SALZBURG
GLOBAL
SEMINAR

CULTURE

The Charter for AI Ethics

Written by Glen Calleja, Mona Gamil, Martin Inthamoussu, Amy Karle, Monica Lopez, Akihiko Mori, and Stephanie Meisl

TABLE OF CONTENTS

Preamble

The advent of Artificial Intelligence (AI) has presented distinct opportunities and challenges for the future of societies. This charter seeks to establish a reference framework of ethical principles addressing issues related to the development, adoption, and application of AI technologies within the Cultural and Creative Sectors (CCS) and beyond.

The document was created at the Salzburg Global Seminar program on **"Creating Futures: Art and AI for Tomorrow's Narratives"** as an initial concept piece and framework for further interviews, investigation, research, and development. This is intended to be an introductory document including additional ethical considerations that arose during the program.

The authors of this charter hail from different socio-political and economic realities including Austria, Argentina, Egypt, Ireland, Japan, Malta, Mexico, the United States, and Uruguay. It represents a collaborative effort among artists, technologists, philosophers, ethicists, writers, and policymakers to address the emerging challenges at the intersection of AI, creative and critical thinking, the arts, culture, and human rights.

The mission of this charter is to elevate both critical and creative approaches without compromising the delicate balance of life on Earth. Whether human or non-human, all species deserve thriving conditions, and this charter endeavors to ensure that AI becomes a catalyst for vibrant cultural expression while safeguarding the rights of those who wish to opt out of AI's integration into multiple fields.

The influence of AI is revolutionary, shaping both tangible and digital landscapes alike. This demands a resilient ethical framework to navigate its use in sensitive, high-stakes domains, ranging from the intricate world of the arts to other far-reaching spheres of impact.

The working definition of the term *"artificial intelligence"* used for the drafting of this document is that used by the **Organization for Economic Co-operation and Development** (OECD); *"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."*[1]

[1] OECD AI Principles." OECD. **https://oecd.ai/en/ai-principles**

Core Principles

Traditional as well as current and significant emerging ethical AI frameworks, such as **UNESCO's Recommendation on the Ethics of Artificial Intelligence**, the **EU AI act**, the **United Nations resolution** on the promotion of "safe, secure, and trustworthy" artificial intelligence (AI) systems, and the **US Blueprint for an AI Bill of Rights** have addressed many of the complexities and potential risks associated with AI, such as cultural plurality, economic variety, and ecological inclusivity. Building on these foundational frameworks and the insights gained at Salzburg Global Seminar, we aim to underscore ethical principles tailored to the current and future possible risks of AI, to better safeguard against potential harms, promote transparency, and ensure that AI systems genuinely align with the various nuances of societal values as the technology continues to evolve.

1. DO NO HARM

This principle emphasizes the development and deployment of AI systems that prioritize human and living systems' safety, well-being, and fairness, ensuring they operate securely, transparently, and without infringing on privacy, safety, or perpetuating bias. This includes preventing the development and proliferation of AI that is designed, implicitly or explicitly, to propagate bias or misinformation/disinformation, and regulate AI research for compliance with ethical standards. It mandates proactive measures to safeguard against potential risks, including robust testing, respecting psychological integrity, and preventing unintended consequences to ensure AI acts as a beneficial tool for living systems, ecosystems, cultures, and society.

2. GOVERNANCE

This principle ensures the responsible development and application of AI technologies by promoting equitable relationships via fair and transparent processes such that the interests of all players, irrespective of their degree of engagement with said technologies, are protected, and their rights to access real and perceived benefits are guaranteed. This can be achieved through the adoption of a comprehensive framework that regulates access to AI literacy and development with a focus on equity between stakeholders including future generations.

The following are some key principles and considerations. Their order does not reflect any assumed hierarchy of priority.

