem Ethikrat oder der Kommission für Ethik sicherheitsrelevanter Forschung der MPG in Anspruch nehmen.

³ Siehe z.B. MPG Code of Conduct, S.2; Hinweise und Regeln der Max-Planck-Gesellschaft zum verantwortungsvollen Umgang mit Forschungsfreiheit und Forschungsrisiken, S. 5, 6, 8; Verhaltensregeln für gute wissenschaftliche Praxis, S. 6, 9, 18.

⁴ Forschung zu KI-Fairness und Bias, die für die Förderung verantwortungsvoller KI-Systeme nötig ist, kann mit diesem Prinzip in Konflikt geraten.

⁵ Diese Leitlinien wurden 2017 aktualisiert.



KI-spezifische Sicherheitsaspekte: Menschliche Aufsicht und menschliche Autonomie

KI-Systeme ermöglichen neuartige Mensch-Maschine-Interaktionen, die außergewöhnliche Chancen bieten, aber auch neue Risiken bergen. Im Diskurs über vertrauenswürdige KI wurden zwei spezifische Risiken hervorgehoben, die in der Reflexion von Forschungsrisiken besondere Beachtung finden sollten: der mögliche Verlust menschlicher Aufsicht und die Gefährdung der menschlichen Autonomie. Forschende, die KI-Systeme untersuchen oder entwickeln, sollten diese Risiken in ihre allgemeine Risikoanalyse einbeziehen.

Menschliche Aufsicht bezieht sich auf die Möglichkeit, in KI-Systeme einzugreifen. Die Bewahrung menschlicher Aufsicht soll sicherstellen, dass KI-Systeme im Einklang mit menschlichen Absichten arbeiten, sodass Menschen die Kontrolle behalten und KI-Outputs bei Bedarf korrigieren können. Dies fördert die menschliche Verantwortung dafür, potenziell riskantes Maschinenverhalten zu begrenzen und ermöglicht, die Konsequenzen von KI-Outputs den Menschen zuzuordnen. Forschende sollten den potenziellen Verlust der menschlichen Kontrolle bewerten, besonders wenn KI-Systeme unbeaufsichtigt arbeiten. Funktionen für menschliche Kontrolle und Intervention, wie Überwachungssysteme oder Notabschaltfunktionen, können dieses Risiko beispielsweise mindern.

Menschliche Autonomie zu schützen bedeutet, Menschen vor unangemessenen Einflüssen oder Einschränkungen durch maschinelle Entscheidungen zu bewahren und sicherzustellen, dass ihre Grundrechte und Freiheiten gewahrt bleiben. KI-Systeme sollten Menschen befähigen, informierte Entscheidungen zu treffen, die mit ihren eigenen Zielen übereinstimmen. Dabei sollen Menschen weder von KI-Systemen untergeordnet, genötigt, getäuscht oder auf ungerechtfertigte Weise manipuliert werden. Wenn Forschende die potenziellen Risiken und den Missbrauch ihrer KI-Forschung und -Entwicklung abschätzen, verdient der Schutz menschlicher Autonomie angesichts der einzigartigen Fähigkeit von KI-Systemen, diese zu bedrohen, besondere Beachtung. Forschende sollten auch Maßnahmen in Betracht ziehen, die sicherstellen, dass Personen, die von KI-Systemen betroffen sind, die Möglichkeit haben, die Ergebnisse dieser Systeme zu beeinflussen.

Transparenz

Neben dem Fokus auf Sicherheit wird im Diskurs über vertrauenswürdige KI auch die Bedeutung der Transparenz von KI-Systemen betont. Auch wenn es oft nicht möglich ist, die Algorithmen eines KI-Systems vollständig zu verstehen, können bei deren Erforschung oder Entwicklung dennoch Maßnahmen ergriffen werden, um die Transparenz zu erhöhen. Dazu gehört die Dokumentation von Daten, Design, Methodik und Prozessen.⁶ Im Kontext von KI bedeutet Transparenz auch, die Unsicherheiten in den Ergebnissen und Vorhersagen der Systeme offenzulegen. Zudem sollte betroffenen Personen ermöglicht werden, die Ergebnisse der KI-Systeme zu hinterfragen.

