

ZWISCHENRÄUME



POLICY BRIEF Nr. 6/2023
Dezember 2023

Open science diplomacy for an ethical and trustworthy global AI

**Written by our cooperation partner
ZSI-Zentrum für soziale Innovation:**

Katja Mayer

Inhaltsverzeichnis

1 IDENTIFYING THE CHALLENGES IN AI GOVERNANCE 4

2 CURRENT INITIATIVES AND GLOBAL RESPONSES 6

3 THE ROLE OF OPEN SCIENCE DIPLOMACY IN AI..... 7

4 RECOMMENDATIONS 8

Summary

Summary: This policy brief is prepared for science diplomats, science and innovation policymakers, and AI experts, drawing upon insights from the 2023 UN Science Summit event on Science Diplomacy in Times of Global Crises, held at the Austrian Cultural Forum on September 15, 2023. The briefing document advocates for the application of science diplomacy as a means to establish a robust, independent, diverse, and inclusive scientific base for AI globally. It furthermore underscores the critical need to integrate a higher level of scientific expertise, aligned with Open Science principles, into the international dialogue and governance of AI. It accentuates the significant influence of Open Science in formulating AI policies that are not only more effective but also equitable on a worldwide scale. This approach counters the prevailing competition and profit-driven imperative by offering a more comprehensive perspective that combines scientific knowledge, ethical and security considerations, and fosters international collaboration.

Prepared by: Dr. Katja Mayer, Zentrum Soziale Innovation (ZSI), December 2023

1 Identifying the Challenges in AI Governance

Artificial Intelligence (AI) technology, at its core, is a set of tools and systems that use statistics and machine learning to analyse data, identify patterns, and make decisions or predictions, often mimicking human cognitive functions. AI is expected to have a huge transformative potential, already expanding the frontiers of scientific research, from accelerating discoveries to automating large-scale experimentation. For instance in fields like molecular and genomic research, or healthcare, AI already is predicting protein structures, or assisting in early detection of diseases. Whereas AI, especially in its rapidly evolving form of generative AI that is recognised as a general-purpose technology, it has been a constant feature in recent news and public discourse is increasingly dominated by its negative aspects and harms. Concerns range from the use of misaligned and discriminatory AI systems in public services, including biased facial recognition and flawed fraud detection, to the emergence of deepfakes and potential for disinformation. Additionally, issues like copyright and privacy breaches with generative AI, and the resource-intensive nature of commercial large language models powering chatbots, have raised alarms. These models have been criticised for being part of large-scale, under-regulated social experiments, perpetuating labour exploitation and reinforcing the detrimental impacts of platform capitalism, which fosters powerful industry oligopolies controlling markets both vertically and horizontally. Hence, AI's transformative impact is currently confined to a limited number of private sector entities within a few states, leading to an unequal distribution of benefits and harms.

Despite concerted efforts to address these issues through regulatory measures, policymakers are encountering significant hurdles. The challenge lies in translating high-level concepts like “AI risks” or “fairness” into practical, enforceable standards. These concepts are complex, context-dependent, and sometimes technically infeasible, making uniform application across various AI applications difficult. Additionally, the rapid pace of AI innovation contrasts sharply with the slower processes of policy consensus-building and regulation, making AI a moving target. As such, there is a pressing need for adaptable policies that can keep pace with AI advancements. Learning from past regulatory challenges, particularly those encountered in governing social media platforms’ impacts, is vital in shaping effective and responsive AI governance frameworks and harmonising approaches globally.

The ascent of social media's influence highlighted a lack of transparency in how these platforms operate, for example in the ways newsfeeds are curated, contact networks are established, and users become targets for commercial tracking and advertising. This lack of transparency continues to be a concern in the realm of AI. Today, much like in the past, regulators and the broader society primarily depend on scientists and civil society organisations, including digital justice advocacy groups, to analyse and

evaluate the AI supply chain underlying current AI products. This reliance is expected to persist, as no commercial audit system or automated clearance interface can match the depth of evaluation provided by scientific scrutiny. Although commercial AI often utilises an ecosystem of open-source code and, to an extent, open data, there is growing concern that transparency is waning. This trend is particularly evident as AI systems become increasingly large and resource-intensive, necessitating specialised infrastructures such as computer chips or cloud services. Such developments are making it increasingly challenging for publicly funded research systems to access equivalent resources, thereby intensifying their dependence on major tech companies. This disparity raises significant concerns about the ability to maintain independent and thorough oversight of AI technologies. The advancement of AI thus primarily hinges on three key resources: data, computational power, and specialised talent, often supported by manual annotation labour. Presently, this triad is largely accessible only to well-resourced nations, and major industry, resulting in a concentration of influence and control in the AI field. Furthermore, the global scarcity of critical hardware like GPUs, coupled with a shortage of top technical talent in AI, exacerbates this concentration.