- Ensure responsible development and application of AI by promoting equitable relationships and transparent processes across the AI value chain.
- Protect rights and ensure fair access to benefits for all stakeholders, irrespective of

their engagement level.
- Foster fairness by preventing biases and appreciating cultural uniqueness.
- Maintain decision-making transparency and hold developers accountable for AI outcomes.
- Respect intellectual property; uphold the privacy of users' data, including but not limited to neural rights.
- Promote decentralized power structures to prevent monopolies and ensure a fair distribution of benefits.
- Implement opt-in and opt-out mechanisms to ensure user autonomy in AI technologies. Individuals or groups must explicitly opt into AI services, providing clear, informed consent before AI is activated, applied, or used. They should also have easy, effective means to opt out at any time without penalties or losing access to essential non-AI services while maintaining fairness and accessibility. Transparency about AI operations, data usage, and potential impacts is crucial. AI systems must respect user preferences, promptly excluding their data upon opting out, and ensuring previously collected data is deleted or anonymized.

3. DIVERSITY

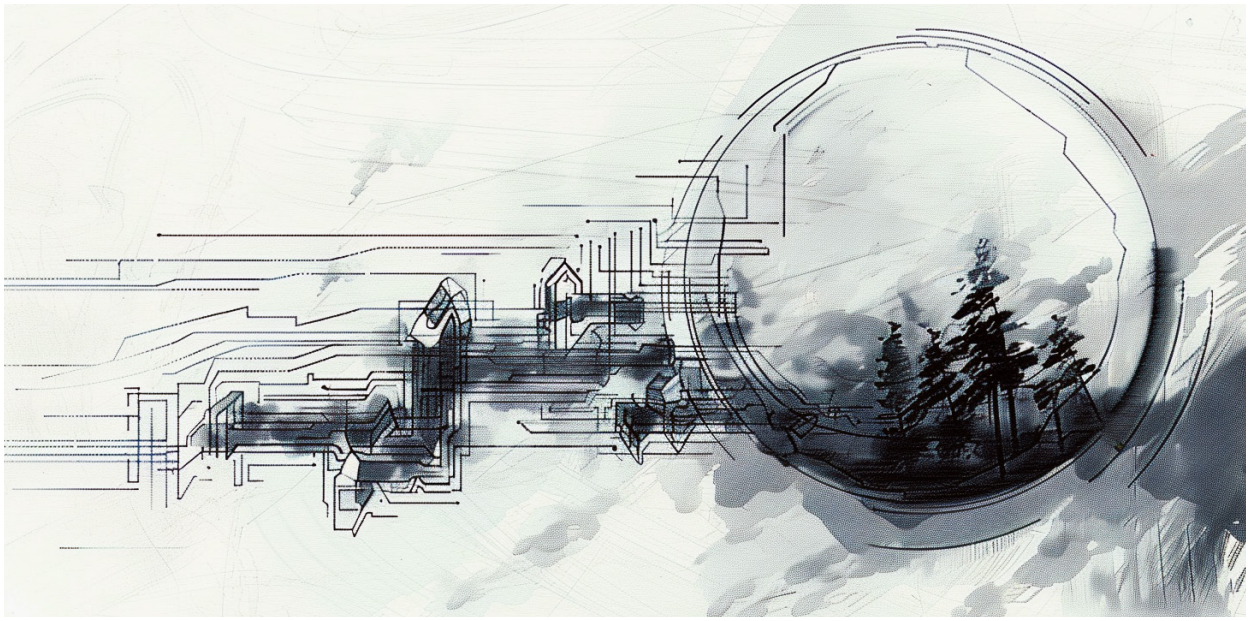
This principle seeks to mitigate the risks of uniformity in decision-making processes and AI outputs, as well as the assumed conformity arising from AI use, proliferation, and imposition. It further seeks to respect and promote a rich pluralism across life systems globally.

- Protect human and non-human species and various forms of intelligence.
- Develop AI in a way that respects all life forms and ecological systems they depend on.
- Counteract the risks of uniform AI outputs and decision-making processes that may cause homogeneity by promoting diversity that reflects the human experience.
- Foster inclusivity and address cultural, ethnic, and gender diversity in AI development.
- Ensure AI technologies are accessible and beneficial across diverse societal segments, providing equitable access.
- Integrate diverse perspectives into AI development, including representation, cultural sensitivity, and routine bias checks.
- Implement technical measures to prevent homogenized outputs and ensure AI respects minority voices and cultural diversity and actively contributes to cultural preservation.
- Develop diverse underlying models and architectures in AI.

