Safeguarding Democracy in the Age of Artificial Intelligence

Policy Brief

Teo, Sue Anne; Bukovska, Barbora

Safeguarding Democracy in the Age of Artificial Intelligence

Challenges and Opportunities for GIZ

Published by:

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Safeguarding Democracy in the Age of Artificial Intelligence

Challenges and Opportunities for GIZ

April 2025

Authors:
Dr. Sue Anne Teo, Associate Research Fellow, Center for European Policy Studies;
Researcher, Raoul Wallenberg Institute of Human Rights

Barbora Bukovska, Senior Director for Law and Policy at ARTICLE 19,
Global Campaign for Free Expression

The publication "Safeguarding Democracy in the Age of Artificial Intelligence" was commissioned by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, Sector Programme Governance, on behalf of the German Federal Ministry for Economic Cooperation and Development (BMZ).

Acknowledgements: We would like to express our sincere gratitude to the Competence Center 4C30 and the Sector Project International Digital Policy for Sustainable Development for their valuable contributions to this Policy Brief. Their insightful comments, critical feedback, and expertise have greatly enriched the analysis and helped shape the final outcome.

Table of Contents

Executive Summary

Artificial Intelligence (AI) is increasingly affecting every aspect of daily life, from personal decisions to political processes. Its growing influence is reshaping governance, with profound consequences for democratic practices and state-citizen dynamics. While AI technologies can enhance citizen engagement and empower civil society initiatives, they also present significant risks, particularly when used for mass surveillance, disinformation, and repression. As global enthusiasm for AI grows, there is also a risk of prioritizing rapid development, investment, and economic gains over ethical and human rights considerations. An evolving geopolitical landscape – marked by shifting alliances, increasing technological competition, and economic pressures – poses new challenges for the international community.

The policy brief explores this dual nature of AI's impact – its potential to strengthen democratic processes and its capacity to undermine them, especially in authoritarian contexts. It argues that effective governance frameworks – built on transparency, accountability, and human rights protection – are essential to ensuring AI's responsible use. In the rush to embrace AI, it is essential to balance technological advancements with democratic safeguards. Additionally, strategic initiatives aimed at bridging the digital divides and empowering marginalised groups, especially women and girls and other vulnerable groups in authoritarian contexts, are crucial. By emphasising transparency, accountability, and inclusive governance, the brief makes a case for German development cooperation to advocate for responsible AI advancement that does not sacrifice human rights in the pursuit for technological progress. Through targeted programs, German development cooperation can play a pivotal role in ensuring that AI becomes a driver for social good, rather than a tool of oppression.

1 Core narrative, premise and the case for action

'AI' includes a wide range of technologies, applications and techniques, each with different levels of complexity and autonomy. These include machine learning, expert systems, domain-specific algorithms, which leverage techniques such as natural language processing or computer vision. AI can be embedded in hardware such as robotics or implemented as software solutions. AI systems are designed to process and analyse large sets of data to derive actionable insights. Early definitions linked AI to human intelligence, but this approach is problematic due to the complexity of human cognition. Instead, Stuart Russell, Professor of Computer Science at UC Berkeley and leading researcher in artificial intelligence, **defines AI** as a computational agent that makes rational decisions based on inputs from its environment. This aligns with the **European Union's Artificial Intelligence Act (EU AI Act)**, which characterises AI as technology capable of a certain degree of autonomy and data-driven inferences. While AI has advanced capabilities that go beyond earlier technologies, it is not 'magic'. It represents an evolution of data-driven systems. Although AI may exhibit some autonomy, all design and deployment decisions are ultimately and currently still made by humans. At the same time, the functioning of some AI systems, notably those reliant upon deep learning, can be a 'black box', where neither the deployer nor the end user knows how the AI system reached a decision or recommendation due to opacity of its inner workings.

