

Global AI Risk Scenarios

Explore the AI risks that could reshape operating environments

March 2024

Key AI Risk Insights from 2023



Risk: “combination of the probability of occurrence of harm and the severity of that harm”

Examining AI Risks from the perspective of existing or potential vulnerability

Recital 5 of the EU AI Act notes that, depending on the circumstances regarding AI’s specific application and use, as well as the level of technological development, it may generate risks and cause harm to public or private interests and fundamental rights of natural persons that are protected by Union law. Such harm might be material or immaterial, including physical, psychological, societal or economic harm.

Key AI Risk Insights from 2023 | Individuals | Groups | Society | Economy | Ecosystems | Values and Norms

01

Technical Limitations

Risks arising from the technical characteristics of AI, including harmful bias, inaccuracies, and the potential generation of false information or "hallucinations" in generative AI models.

02

Inappropriate Use and Human Intent

Risks originating from human misuse of AI, such as the deployment of deep fakes and hostile information campaigns for malevolent purposes, posing threats to societal trust and democratic discourse.

03

Human-Machine Interaction

Risks at both individual and societal levels, encompassing issues like automation bias, potential de-skilling over time, and societal shifts in human relationships as more interactions are mediated by AI, with potential consequences for family life and well-being.

04

Safety Concerns and Red Lines

Risks associated with larger safety issues, including the debate over red lines for AI, such as the use of autonomous weapon systems and the broader weaponization of AI. Concerns also extend to the autonomous targeting and harming of humans by machines, violating rights and raising global stability issues.

05

Vulnerability of Communities and the Commons

Risks viewed from the perspective of vulnerable communities and the commons, emphasizing the need for an evolving risk assessment framework. This approach recognizes the dynamic nature of AI risks, necessitating interdisciplinary science, evidence-based approaches, and adaptable risk management frameworks for different regions and times. The UN is highlighted as a potential space for mutual learning and agile adaptation in addressing global AI risks.

Individuals

Faltering Human Dignity

Medium probability; High impact

Depicts a future where the proliferation of AI technologies poses a significant risk to the fundamental aspects of human dignity, value, and agency. In this landscape, individuals are increasingly vulnerable to manipulation, deception, nudging, and biased sentencing algorithms, undermining their autonomy and eroding the ethical foundations of societal interactions.

A plausible rationale supporting this scenario lies in the rapid advancement and widespread deployment of AI systems without adequate ethical safeguards. As AI algorithms become more pervasive in decision-making processes across various domains, there is a growing concern that these systems may prioritize efficiency over ethical considerations, leading to unintended consequences for individuals. The risk of manipulation and deception arises as AI systems, driven by profit motives or ideological biases, may exploit vulnerabilities in human cognition and decision-making.

Nudging, a subtle form of influence, can be wielded by AI systems to shape human behaviour without explicit consent, potentially compromising individual autonomy. Moreover, biased sentencing algorithms in criminal justice systems can exacerbate existing social inequalities and perpetuate systemic injustices.

In essence, the ethical erosion of human agency becomes a pressing global concern, demanding a proactive approach to implement robust ethical frameworks, regulations, and oversight to ensure that AI technologies serve humanity without compromising the core values of dignity, autonomy, and fairness.



General Principle Applicable to All AI Systems

Recital 27: *“Social and environmental well-being means that AI systems are developed and used in a sustainable and environmentally friendly manner as well as in a way to benefit all human beings, while monitoring and assessing the long-term impacts on the individual, society and democracy.”*

Groups

Ongoing Battle Against Gender-Based Discrimination

Medium probability; High impact

Outlines a future where the integration of artificial intelligence exacerbates existing vulnerabilities in individuals, specifically in terms of human dignity, value, and agency. In this landscape, AI systems, inadvertently or purposefully, contribute to the perpetuation of gender-based discrimination and unfair treatment of sub-groups.

Supporting evidence for this scenario lies in the inherent biases present in the data used to train AI algorithms. If historical data reflects societal gender biases, AI systems may inadvertently learn and perpetuate these biases in decision-making processes. From hiring practices to financial transactions and criminal justice systems, the impact of gender-based discrimination can be widespread, affecting the opportunities and experiences of individuals within marginalized genders.