4. POSITIVE IMPACTS

The aim is to ensure that AI not only avoids harm but actively contributes to human, ecological, societal, and global well-being, fostering a holistically sustainable and thriving future. Positiveness in AI is defined by its capacity to generate beneficial outcomes that enhance the quality of life and thrive across multiple dimensions and domains. This includes improving human experiences, supporting ecological balance, supporting various cultures and intelligences, advancing societal justice, and promoting global sustainability as well as creative and critical thinking.

- Aim for AI to actively contribute and promote human, ecological, societal, and global well-being and flourishing, promoting a sustainable and thriving future.
- Commit to sustainability and supporting cultures and human thriving by supporting and enhancing living systems and ecosystems, environmental health, and biodiversity.
- Support and enhance creative and critical thinking.
- Design AI systems that enhance and augment natural and human capabilities without replacing human roles.
- Ensure that AI innovations advance the common good, inclusivity, and justice, fostering ethical innovation.
- Promote AI as a tool for cultural enrichment and preservation, ensuring it respects and promotes cultural diversity and heritage. Encourage ethical AI innovation to tackle global challenges, and work towards a more sustainable and thriving future.
- Enhance democratic engagement and community participation through AI, promoting transparency, accountability, and inclusivity in decision-making processes.
- Leverage AI to tap into and support diverse forms of intelligence across cultures and species, recognizing and valuing diverse informational systems.



Opportunities and Challenges

The development and deployment of AI presents unprecedented opportunities, including but not limited to, the advancement of an accelerated analytical tool, efficiency, precision, scalability, innovation, novel forms of expression, enhanced creative processes, and personalization.

Accelerated Analytical Tools: AI technologies offer advanced analytical capabilities that have the potential to enhance the understanding and interpretation of various forms of content, including but not limited to artistic content. Machine learning algorithms can analyze vast amounts of data, identify patterns, and provide valuable insights into user preferences, trends, and cultural influences, allowing professionals to expedite their analytical processes and increase the scale of their work.

Enhanced Precision: Automation technologies excel at executing repetitive tasks with unmatched accuracy, as they are not prone to fatigue or human error. This human-machine collaboration could enable humans to achieve greater precision in their work.

Augmented Creativity: AI has the potential to augment the creative process by leveraging its analytical capabilities to generate novel insights and offer professionals new perspectives and ideas. This human-machine collaborative process has the potential to facilitate innovative approaches to problem-solving and content creation.

AI as a Medium: By integrating AI into their creative process, artists can explore new ways of generating and sketching out ideas, prototyping different aesthetic possibilities, engaging with emergent technologies, and responding to contemporary socio-cultural issues in dynamic and impactful ways.

Increased Personalization: AI has the potential to enhance quality of life by personalizing user experiences through data analysis, tailored content, and dynamic interfaces. Anticipatory design and micro-personalization are among the services that could improve users' quality of life.

Healthcare and Medical Research: AI can revolutionize healthcare by providing precise diagnostics, personalized treatment plans, and accelerating medical research, leading to better patient outcomes and innovative medical solutions.

Education and Learning: AI can offer personalized learning experiences, adaptive learning platforms, and new educational tools that cater to the diverse needs of students, fostering a more inclusive and effective educational environment.