As AI evolves, it is increasingly viewed as a tool for addressing many global challenges, including the climate crisis or healthcare improvements. AI is seen as an enabler of the **Sustainable Development Goals (SDGs) – estimates show that AI can positively impact 134 SDG-targets (79%), likewise, it may impact 35 targets negatively**. At the same time, AI can also negatively impact environmental sustainability, and the lack of transparency from tech companies make it difficult to address these concerns. Moreover, many states and corporations are eager to push forward with AI development, seeing regulations as roadblocks that could slow down innovation. Historically, technology has often raced ahead of the laws meant to guide it, leaving societies to play catch-up. This dynamic has been further intensified by geopolitical shifts and policy decisions, particularly in leading AI-developing countries, where regulatory reluctance and industry lobbying have reinforced a hands-off approach. In this context, policymakers face the challenge of finding a way to encourage AI's benefits while also putting in place the necessary safeguards to protect democratic governance and social cohesion.

Use of AI can lead to both intended and unintended harms in democratic and authoritarian contexts. Importantly, deployment of AI tools can significantly impact democracy, which can be understood as more than periodic elections and voting. It refers to a political system that genuinely represents and serves the interests of its citizens. This broader view requires conditions that support democratic deliberation, strong institutions, independent media, an empowered civil society, and an informed population. At the same time, AI tools to aid in democratic deliberation may help citizens to navigate an increasingly complex world.

In both democratic and authoritarian contexts, use of AI significantly impacts public discourse and governance. In democracies, it can enable disinformation, influencing public trust in information sources and potentially erode confidence in institutions. While it can empower civil society by improving information collection and detecting abuses of power, it also raises ethical concerns related to bias and accountability that threaten human rights. In authoritarian regimes, AI is often misused for mass surveillance and disinformation campaigns, reinforcing state control and deepening power imbalances. This is critical in the light of the erosion of democratic norms and a **15-year consecutive decline in democracy worldwide** (e.g. currently, **71% of the world's population or 5.7 billion people are living in autocracies**, an increase by 48% compared to ten years ago).

Additionally, the growing power of big tech companies in a largely unregulated global space further complicates these challenges. The economic monopolies of these companies translate into political power as they aim to influence and dictate both the shape and implementation of regulation. At the AI Action Summit in Paris in February 2025, **US Vice-President JD Vance called out against the 'excessive regulation' of AI in the EU** for being detrimental to innovation; echoing **similar objections from tech companies**. In the US, the new Trump administration has seen the tech and politics closely linked, with figures like Elon Musk gaining unprecedented power to shape tech-powered infrastructures and challenging long-standing institutions. The jostling of political influence can also be seen through Meta's dismantling of platform protections and Tik Tok's presence in the US.

All these developments are situated against the backdrop of increasing deregulation of technology, including AI, and 'race' to AI-dominance and supremacy. In January, the **US President announced a 500 billion USD venture called 'Stargate**,' a joint-venture between the public and private sector to propel AI innovation in the US. Due to the dominant global presence of platforms and technologies, this confluence of tech and political power means that effects go beyond the US, shaping the trajectory of governance, industrial policies and economic interests worldwide. However, the implications of AI on state and society differ between different country contexts.

2 Contextual analysis of AI

2.1 Authoritarian contexts

In authoritarian regimes, AI-driven tools can be misused to strengthen the regime's grip on power in several ways; for example through:

- **Using AI-driven facial recognition systems for mass surveillance.** For instance, in China, an extensive facial recognition system that **logs nearly every citizen in the country** through a vast network of cameras, is used for mass surveillance, public shaming, and control of behavior. It has been accused of being used to commit atrocities against Uyghur Muslims. **India has at least 170 facial recognition systems**, with 20 currently operational in locations such as Telangana, Delhi, and Maharashtra. Such technologies enable real-time monitoring, disproportionately affecting dissidents and human rights defenders while fostering a climate of fear that leads to self-censorship among the general population. The covert nature of these surveillance tools makes it difficult for people to recognize or resist their deployment, underscoring the broader risks associated with AI in authoritarian context.

- **Monitoring social media and other platforms** to identify and preempt popular discontent movements or to track and suppress dissent. **Research has shown that surveillance is increasing across Africa**, including social media surveillance. Surveillance used to be a labour-intensive exercise, but AI can now automate the process and at the same time, increases the scale of surveillance. This allows authorities to map patterns of dissent and predict potential uprisings. Even strong privacy measures, like end-to-end encryption, can be bypassed by spyware such as Pegasus, which can infiltrate devices without users' knowledge. This kind of surveillance creates a chilling effect, making people hesitant to speak out or engage in activism for fear of being monitored or punished.