Finally, the AI supply chain is deeply entwined with extensive resource extraction, encompassing rare earth elements, water, and significant energy consumption. The full extent of these resource requirements has remained largely under the radar, with only recent investigations beginning to shed light on the magnitude of resources necessary for training and maintaining AI systems. This emerging area of study is crucial, as it uncovers the environmental footprint of AI technologies, which is often overlooked in discussions focused solely on their computational prowess and applications. Understanding and quantifying these resource demands is vital, not only for evaluating the environmental impact of AI but also for developing strategies to minimise this impact, which is another important dimension of global AI governance. The AI supply chain furthermore often involves labour under conditions that reflect broader systemic inequities and exploitative practices reminiscent of those seen in social media platform capitalism. This exploitation is not just in terms of physical labour and the respective psychological effects - e.g. with the moderation or annotation of sensitive content - but also in the harvesting and use of data, where populations may unwittingly become data sources without fair compensation or acknowledgment. Once again, lessons from the lack of governance of social media platforms highlight the need for more equitable AI policies that protect against such exploitation and ensure that the benefits of AI technology are distributed more fairly. These insights are crucial in shaping AI governance that is sensitive to the needs and rights of diverse global populations.

While there are many commendable exceptions, such as AI systems effectively used in personalised medicine or climate change satellite image analysis, the general landscape of AI is currently not synonymous with ethical or trustworthy practices.

Diplomacy has already been instrumental in elevating the discourse on AI governance from national to international arenas, highlighting its significance in global negotiations. However, what is increasingly necessary is the inclusion of science as a key stakeholder in these discussions. The involvement of the scientific community is essential for ensuring that AI governance is not only internationally coordinated but also grounded in rigorous, ethical, and evidence-based practices, bridging the gap between technological advancement and its responsible application.

2 Current Initiatives and Global Responses

The increasing recognition of AI's impact has led to a multi-voiced call for regulation and governance at an international level. The OECD has installed an AI policy observatory. The last G20 AI Summit resulted in a statement emphasising the use of AI for public good, protecting people's rights and safety, and reducing environmental impacts. This statement draws upon existing guidelines and recommendations, such as UNESCO's Recommendations on AI Ethics. The G7's 'Hiroshima AI process' aims to regulate generative AI, while the EU and the US are collaborating on voluntary codes of conduct to avoid regulatory divergence and promote responsible, human-centric AI. AI industry, NGOs and research organisations have formed the multi-sector global "Partnership on AI", which has set up a theory of change for AI related institutions, as well as guidelines for AI deployment, an example of a voluntary code of conduct. Meanwhile the UN Secretary-General is convening a High-level Advisory Body on AI to analyse and recommend strategies for international AI governance, aiming for a globally inclusive, multi stakeholder approach that aligns with human rights and the Sustainable Development Goals.

Numerous countries are enacting AI-related legislation focusing on privacy, consumer protection, non-discrimination, misinformation, and education. The EU's AI Act, which regulates AI with a focus on fundamental rights, employs a risk-based approach, categorising AI into minimal, limited, high risk, and banned applications, with high-risk applications in sectors like education and healthcare, and limited risk applications, such as generative AI based chatbots, requiring transparency. General purpose AI models, or "foundation models" like large language models are required to make their parameters, including the weights, the information on the model architecture, and the information on model usage publicly available. Notably, the Act exempts AI systems used solely for research and innovation, facilitating scientific advancement without the constraints of commercial applications, though compliance becomes necessary when commercial products are developed. Additionally, the Act recognises the potential of openness by exempting AI systems under free and open source licences from compliance, except when they are high-risk. Models below the systemic risk threshold are exempt from transparency and documentation requirements but must comply with copyright law and provide detailed summaries of training content.

The comprehensive AI regulation, while encouraging AI research and embracing the potential of open source, has sparked a broad debate about its open source exemption, which is still to be detailed in the final version of the AI Act to come. The AI sector, rooted in a tradition of open source and data accessibility, illustrates how openness can paradoxically lead to power concentration, especially as major tech companies like Google, Microsoft, and Meta spearhead development and maintenance of key open source tools. TensorFlow and PyTorch, for example, have become standard tools, evolving into crucial infrastructure. This, combined with their powerful computing, storage capabilities, and financial resources, enables these companies to penetrate the market multi-dimensionally, creating a dependency among businesses and users on these gatekeeper platforms, thus centralising power. Amidst scepticism towards self-regulation and voluntary codes of conduct, there's a growing call for independent scientific assessment and oversight, as emphasised by Ursula von der Leyen at the AI safety summit in November 2023. She stressed the necessity of an independent scientific community with access to resources for unbiased AI risk evaluation. The challenge lies in fostering this independence and community across borders in a globally coordinated manner, and integrating these independent scientific voices into global AI governance. This inclusion of diverse stakeholders is crucial for shaping effective and comprehensive AI governance, underlining the vital roles of both science and science diplomacy in these efforts.