These advancements, however, also bring forth several challenges:

- The deployment of AI systems may contribute to a **concentration of power**, with major corporations and countries from the Global North potentially exacerbating the North-South divide.
- The use of AI can be seen as a form of **digital colonization**, where technologies developed primarily by entities in the Global North may impose their models, values, and priorities on other parts of the world.
- AI-based technologies can lead to **market concentration**, potentially creating monopolies and barriers that hinder competition and the entry of independent players into the AI ecosystem.
- The use of AI in various industries might accelerate **job displacement and loss**, raising concerns about the future of employment and the sustainability of workers' livelihoods.
- AI-based technologies may promote **homogeneity** in cultural outputs, potentially stifling diversity. They may also promote homogeneity of thought and loss of critical and creative thinking.
- AI systems might contribute to **cultural and language erasure**, as their widespread use can marginalize non-dominant cultures and languages.
- The unmonitored use of cultural data to train AI software can lead to instances of systematic cultural appropriation and abuse unless systems are put in place to counteract it. AI-based technologies can perpetuate **biases, stereotypes, and discrimination** if not carefully designed, reflecting and reinforcing existing social prejudices.
- AI technologies may facilitate the spread of **disinformation and misinformation**, potentially eroding democratic processes, undermining trust in institutions, and exacerbating polarization.
- The increasing reliance on AI technologies might lead to **dehumanization**. Overall, the pervasive integration of AI into daily life and decision-making processes risks reducing human interactions, experiences, and emotions to mere data points, potentially eroding the intrinsic value of human dignity and the richness of the human experience and human intelligence.
- The unconsented use of data extracted from cultural expressions such as paintings, photographs, texts, audios or videos, and biometrics, among others, to train AI systems may pose **threats to the moral and material rights of artists**.
- The deployment of AI can lead to enhanced surveillance capabilities, raising significant concerns about **privacy, surveillance, security**, and the potential misuse of personal information.
- It is unclear if and how AI systems making critical decisions in areas such as healthcare, criminal justice, and employment are designed to uphold ethical standards

and ensure health and well-being, fairness, accountability, and transparency. These systems may introduce **bias and danger to health and well-being** without transparency and accountability.

- The **energy consumption and environmental footprint** of AI technologies, particularly those involving large-scale data centers and computational resources, are immense and may be an existential environmental threat.
- AI systems that do not remain under human control pose significant threats to safety and ethical standards. **Unsupervised autonomous behaviors** can lead to unpredictable and potentially harmful outcomes, undermining human oversight and accountability.

AI systems that l**ack transparency and explainability** create substantial risks by obscuring how decisions are made.

Priority Actions

Priority actions are essential to addressing the current ethical gaps highlighted in this charter. Proactive measures such as the continuous engagement of multiple stakeholders, the holistic implementation of core principles delineated above, and the enforcement of emerging regulations are critical to ensuring AI-enabled technologies are responsibly developed and integrated within society. Without prompt intervention by all, we stand to continue exacerbating socioeconomic inequalities. As such, we propose the following set of priority actions:

INCENTIVIZE LOCAL PRODUCTION

Fundamental to the designing and development of fair and equitable AI for everyone demands a recognition of local communities and their unique needs and solutions. Specific ways to collaborate include:

- AI awareness and literacy: Enhance understanding of AI technologies and their ethical implications at the community level through educational initiatives.
- Support grassroots initiatives and community-led and owned engagement across the entire AI lifecycle: : Foster grassroots participation in AI development to ensure local contexts and needs are addressed.
- Intellectual property of developed products and services: Protect the rights and ownership of locally developed AI innovations to encourage sustainable economic growth.

SAFEGUARD HUMAN AUTONOMY AND CURRENT AND EMERGING SOCIAL STRUCTURES

Human autonomy should be prioritized over AI automation. AI-enabled systems should be designed to complement and augment human decision-making capabilities rather than replace them. Moreover, preserving social structures involves mitigating the potential for AI-driven disruptions to power dynamics, the job market, and employment. This entails the following steps:

- Design AI to support, not supplant, human roles: Ensure AI systems enhance human capabilities without rendering human roles obsolete.
- Mitigate AI-driven disruptions: Develop policies and frameworks to address potential job displacement and changes in power dynamics caused by AI deployment.
- Encourage a variety of business models: Ensure that different perspectives and values are incorporated into AI development and integration, leading to more inclusive and ethical outcomes for all.
- Protect the rights of artists/creators whose works have been used to train AI systems.