Zwischen den Prinzipien der Transparenz und der Sicherheit bestehen potenziell Spannungen: zum Beispiel kann die detaillierte Offenlegung von Methoden das Missbrauchsrisiko erhöhen. In Konfliktfällen sollte Sicherheit Vorrang haben, zum Beispiel durch die Anpassung der Kommunikation von Ergebnissen, um Missbrauch zu vermeiden.⁷ Letztlich erfordert Transparenz nicht, jegliche Dokumentation jederzeit öffentlich zugänglich zu machen.⁸ Vielmehr geht es darum, Daten und Methoden nachvollziehbar zu dokumentieren, sowie potenzielle Risiken zu identifizieren und zu erfassen.⁹

Datenschutz und Urheberrecht

Alle geltenden Gesetze sind selbstverständlich einzuhalten. Gegebenenfalls sind dem Datenschutz, dem IT- und dem Urheberrecht besondere Aufmerksamkeit zu widmen. Die MPG bietet Unterstützung durch die *Datenschutzleitlinie* sowie die Rechtsabteilung.

6 Verhaltensregeln für gute wissenschaftliche Praxis, S. 12

7 Hinweise und Regeln der Max-Planck-Gesellschaft für einen verantwortungsvollen Umgang mit Forschungsfreiheit und Forschungsrisiken, S. 10 f.

8 Ebd., S. 9

9 Ebd., S. 12



ANHANG (Anlaufstellen bei weitergehenden Fragen)

Ethikrat der MPG

- Beratung zu ethischen Fragen von Forschungsprojekten.
- Prüft Projekte aller Sektionen und gibt Empfehlungen oder Ethikvoten ab.
- Tritt in der Regel viermal im Jahr zusammen, dringende Anfragen können jedoch auch außerhalb der regulären Termine behandelt werden.
- Kontakt: ethikrat@mpg.de

Ethikkommission für sicherheitsrelevante Forschung (KEF) der MPG

- Beratung in Fragen der Forschungssicherheit und des potenziellen Missbrauchs von Forschung, insbesondere in Fällen von „dual use“ (wenn Forschungsergebnisse für militärische oder terroristische Aktivitäten verwendet werden könnten).
- Unterstützt die Einhaltung der „Hinweise und Regeln der Max-Planck-Gesellschaft für einen verantwortungsvollen Umgang mit Forschungsfreiheit und Forschungsrisiken“.
- Beurteilt, ob ein Projekt mit den MPG-Leitlinien in Einklang steht und gibt Empfehlung ab, wie Projekte angepasst werden können.
- Kontakt: kef@mpg.de

Ombudspersonen

- Anlaufstelle für vertrauliche Mitteilungen bei Verdacht auf wissenschaftliches Fehlverhalten oder bei Konflikten, die die Regeln guter wissenschaftlicher Praxis betreffen.
- Forschende sollten sich zunächst an die Ombudsperson ihres MPI wenden. Bei Bedarf können sie sich aber auch an die Ombudsperson ihrer Sektion wenden.
- Kontakt bei Fragen zum Ombudswesen: ombudswesen@mpg.de

Rechtliche Grenzen und Compliance

- Bei Fragen zu den rechtlichen Grenzen von Forschung können sich Forschende an die zuständige Stelle in der Generalverwaltung wenden:
 - **Datenschutz-, IT-, und Urheberrecht:** Abteilung VII, Referat VIIe: it-recht@gv.mpg.de
 - **Datenschutz: MPG-Datenschutzbeauftragte:r:** datenschutz@mpg.de
 - **IT-Sicherheit:** MPG-Beauftragte:r für IT-Sicherheit: it-sicherheit@mpg.de
 - **Forschungsrecht und Forschungsethik:** Abteilung VII, Referat VII c4



ANHANG (Weiterführende Literatur)

- "Interim Report: Governing AI for Humanity" by UN Advisory Body on AI (December 2023)
- "Stanford HAI Artificial Intelligence Bill of Rights" (January 2022)
- "The OECD AI principles" (adopted in 2019)
- "Ethics Guidelines for Trustworthy AI" by the High Level Expert Group set up by the European Commission (2019)
- "Recommendation on Ethics of Artificial Intelligence" by the UNESCO (2022)
- "Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI systems" by the G7 states (October 2023)
- "The Bletchley Declaration", signed by 28 states after the AI Safety Summit (November 2023)
- "G20 AI Principles", Annex to ministerial Statement on Trade and Digital Economy (2019)
- "Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law" (adopted in 2024)



ENGLISH VERSION

Responsible Research and Development of Artificial Intelligence

GUIDANCE FOR SCIENTISTS OF THE MAX PLANCK SOCIETY¹

Artificial Intelligence (AI) holds immense transformative potential, offering significant opportunities not only for basic research, but also for improving welfare and tackling societal and economic challenges. As a leading research institution, the Max Planck Society (MPG) is in an excellent position to promote human-centered AI systems through interdisciplinary research and thereby develop AI's enormous potential for the benefit of humanity. The existing guidelines of the MPG provide the framework for research and development of trustworthy AI. The guidance presented here complements these guidelines to ensure responsible research and development of AI in the MPG.