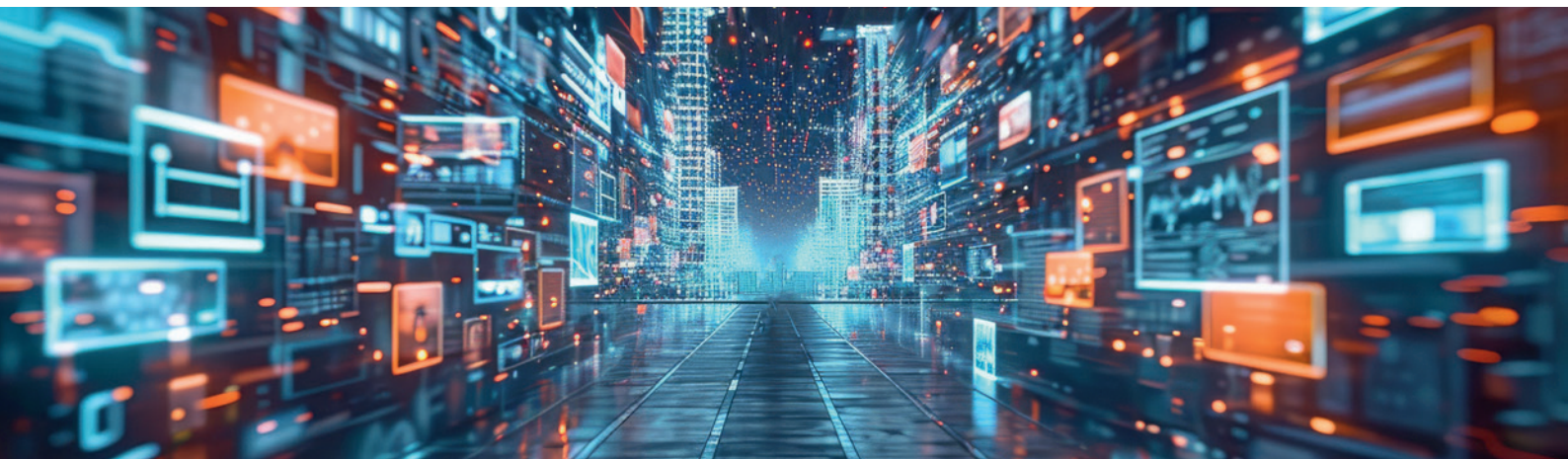
- **Exploiting AI technologies to conduct disinformation and propaganda campaigns** that aim to discredit political opponents and undermine civil society or engaging in foreign information manipulation and interference (FIMI). These campaigns can operate across borders, employing tactics like using bot accounts and astroturfing (an orchestrated response that gives the impression of fake support or widespread discontent). Bots can flood online forums with deceptive content, and generating false text, audio, and images. A recent infamous case is the November 2024 Romanian presidential election, which was annulled on grounds of astroturfing and other forms of algorithmic manipulation. AI-facilitated disinformation can also erode trust in social structures and the information environment and leads to a "**liar's dividend**," where people struggle to distinguish between true and false information, ultimately doubting all sources. In functioning democracies, journalists and civil society groups can counter disinformation efforts. However, **in autocratic regimes, the use of AI** allows for rapid, large-scale manipulation by combining older methods of repression – such as state-controlled media and mass surveillance.

- Misusing **AI in conflict situations:** While discussions often focus on lethal autonomous weapons (LAWS) or 'killer robots', AI in conflict extends AI-enabled intelligence gathering, such as through the use of drones and **AI facilitated identification of potential targets**. Authoritarian contexts and conflict situations intertwine as the theatre of war has been used as testing grounds for technologies, including AI. In turn, the access to and utilisation of AI tools by authoritarian governments can also serve as a tool of intimidation towards its populace.