MITIGATE A CONCENTRATION OF POWER

Develop and enforce regulations and policies that promote decentralized power structures within the AI industry. Encourage a variety of business models including the growth of small and medium-sized enterprises (SMEs) and startups, fostering a competitive and diverse AI ecosystem. This can prevent a few large corporations from monopolizing the industry and ensure that the benefits of AI are widely distributed.

DEVELOP MECHANISMS FOR ONGOING SYSTEM MONITORING

Developing mechanisms for ongoing monitoring of AI systems, and the standardization and normalization of such, is crucial to ensuring their ethical development and use and mitigating potential harms. Mechanisms enable the continuous evaluation of AI systems' behavior, performance, and impact on society and its cultures, allowing for timely intervention in the case of ethical violations or unintended consequences as systems change across time. Establishing transparent and accountable monitoring frameworks enhances trust in AI technologies, fosters responsible innovation, and promotes the alignment of AI development with ethical principles and societal values. Specific ways to ensure ongoing monitoring include:

- **Continuous evaluation of AI systems:** Implement standardized protocols for the regular assessment of AI technologies.
- **Timely intervention mechanisms:** Establish rapid response strategies for addressing ethical violations and unintended consequences.
- **Transparent and accountable monitoring frameworks:** Create clear and accessible reporting structures to maintain public trust and promote ethical AI practices.

ENHANCE AND PROMOTE AI UNDERSTANDING AND LITERACY

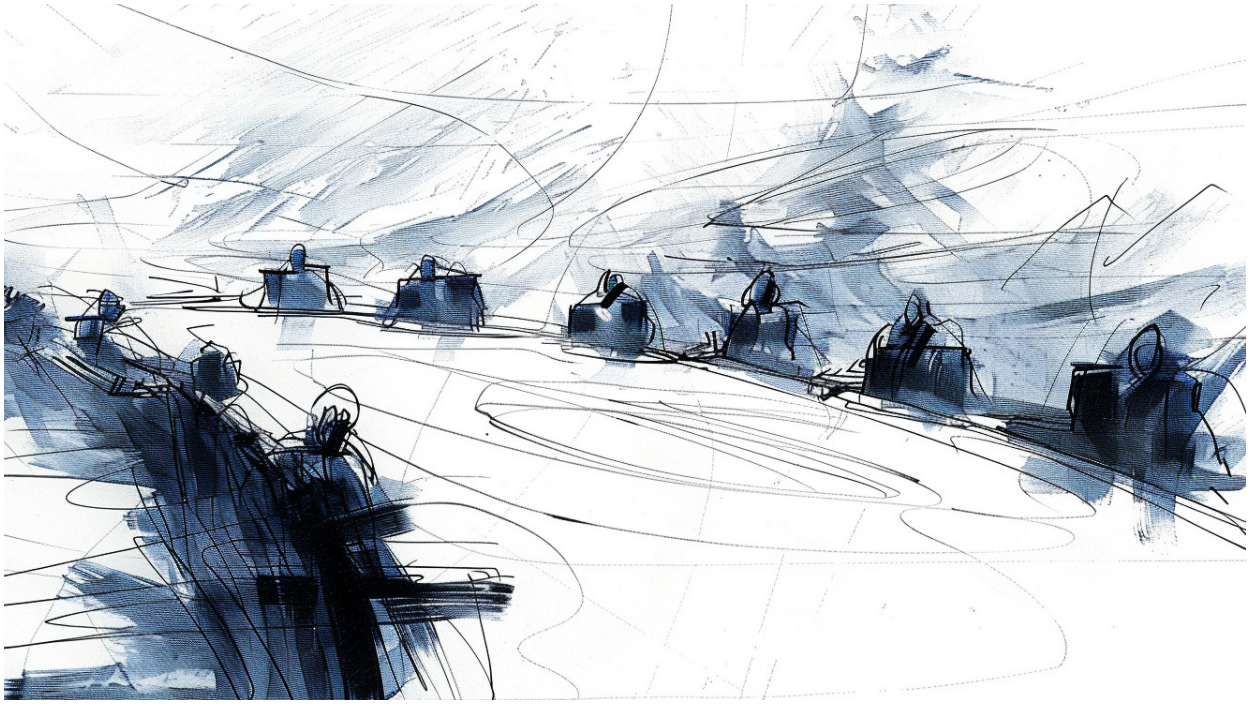
Promote continuous education and literacy programs about AI technologies to ensure that all societal segments are well-informed about AI's capabilities, limitations, and ethical considerations. Develop initiatives to improve public understanding of AI, emphasizing transparency in AI operations and the importance of ethical AI deployment.