The rapid advances in AI technology and its transformative potential have prompted conversations about the socially desirable characteristics of these systems. International bodies and global forums have regularly highlighted key principles that should guide the responsible development of AI.² With the introduction of the European AI Act, norms have been formally recognized. While AI systems developed solely for scientific purpose are exempt from this regulation, there is an expectation that such research and development upholds established ethical and professional standards.

For the Max Planck Society, this is an important moment to reflect on how its existing guidelines and rules foster responsible research practices, especially as it makes seminal contributions to cutting-edge AI research and development at the forefront of science. It is an opportunity to reaffirm the commitment to approaching the freedom of research with ethical integrity, and to illustrate how this extends to research and development of AI systems. This document seeks to offer an overview of the international discourse on responsible AI and place it in the context of MPG's research practices. It outlines key principles, that are particularly relevant to responsible research and development of AI systems. Moreover, it provides guidance on how MPG's established guidelines and rules uphold and integrate these principles into research practices.

¹ The Senate of the MPG has decided that the paper "Responsible Research and Development of Artificial Intelligence" will be reviewed every five years to ensure that it is up to date and adapted if necessary.

² The key policy papers that reflect this discourse are listed in the appendix for reference.



For the Benefit of Humanity

The international discourse on desirable properties of AI is motivated by one overarching goal: AI systems should be developed to serve the benefit of humanity. This guiding principle, which is shaping international (soft) law on AI, aligns directly with the mission of the MPG. At the core of MPG's approach to research is the dedication to advancing knowledge for the greater good. As the first principle of the MPG Code of Conduct states: "Our mission is to conduct cutting-edge, fundamental research for the benefit of all humankind". This commitment is consistently reflected throughout all MPG guidelines.³

Protection against Discrimination

A key concern in the discussion around trustworthy AI is the systems' potential for discrimination and for perpetuating existing inequalities and stereotypes. Since AI systems often make predictions based on training data, there is a risk of reproducing biases or underrepresentation present in the data. This underscores the need for researchers to rigorously evaluate training datasets and to ensure they are, as much as possible, diverse, representative, and free from bias. In line with MPG's mission to ensure that its research serves the benefit of humanity, it is essential to prevent any group from being disadvantaged by research activities, with a particular focus on marginalized and underrepresented groups and communities, thus avoiding any form of discrimination.⁴

Safety

The principle of conducting research for the benefit of humanity implicitly requires that developed AI systems must be safe. This requirement is also explicitly emphasized in international policy documents on AI. Ensuring safety entails preventing harm, which in the case of AI also encompasses robustness and security. Robustness refers to a system's ability to sustain consistent performance and accuracy under changing conditions, such as data fluctuations. Security focuses on protecting systems and data from unauthorized access or malicious attacks.

To address research risks and promote the safety of research outcomes, the MPG has established "*Guidelines and Rules of the Max Planck Society on a Responsible Approach to Freedom of Research and Research Risks*" in 2010.⁵ They provide a framework for researchers to balance research freedom and risks both before the start of and continuously during an ongoing research project. This requires reflection on the relevant and probable risks arising from research. Given AI systems' capacity for autonomous operation and self-learning, and their broad applicability across various contexts, it is essential for researchers to continuously assess whether their AI research could reasonably cause significant harm and to establish appropriate measures to minimize risk.

In essence, these established guidelines offer a framework for addressing the risks specifically associated with AI research, and give MPG scientists an appropriate basis for evaluating their research projects. While researchers, given their expertise, hold primary responsibility for defining the ethical boundaries of their work, they can further seek support from their institute's management, the Ethics Council, and the Committee for Ethics in security relevant Research of the MPG.

³ These guidelines were revised in 2017.

⁴ Research on AI fairness and bias, which is necessary for promoting trustworthy AI systems, may conflict with this principle.

⁵ These guidelines were revised in 2017.



AI-specific Safety Considerations: Human Oversight and Human Autonomy

AI has introduced novel forms of human-machine interaction, offering unprecedented opportunities but also presenting distinct risks. Two AI-specific risks have been highlighted in the debate surrounding trustworthy AI and require special consideration when assessing research safety: the potential loss of human oversight and the risk to human autonomy. Researchers developing or studying AI systems should give careful consideration to these risks as part of their broader risk assessment.

Human oversight refers to the ability to intervene when necessary. Safeguarding human oversight aims to align the behavior of AI systems with human intent, ensuring that humans retain control and are able to stop or correct the outcomes of AI systems, if necessary. It also fosters human responsibility to limit potentially risky machine behavior and ensures accountability of humans for outcomes generated by AI.

Researchers should evaluate the potential loss of human oversight, especially when AI systems operate in an unsupervised manner. To mitigate this risk, they can for instance install mechanisms that enable human control and intervention, such as monitoring systems or emergency shutdown features.