- Particular **risks to women and girls:** In authoritarian regimes, women and girls can face specific risks from AI misuse that exacerbate existing inequalities. For example, in Iran, the government has deployed AI technologies, such as facial recognition and automated surveillance tools, to enforce strict morality laws and crack down on women's rights movements. Following the killing of Mahsa Amini, AI-assisted repression has intensified, with authorities utilising these technologies to monitor and punish women who defy hijab regulations, thereby infringing on their personal freedoms and safety. The pervasive use of AI in these contexts not only threatens individual rights but also reinforces systemic discrimination against women and girls. As segments of society benefit from advancements in AI technology, those who are digitally marginalized –often including women and girls – risk being left further behind, impeding their access to opportunities and threatening their fundamental human rights. Addressing these challenges is crucial for promoting gender equity and safeguarding the rights of women and girls within oppressive regimes.

The role of private companies, particularly tech firms, in enabling state repression is a growing concern in authoritarian regimes. In countries like China, major tech firms such as Huawei and ZTE are not only state-owned but also actively develop and deploy AI surveillance systems that monitor citizens and suppress dissent. Even Western tech companies can inadvertently support authoritarian practices; for example, the **French surveillance companies Amesys and Nexa Technologies** were indicted for allegedly providing surveillance technology to governments in Libya and Egypt, potentially enabling human rights abuses.

At the same time, even in authoritarian contexts, AI systems can empower civil society, political activists, and human rights defenders. They have significant potential to uncover abuses of power and detect or forecast human rights violations. Civil society can enhance its open-source investigation (OSINT) efforts by using AI to document human rights abuses for future accountability in international forums. For instance, the **Syrian Archives**, which hosts over 3 million videos, may play a crucial role in transitional justice and reconciliation efforts. AI can streamline the analysis of such content, saving time and costs while relying on human verification as a safeguard. Additionally, AI can help anticipate cyberattacks and identify deepfakes or other manipulated content.



Implications of AI misuse in authoritarian regimes for German development cooperation

Key challenges

- **Human rights violations:** The use of AI for surveillance and repression can undermine the mission to promote human rights. Additionally, the potential compromise of privacy can complicate advocacy for human rights in partner countries.
- **AI-driven disinformation campaigns** distort narratives both around socio-political issues but also on the work of development agencies themselves. In consequence, it can complicate international dialogue and collaboration efforts. A focus on countering AI misuse may divert resources from other essential development priorities.
- **State control over dissent** can hamper the ability to support civil society and grassroots movements. Covert surveillance tools can also affect the ability to monitor human rights abuses effectively.
- **Ethical dilemmas:** Collaborations with tech companies that knowingly deploy tools enabling political repression and human rights violations pose ethical challenges for partnerships (e.g. OpenAI and Google are increasingly adopting stances that support the national security narrative, including by removing specific provisions in their AI principles to that effect).
- **Conflict escalation risks:** The potential for AI to exacerbate conflicts necessitates proactive engagement and intervention strategies.

Key opportunities

- **Promoting the responsible use of AI technologies** and advocating for the development of ethical frameworks that prioritise democratic principles – such as participation and transparency – while safeguarding human rights and ensuring that AI technologies are not used for repression.
- **Supporting civil society** in leveraging AI tools to document and expose human rights abuses, countering state-sponsored narratives. This also includes support through training to civil society organisations on effectively using AI for open-source investigations to gather evidence of abuses.
- **Strengthening private sector support for ethical and human rights respecting AI**, including by encouraging the adoption of ethical principles or through using existing frameworks to gauge human rights impacts of AI.
- **Fostering technological innovation** that counteracts AI misuse, such as tools for detecting mis- and disinformation and protecting digital rights.
- **Harnessing AI for gender equity**, including targeted programs that empower women and girls through digital literacy training, support for NGOs, advocacy for gender-sensitive AI policies, initiatives to bridge the digital divide, research on the impact of AI, and collaboration with international partners to promote responsible and equitable AI use.

2.2 Democratic and semi-democratic systems

AI used in democratic contexts can help to empower individuals and collectives, strengthen institutions and contribute to societal resilience.

Deployment of AI tools enhances **citizen engagement** through tools like Taiwan›s Pol.is, which facilitates real-time public feedback on policy-making or other efforts **tested in various jurisdictions**. AI can also help media and civil society to detect **mis- and disinformation** and build trust in journalism with platforms like **Co-facts**, promoting informed discourse. Moreover, it can be used to analyse data for **policy effectiveness**, as demonstrated by **Open Knowledge Brazil's use of natural language processing** to create actionable insights. Additionally, AI tools can support civil society in identifying **abuses of power** and advocating for marginalised groups, exemplified by a **custom tool in Kenya** that informs citizens about the impacts of controversial legislation.