FOSTER CONTINUOUS COLLABORATION BETWEEN STAKEHOLDERS

Fostering continuous collaboration between diverse stakeholders is pivotal for the ethical development of AI, as it encourages knowledge gathering, sharing, and cross-pollination, facilitates harmonization of regulatory frameworks, and promotes representative global standards. By working together, governments, industries, academia, and civil society can address common ethical challenges. This collaborative approach not only enhances trust and confidence in AI-enabled technologies but also ensures that benefits are shared equitably and that potential risks are effectively managed at a local and global

scale. Specific actions to support mechanisms for multi-stakeholder engagement and collaboration include:

- **Interdisciplinary forums and conferences:** Organize regular meetings to discuss AI ethics across different sectors.
- **Joint research initiatives:** Promote collaborative projects that address ethical issues in AI development.
- **Global regulatory harmonization efforts:** Work towards creating unified ethical standards and regulations for AI technologies.



Stakeholders

The rapid rise of AI necessitates a robust ethical foundation. This section delineates the stakeholders who are integral to embedding ethical considerations and guidelines into AI: they are categorized into creation and application, policy and regulation, education and advocacy, media, and end users.

A. CREATION AND APPLICATION

The domain of creation and application encompasses a broad range of participants. Multidisciplinary engagement is crucial for fostering an ethical framework that supports innovation while safeguarding the rights and well-being of all involved in the AI lifecycle and ecosystem. This collaborative effort spans various sectors, ensuring that AI technologies are developed and applied in ways that enhance cultural, creative, and critical practices while addressing ethical considerations such as transparency, fairness, accountability, and safety. By integrating diverse perspectives and expertise, this approach promotes a balanced and inclusive AI ecosystem that respects the unique needs and contributions of all stakeholders, ultimately driving positive societal impact, cultural enrichment, and the well-being of all life.

B. POLICY AND REGULATION

Policymakers and regulators, including government bodies and officials, are tasked with formulating and enforcing rules and guidelines for AI deployment. This includes the judicial system and legal professionals addressing the complex legal facets of AI, such as privacy, safety, copyright, and moral rights. International organizations, including but not limited to the United Nations and UNESCO, significantly influence global policies and standards for AI. Civil society organizations, comprising of various non-profits and advocacy groups, play a crucial role in advocating for the responsible use of AI, ensuring that ethical, social, and cultural considerations are meticulously and rigorously addressed. These stakeholders must work together to establish a comprehensive regulatory framework that promotes safe, secure, ethical, and trustworthy AI systems.

C. EDUCATION AND ADVOCACY

Academic and educational stakeholders are fundamental in studying AI's impact and developing new methodologies for its ethical application. Universities and research institutions conduct pivotal research, while educational institutions at all levels prepare students to navigate AI's complexities with a solid ethical foundation. Advocacy groups actively promote ethical standards and protect the rights of those whose information has been utilized or exploited, including AI users, the public, cultures, various life forms, ecosystems, and intelligences. These stakeholders must foster a deep understanding of AI ethics and ensure that future generations are equipped to handle AI's challenges responsibly.

D. MEDIA

Media entities play a critical role by reporting on AI advancements and their implications, shaping public opinion, and disseminating vital information. Media's influence is paramount in how AI integration is perceived and understood by the broader public. It is essential for media to engage in ethical reporting, ensuring accuracy, transparency, and a balanced perspective on AI's benefits and challenges. Media should also advocate for the responsible use of AI and highlight both positive impacts and potential risks, fostering an informed and critical public discourse.