Safeguarding **human autonomy** aims to protect individuals from being unduly influenced or constrained by machine decisions, ensuring their fundamental freedoms and human rights are protected. AI systems should empower individuals to make informed decisions that align with their intended goals. Individuals should be protected from being subordinated, coerced, deceived or unjustifiably manipulated by AI systems.

When scientists assess the potential risks and misuse of their AI research and development, the protection of human autonomy deserves particular consideration, given AI systems' unique capacity to threaten it. Researchers should also consider implementing measures that ensure individuals affected by AI systems have the ability to influence the outcomes of those systems.

Transparency

In addition to the emphasis on safety, the discourse on trustworthy AI highlights the importance of transparency in AI systems. While fully understanding the algorithms of an AI system is often not feasible, researchers can still take meaningful steps to enhance transparency when examining or developing AI systems. This includes documenting data, design, methodology and processes.⁶ In the context of AI, transparency also entails clearly documenting uncertainties in the system's outcomes and predictions, and providing information, so individuals affected by AI systems can understand the basis of the system's behavior and challenge its outcomes.

It is important to acknowledge potential tensions between transparency and safety. In particular, sharing detailed information may increase the risk of misuse. In cases of conflict, prioritizing safety is essential, such as by restricting access to certain components to prevent misuse.⁷ Ultimately, transparency does not require making all information publicly accessible at all times.⁸ Instead, it involves documenting data and methods comprehensively to facilitate auditability, as well as identifying and recording potential risks.⁹

Data Protection and Copyright

Compliance with applicable laws is strictly expected in all research and development. Special attention should be paid to data protection, IT law, and copyright law, where applicable. The MPG provides support through its *Data Protection Guideline* and its Legal Department.

⁶ Rules of conduct for good scientific practice, pp. 40

⁷ Guidelines and Rules of the Max Planck Society on a Responsible Approach to Freedom of Research and Research Risks, p.11

⁸ Ibid, p.11

⁹ Ibid, p.13



APPENDIX (Contact Points for further Questions)

MPG Ethics Council

- Advises on ethical questions regarding research projects.
- Reviews projects across all sections, provides recommendations or ethics vote.
- Typically meets four times a year, but urgent requests can be reviewed outside the regular schedule.
- Contact: ethikrat@mpg.de

MPG Committee for Ethics in security relevant Research (Kommission für Ethik sicherheitsrelevanter Forschung, KEF)

- Advises on research security and potential misuse of research, especially in cases of “dual use” (where research results could be repurposed for military or terrorist activities).
- Supports compliance with the “Guidelines and Rules of the Max Planck Society on a Responsible Approach to Freedom of Research and Research Risks”.
- Assesses if a project aligns with MPG guidelines and offers recommendations on how to adjust research activities.
- Contact: kef@mpg.de

Ombudspersons

- Point of contact for confidential reports in cases of suspected scientific misconduct or conflicts concerning the rules of good scientific practice.
- Researchers should first contact the ombudsperson at their MPI. But if needed, they can reach out to the ombudsperson for their section.
- Contact for questions about the ombudsperson system: ombudswesen@mpg.de

Legal Boundaries and Compliance

- For questions about the legal limits of research, researchers can contact the respective office in the Central Administration:
 - **Data Protection, IT, and Copyright Law:** Contact Abteilung VII, Referat VIIe: it-recht@gv.mpg.de
 - **Data Protection:** Contact the MPG Data Protection Officer: datenschutz@mpg.de
 - **IT Security:** Contact the MPG Officer for IT Security: it-sicherheit@mpg.de
 - **Research Law & Research Ethics:** Contact Abteilung VII, Referat VII c4



APPENDIX (Further Reading)

- "Interim Report: Governing AI for Humanity" by UN Advisory Body on AI (December 2023)
- "Stanford HAI Artificial Intelligence Bill of Rights" (January 2022)
- "The OECD AI principles" (adopted in 2019)
- "Ethics Guidelines for Trustworthy AI" by the High Level Expert Group set up by the European Commission (2019)
- "Recommendation on Ethics of Artificial Intelligence" by the UNESCO (2022)
- "Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI systems" by the G7 states (October 2023)
- "The Bletchley Declaration", signed by 28 states after the AI Safety Summit (November 2023)
- "G20 AI Principles", Annex to ministerial Statement on Trade and Digital Economy (2019)
- "Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law" (adopted in 2024)

HERAUSGEBER

Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V.
Hofgartenstr. 8, 80539 München

PUBLISHER

Max Planck Society for the Advancement of Science
Hofgartenstr. 8, 80539 München

November 2024