3 The Role of Open Science Diplomacy in AI

Open science is the principle of making scientific research and its dissemination as open and accessible as possible, while being as closed as necessary for privacy, confidentiality, and security reasons. For the purpose of this policy brief, it refers to the practice of making artificial intelligence research and development transparent, accessible, and collaborative. This encompasses sharing AI algorithms, datasets, and research findings openly, while ensuring ethical considerations and data privacy are upheld. It fosters international collaboration, inclusivity, and participation, ensuring scientific research and its outputs are accessible to a global audience. This approach can democratise the production of scientific knowledge and supports a more collaborative and participatory scientific community, aligning with global efforts to ensure equitable and sustainable scientific advancement.

In essence, while science diplomacy is about the intersection of science and foreign policy, open science diplomacy goes a step further by infusing this intersection with the principles of openness, inclusivity, and transparency. It's not just about collaboration between nations but also about making the processes and results of these collaborations openly accessible to a wider audience for their benefit. Open science diplomacy plays a pivotal dual role in the realm of AI governance. Firstly, it is

instrumental in creating awareness for the principles of open science among policymakers and stakeholders. This involves advocating for the importance of transparency, collaboration, and accessibility in AI research. By highlighting how open science can lead to more ethical, inclusive, and effective AI solutions, open science diplomacy encourages policymakers to integrate these principles into legislative and regulatory frameworks. This advocacy is crucial in ensuring that laws and policies governing AI are not only technically sound but also ethically grounded and socially responsible.

Secondly, open science diplomacy leverages the very principles it espouses to negotiate and facilitate a global framework for the development of trustworthy AI and the ethical governance of AI. By fostering international collaborations and open exchanges in AI research and policy, open science diplomacy ensures that a diverse range of perspectives and expertise are considered in decision-making processes as well as in the set up of international advisory bodies or governance institutions. This inclusive approach is key in addressing global challenges posed by AI, such as security and privacy concerns, algorithmic bias, digital divide issues, as well as fraud or misinformation. Moreover, through open science diplomacy, nations can come together to establish common open and thus freely reusable standards and practices for AI development and implementation, ensuring that AI advancements are not only harmoniously aligned with global ethical norms and contribute positively to societal progress, but that they are also monitored and critically reflected over time. In essence, open science diplomacy not only champions the cause of open science in AI policy circles but also actively employs its principles to steer the international community towards a more ethical, equitable, and transparent future in AI governance.

4 Recommendations

The urgent need for open science diplomacy in the governance of ethical and trustworthy global AI stems from the rapidly evolving landscape of AI technologies and their profound impact on society. Adopting an open science diplomacy approach for global AI governance involves a multi-faceted strategy focusing on transparency, ethical standards, international collaboration, informed policy-making, capacity building, and fostering innovation, all underpinned by the principles of open science. This approach is crucial for ensuring transparent, inclusive, and collaborative development and use of AI, transcending national borders and sectoral divides. Open science diplomacy facilitates the sharing of AI research, methodologies, and data across nations, and across markets, promoting universal access and preventing knowledge silos. This transparency is essential for building trust, mitigating risks such as bias and ethical lapses, and ensuring that AI technologies align both with global ethical standards and are sensitive to local requirements. Moreover, it encourages diverse, international collaboration, drawing on a wide range of perspectives to

address complex challenges like data privacy, security, and the socio-economic implications of AI. By integrating open science principles into AI governance, open science diplomacy paves the way for more democratic, accountable, and socially beneficial AI advancements, crucial for addressing global challenges and achieving sustainable, equitable innovation.

As we address the challenges in AI governance, the following are essential priorities for open science diplomacy, which science diplomats can support and advocate for, to ensure the responsible and equitable development of AI technologies:

1. Transparency:

Advocate for comprehensive transparency in AI governance, supporting open-source and open science initiatives for greater disclosure in AI processes including data, code, documentation, and evaluations, as well as environmental footprints and ethical considerations.