E. END USERS

Users, consumers, and audiences impacted by or engaging with AI tools and AI-generated content hold significant sway. Their feedback drives improvements and informs ethical considerations in AI development, use, and applications. These stakeholders must actively express their concerns, advocate for their rights, and demand transparent and accurate communication about AI. Promoting informed public discourse, end users play a crucial role in ensuring AI technologies are developed and deployed in ways that respect and enhance human dignity, cultural diversity, and societal well-being. By voicing their experiences and expectations, end users contribute to the ongoing refinement and ethical evolution of AI systems.

Engage this Resource

ChatGPT BOT was developed to engage this resource and ask questions. Access it here: https://chatgpt.com/g/g-jIG9JgPbb-ai-and-art-ethics

References

All references were accessed from May 8 to June 21, 2024.

- "AI Act Explorer." Artificial Intelligence Act. https://artificialintelligenceact.eu/ai-act-explorer/
- "AI Ethics Framework for the Intelligence Community 1.0." Intelligence.gov. https://www.intelligence.gov/images/AI/AI_Ethics_Framework_for_the_Intelligence_Community_1.0.pdf
- "Artificial Intelligence Act." https://artificialintelligenceact.eu/
- "Artificial Intelligence." OECD. https://www.oecd.org/digital/artificial-intelligence/
- "Blueprint for an AI Bill of Rights." Whitehouse.gov. October 2022. https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf
- "Checklist for AI Deployment." USAID. https://www.usaid.gov/digital-development/checklist-ai-deployment
- "Creating Futures: Art and AI for Tomorrow's Narratives." Salzburg Global Seminar. https://www.salzburgglobal.org/multi-year-series/culture/pageId/10962
- "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." Whitehouse.gov. October 30, 2023. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/
- "Guest Speaker: Rafael Yuste, Columbia University." Harvard University School of Engineering and Applied Sciences Events. https://events.seas.harvard.edu/event/guest_speaker_rafael_yuste_columbia_university
- IEEE. "Ethically Aligned Design, Version 2." https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf
- NeuroRights Foundation. https://neurorightsfoundation.org/
- "Neurotechnologies, AI, and Human Rights." AI for Good. https://aiforgood.itu.int/event/neurotechnologies-ai-and-human-rights/
- "OECD AI Principles." OECD. https://oecd.ai/en/ai-principles
- "Principles of Artificial Intelligence Ethics for the Intelligence Community." Intelligence.gov. https://www.intelligence.gov/principles-of-artificial-intelligence-ethics-for-the-

intelligence-community https://www.intelligence.gov/images/AI/Principles_of_AI_Ethics_for_the_Intelligence_Community.pdf

- "Recommendation on the Ethics of AI." UNESCO. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics
- "United Nations Unanimously Adopts First Resolution Promoting Safe, Secure, and Responsible AI." The National Law Review. https://natlawreview.com/article/united-nations-unanimously-adopts-first-resolution-promoting-safe-secure-and
- "UN Docs A/78/L.49." United Nations Documents. https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf?token=G7qBhgwJKrdLKDQrv5&fe=true
- UN Resolution "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development": https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf
- UNESCO. "Forum Ethics AI." https://www.unesco.org/en/forum-ethics-ai
- UNESCO. "Ethics of AI." https://www.unesco.org/ethics-ai/en
- UNESCO. "Recommendation on the Ethics of AI" https://unesdoc.unesco.org/ark:/48223/pf0000381137
- "Responsible Business Conduct and Artificial Intelligence." OECD. https://mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf
- HOME | Índice Latinoamericano de Inteligencia Artificial (indicelatam.cl)
- "WHO Releases AI Ethics and Governance Guidance for Large Multi-Modal Models." WHO. January 18, 2024. https://www.who.int/news/item/18-01-2024-who-releases-ai-ethics-and-governance-guidance-for-large-multi-modal-models
- "WHO Outlines Considerations for Regulation of Artificial Intelligence for Health." WHO. October 19, 2023. https://www.who.int/news/item/19-10-2023-who-outlines-considerations-for-regulation-of-artificial-intelligence-for-health