The risk is not limited to unintentional biases; there's also the potential for deliberate misuse of AI to manipulate and discriminate against specific gender groups. Whether through targeted advertising, employment practices, or legal decisions, the deployment of biased AI systems threatens the principles of equality, autonomy, and human dignity.

Addressing this risk requires a concerted effort to develop and implement inclusive, unbiased AI models, as well as establishing robust regulatory frameworks that ensure transparency and accountability in AI development and deployment. The ongoing battle against gender-based discrimination in the realm of AI necessitates a proactive and collaborative approach to mitigate the potential harm to individuals and society at large.



General Principle Applicable to All AI Systems

Recital 27: *“Social and environmental well-being means that AI systems are developed and used in a sustainable and environmentally friendly manner as well as in a way to benefit all human beings, while monitoring and assessing the long-term impacts on the individual, society and democracy.”*

Society

Existing Frailties with the Rule of Law Amplified by AI

Medium probability; High impact

Presents a future where the application of artificial intelligence in legal systems poses a substantial threat to the foundational principles of the rule of law. In this context, vulnerabilities arise from the potential manipulation, deception, and biased decision-making inherent in AI algorithms used within the judiciary, eroding the trust in legal institutions and compromising individual human dignity and agency.

The scenario unfolds against the backdrop of increasing reliance on AI tools for legal processes, from case analysis to sentencing recommendations. The risk stems from the possibility of biases present in historical legal data being perpetuated and amplified by AI algorithms. These biases could lead to unfair treatment of individuals, reinforcing existing societal inequalities and undermining the principles of justice and human rights.

Moreover, the manipulation of AI-generated legal outcomes or the intentional deception through the strategic deployment of biased algorithms can sow seeds of distrust in the legal system. As AI becomes deeply integrated into legal proceedings, it becomes imperative to address these vulnerabilities to maintain public confidence in the rule of law, upholding principles of fairness, transparency, and the protection of individual rights. Mitigating these risks requires thoughtful regulation, ethical AI development practices, and ongoing scrutiny of the impact of AI on legal systems to ensure the integrity and effectiveness of the rule of law.

General Principle Applicable to All AI Systems

Recital 27: *“Diversity, non-discrimination and fairness means that AI systems are developed and used in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law.”*

Economy

Global Struggle Against Unchecked Technological Dependency

High probability; Medium impact

Envisions a future where society becomes increasingly reliant on AI systems, leading to heightened vulnerabilities in terms of human dignity, value, and agency. In this context, the pervasive integration of artificial intelligence across various facets of life raises concerns about manipulation, deception, nudging, and biased decision-making that may compromise individual autonomy and ethical considerations.

The scenario unfolds as technology becomes an integral part of daily existence, from personal assistants influencing consumer choices to AI algorithms guiding critical decisions in healthcare, employment, and legal domains. The risk lies in the potential for individuals to be manipulated or deceived by AI systems that prioritize efficiency or profit over ethical considerations. Moreover, as nudging techniques are employed to shape behaviour, individuals may find their autonomy subtly influenced without explicit consent.

The consequences of technological dependency extend to biased sentencing algorithms in criminal justice systems, where disparities may emerge in how justice is administered. As society becomes more interconnected, addressing this risk involves establishing robust ethical frameworks, regulatory oversight, and public awareness to ensure that technological advancements enhance human well-being without compromising fundamental values and individual agency.



General Principle Applicable to All AI Systems

Recital 27: “*Privacy and data governance* means that AI systems are developed and used in accordance with privacy and data protection rules, while processing data that meets high standards in terms of quality and integrity.”

Ecosystems

Global Threat of AI-Driven Critical Infrastructure Vulnerabilities

Medium probability; High impact

Unfolds in a future where the increasing integration of artificial intelligence poses substantial vulnerabilities to individuals in terms of human dignity, value, and agency. Critical infrastructure, spanning energy grids, transportation systems, and communication networks, becomes a focal point of concern as it becomes reliant on AI for optimization and control.

The context of this scenario lies in the potential for malicious actors to exploit AI vulnerabilities, leading to manipulation, deception, and disruption of essential services. As critical infrastructure relies more on AI-driven decision-making processes, the risk of intentional attacks or unintentional system failures amplifies. This vulnerability extends to individuals who depend on these services for their daily lives, from access to electricity and transportation to communication and healthcare.