At the same time, as documented by **several studies**, the deployment of AI tools can be a **potential threat** to the integrity of democratic processes. Even in democracies, there is a risk that AI technologies could be misused by governments for mass and other surveillance or for suppressing dissent under the guise of public safety or efficiency. This is particularly evident in hybrid regimes or semi-democratic systems, where democratic principles are undermined by practices that limit political freedoms and weaken key institutions.

In addition to challenges mentioned in the previous section, the key problems include:

- Algorithmic **discrimination**, where AI systems in public administration may unfairly target marginalized populations, leading to biased outcomes in areas like social welfare or tax fraud detection. An example for this was seen in the Netherlands (**an AI system used to monitor the allocation of child welfare benefits** flagged those holding dual nationalities as being at a higher risk), **South Africa** (predictive policing have been found to disproportionately target low-income communities) or in **Kenya** (biased outcomes in fintech was found to particularly affect women with limited internet access).

- **Lock-in effects** arising from over-reliance on historical data in AI systems, which can entrench policy measures and stifle innovation, ultimately diminishing individual autonomy and political contestation that is essential for thriving democracy. Historical data can also exacerbate algorithmic discrimination as such data can reflect long-standing societal bias.

- **The lack of transparency and accountability** surrounding AI systems, often sourced from private companies, complicates oversight and limits individuals› understanding of (and trust in) decision-making processes.

- **The erosion of public trust**, exacerbated by AI-fuelled disinformation campaigns undermines confidence in information sources and democratic institutions. High-profile incidents like the **Cambridge Analytica scandal** and recent **disinformation in Romania's presidential election** illustrate how AI can significantly impact electoral integrity.

- **Evolving challenges:** As seen during the super election year of 2024, traditional election monitoring practices struggle to combat the rapid spread of mis- and disinformation on social media platforms, for instance with electoral management bodies (EMBs) often being under-resourced and ill-equipped. The evolving nature of information pollution (mis- and disinformation, hate speech) practices means that vigilance and multi-pronged responses in strengthening informational integrity is essential.

- Uncritical adoption of AI across societal institutions risks **deepening the digital divides**, leaving marginalized populations further behind and threatening social cohesion. This can take different forms such as gender-based, rural-urban or age-related digital divides. Digital divides undermine inclusivity, essential for a functioning democracy. While AI holds potential benefits for enhancing democratic processes, its unchecked use raises serious concerns about fairness, accountability, and the overall health of democratic systems.

- AI systems must also account for **regional and local realities**, meaning that partner countries have to address biases in imported AI models and promote locally-driven AI development.

While many countries are now embracing AI, as evidenced by the number of national **AI strategies worldwide**, these efforts often reflect political agendas and lack public involvement in shaping their scope and direction. The focus of AI strategies often tends to be on its economic potential rather than prioritizing public good, human rights, and the needs of citizens. Democratising AI involves not only increasing access but also incorporating collective public input to ensure safe, secure and trustworthy AI.

Indeed, the growth of AI and a robust tech industry can stimulate economic development, and over-regulation may hinder innovation. Engaging the public in AI decision-making ensures that technology aligns with community needs and serves the public good. Nobel Prize-winning economists Daron Acemoglu and Simon Johnson emphasize that inclusive democracies, where citizens participate in decision-making, are more stable over time. This stability, in turn, creates more conducive environment for the private sector and sustainable development. Similarly, this inclusivity is crucial fo ensuring equitable distribution of AI benefits.

Effective regulation can mitigate AI-related harms, enhance trust in technology, and promote sustainability. As AI is a sociotechnical system, its deployment must consider existing institutions that ensure accountability for any negative impacts. AI has the potential to both support and undermine democracy. Therefore, a multi-level approach is essential. Key questions to consider include: Is AI necessary for the task? Does AI empower individuals and communities? Are there robust institutions in place that are trustworthy and accountable? Can AI foster societal resilience and cohesion? How does the AI solution compare to traditional, more environmentally friendly or 'low-tech' methods in terms of environmental impact and resource efficiency?