2. Inclusivity and diversity:

Promote the inclusion of a diverse range of stakeholders in AI governance, facilitating representation from various industry sectors, academia, and user groups. Encourage collaborations between AI researchers and foreign policy frameworks to enhance the dialogue between scientific and diplomatic sectors. Support the integration of diverse geographical, cultural and scientific approaches in AI research and policy development, amplifying the voices of underrepresented groups.

3. AI literacy:

Support initiatives to enhance AI literacy across all levels of society. Encourage the development and dissemination of accessible resources and training programs that cater to diverse audiences, ensuring a broad understanding of AI's capabilities, limitations, and ethical implications. This effort is crucial for empowering informed decision-making and fostering a knowledgeable public that can actively participate in AI-related discourse and development.

4. Awareness:

- Adapting to a Multipolar World: Advocate for the use of open science diplomacy in Europe as a strategic tool in AI, balancing openness with necessary protections. Highlight the need to be vigilant against potential foreign interference in scientific cooperation, while maintaining the principles of open science to be as open as possible.
- Support initiatives that enhance public understanding of AI, addressing sensationalism and misconceptions, and raising awareness about external influences in AI scientific collaboration.

5. Coordination and Collaboration:

- Endorse “regulatory sandboxing” as a means to balance AI innovation and regulation.
- Support the creation and implementation of joint collaborative research infrastructures, such as high performance computing and public sector cloud services and promote equitable access to advanced technologies
- Encourage robust global connections within the AI community, emphasising the essential role of science in decision-making and the integration of a global Open Science framework in AI governance.
- Facilitate international conversations and science communication on AI research and governance issues, help to pool expert knowledge
- Suggest establishing an International Panel on Artificial Intelligence or similar for comprehensive oversight and to monitor the impact of AI on society and on the environment

6. Transdisciplinarity:

Endorse interdisciplinary collaborations, nurturing partnerships across science, policy, diplomacy, and industry. Highlight the importance of cross-disciplinary research and workforce diversity in AI to address global challenges effectively. Highlight research at the intersection of AI, social sciences, and humanities to explore the broader impacts and potentials of AI and promote citizen science or participatory research.

7. Participation:

Advocate for community-driven methodologies in AI to achieve the SDGs. Emphasise the importance of transparency, inclusivity, and ethics in AI deployment, and support grassroots involvement in AI solution development. This approach involves leveraging mobile apps and AI platforms for local data reporting, developing region-specific knowledge bases, and encouraging participatory AI project evaluation, alongside fostering the acknowledgement of citizen generated data.

8. Responsibility:

- Encourage research organisations as well as policy makers to acknowledge their role in disseminating knowledge and countering disinformation. Support transparency in public sector AI applications and promote ethical AI practices.
- Highlight and support open infrastructures and organisations dedicated to open AI research, focusing on their efforts to achieve sustainability. Emphasise the importance of stable and reliable open research platforms in the AI field, and advocate for the necessary backing, whether financial, policy, or community support, to ensure their longevity and effectiveness.

9. Trust:

Foster trust in science by supporting transparent, inclusive, and well-informed AI development. Encourage dialogue to address science scepticism and uphold

democratic values. Support evaluation and critical reflection of effectiveness, impact, and safety of open research practices and participatory methods (e.g. for open model development). Observe international monitoring, evaluation and audit activities, for accountability and compliance developments, and codes of conduct and self-regulation procedures.

Further Resources or Recommended Reading:

- UN AI Advisory Board <https://www.un.org/en/ai-advisory-body>
- UNESCO Open Science Recommendations <https://www.unesco.org/en/open-science>
- OECD AI Policy Observatory <https://oecd.ai/en/>
- AI Index Annual Report (HAI Stanford) <https://aiindex.stanford.edu/report/>
- Mayer, K. (2020). Open Science Diplomacy. In M. Young, T. Flink, & E. Dall (Eds.), Science Diplomacy in the Making: Case based insights from the S4D4C project (pp. 133–215). <https://www.s4d4c.eu/>
- Interim Report- Governing AI for Humanity: [ai_advisory_body_interim_report.pdf \(un.org\)](https://www.un.org/ai_advisory_body_interim_report.pdf)
- Science Diplomacy Event at the Austrian Cultural Forum NY in September 2023 [Science Diplomacy in Times of Global Crises - Symposium - Austrian Cultural Forum New York \(acfnyc.org\)](https://www.acfnyc.org/science-diplomacy-in-times-of-global-crises-symposium-austrian-cultural-forum-new-york)

Disclaimer: The views expressed in the brief are those of the authors and do not necessarily reflect the position of the publishing organisation or its affiliates.

Date of Publication: December 2023

austrian cultural forum^{nyc}

#TechDiplomacyTalks

University for
Continuing
Education Krems