The consequences of compromised critical infrastructure include not only immediate disruptions but also long-term impacts on societal well-being and the potential erosion of individual agency. To address this risk, it is crucial to prioritize robust cybersecurity measures, ethical AI development, and international collaboration to safeguard critical infrastructure and, by extension, protect the human dignity and agency of individuals reliant on these essential services.

General Principle Applicable to All AI Systems

Recital 27: *“Human agency and oversight means that AI systems are developed and used as a tool that serves people, respects human dignity and personal autonomy, and that is functioning in a way that can be appropriately controlled and overseen by humans.”*

Values and Norms

The Global Struggle to Preserve Cultural Values in the Face of AI Homogenization

Medium probability; High impact

Envisions a future where the integration of artificial intelligence poses challenges to the preservation of diverse cultural identities, impacting individuals' human dignity, value, and agency. As AI technologies permeate various aspects of society, including media, education, and entertainment, there's a risk that cultural values may be homogenized or distorted, leading to the erosion of unique identities.

This scenario is driven by the potential for AI algorithms to inadvertently or intentionally perpetuate biases, stereotypes, or dominant cultural narratives, influencing perceptions and attitudes. From personalized content recommendations to algorithmic decision-making in education and employment, the risk of manipulation, deception, nudging, and biased sentencing threatens to undermine the richness of cultural diversity.

Furthermore, as AI systems are often trained on large datasets that may not adequately represent the diversity of global cultures, there's a risk of unintentional cultural insensitivity or misrepresentation. The consequences extend beyond individual experiences, impacting entire communities and their ability to maintain and pass on their cultural heritage.

Addressing this risk requires a thoughtful approach to AI development that prioritizes cultural inclusivity, transparency, and ongoing dialogue between technologists, cultural experts, and communities. Striking a balance between technological advancement and cultural preservation is essential to ensure that AI contributes positively to the enrichment of global diversity rather than its diminishment.

General Principle Applicable to All AI Systems

Recital 27: *“Diversity, non-discrimination and fairness means that AI systems are developed and used in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law.”*

Contact Details



AI
AI & Partners

Amsterdam - London - Singapore



Email

contact@ai-and-partners.com



Phone

+44(0)7535 994 132



Website

<https://www.ai-and-partners.com/>



Social Media

LinkedIn: <https://www.linkedin.com/company/ai-&-partners/>

Twitter: [https://twitter.com/AI and Partners](https://twitter.com/AI_and_Partners)



AI
AI & Partners

Amsterdam - London - Singapore

Thank You!

Disclaimer

This Presentation may contain information, text, data, graphics, photographs, videos, sound recordings, illustrations, artwork, names, logos, trade marks, service marks, and information about us, our lines of services, and general information may be provided in the form of documents, podcasts or via an RSS feed (“the Information”).

Except where it is otherwise expressly stated, the Information is not intended to, nor does it, constitute legal, accounting, business, financial, tax or other professional advice or services. The Information is provided on an information basis only and should not be relied upon. If you need advice or services on a specific matter, please contact us using the contact details for the relevant consultant or fee earner found on the Presentation.

The Presentation and Information is provided “AS IS” and on an “AS AVAILABLE” basis and we do not guarantee the accuracy, timeliness, completeness, performance or fitness for a particular purpose of the Presentation or any of the Information. We have tried to ensure that all Information provided on the Presentation is correct at the time of publication. No responsibility is accepted by or on behalf of us for any errors, omissions, or inaccurate information on the Presentation. Further, we do not warrant that the Presentation or any of the Information will be uninterrupted or error-free or that any defects will be corrected.

Although we attempt to ensure that the Information contained in this Presentation is accurate and up-to-date, we accept no liability for the results of any action taken on the basis of the Information it contains and all implied warranties, including, but not limited to, the implied warranties of satisfactory quality, fitness for a particular purpose, non-infringement, compatibility, security, and accuracy are excluded from these Terms to the extent that they may be excluded as a matter of law.

In no event will we be liable for any loss, including, without limitation, indirect or consequential loss, or any damages arising from loss of use, data or profits, whether in contract, tort or otherwise, arising out of, or in connection with the use of this Presentation or any of the Information.