Key issues for German development cooperation in supporting democracy through AI integration

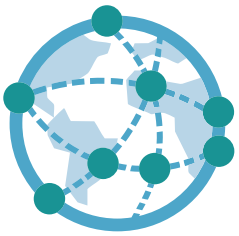
Key challenges

- **Integrity of information ecosystem:** The potential of AI to generate disinformation presents a critical challenge for work in supporting democratic integrity. Disinformation can manipulate public opinion and erode trust in democratic processes, making it vital to develop strategies to counteract these risks.
- **Bias in algorithms:** The risk of bias in AI algorithms poses a significant challenge for initiatives aimed at promoting inclusive governance. If AI systems favour certain demographic groups, they can exacerbate social inequalities and disenfranchise marginalised communities, undermining the democratic principle of equal representation.
- **Opacity of decision-making:** The lack of transparency in AI decision-making complicates accountability in democracy support efforts. Without clear insights into how decisions are made, effective oversight can be thwarted. Citizens may struggle to understand or seek accountability, including in attaining possible remedies. Transparency and explicability of AI systems that govern the lives of citizens are an essential part of respect for the rule of law and a stable democracy.
- **Regulatory frameworks:** Collaborating with governments to establish effective regulatory frameworks for responsible AI (that accounts for local realities) use is essential. This ensures that AI technologies enhance rather than hinder democratic governance, protecting civil liberties, promoting accountability, and increasing trust in technologies and institutions.

Key opportunities

- **Public engagement:** Development cooperation can lead initiatives to involve citizens in shaping AI strategies, ensuring that technologies prioritise democratic principles, public good and human rights.
- **Inclusive access:** Promoting equitable access to AI can empower underrepresented communities, fostering inclusive democracies where all voices are heard.
- **Strengthening regulations, policy and governance measures:** Supporting partner countries in developing AI strategies and regulation. Collaborating with governments to develop effective, tailor-made regulations can mitigate AI-related harms and enhance transparency, building public trust in technology and democratic institutions. Engaging with partner countries to adopt inclusive and human rights respecting AI strategies – moving beyond the latter's typical focus on economic growth.
- **Economic development:** Development cooperation can leverage AI to stimulate economic growth, including through public incentives, while aligning initiatives with community needs, ensuring equitable distribution of benefits that reinforce democratic stability.
- **Strengthening state incentives** towards the adoption, procurement and deployment of ethical and rights respecting AI.
- **Enhancing resilience:** Utilizing AI to empower communities in addressing local challenges can strengthen societal cohesion and improve public services, ultimately supporting a robust democratic framework. An example is using AI as a tool to detect, analyse and combat mis- and disinformation

3 Global developments: Power dynamics and governance



The rapid evolution of AI is not just a technological advancement; it is a pivotal factor in the **global competition for power and influence**, particularly between the United States and China. This competition often manifests as a "winner-takes-all" dynamic, where national security concerns overshadow democratic values. While China leads in AI patents, the U.S. excels in developing **significant machine learning models.** The European Union, on the other hand, aimed to carve out a unique role by emphasizing human-centric AI governance, which seeks to balance innovation with the protection of fundamental rights. However, the sentiment **after the 2025 Paris AI Summit seems to converge on the narrative of competition, even for the EU – in racing towards building capable AI**. Partner countries in the Global South often find themselves navigating through a complex landscape shaped by the competing influences, making the adoption of a value-based approach an essential alternative that should be advanced through international development cooperation.

This is particularly true since the development and use of AI goes beyond purely technical applications. Different actors are defining standards and norms and attempting to enforce them in the context of geopolitical competition, aiming to establish or influence these standards globally. This effort encompasses more than just technical standards, such as on data and network security. It also includes governance aspects such as data protection, ethical principles, and governance of the social impacts of AI. A significant risk for partner countries emerges when authoritarian approaches become globally normalized and align with centralized structures and comprehensive control mechanisms in countries with weak democratic foundations. This highlights that technological innovations cannot be viewed in isolation but rather aligned to strategic political objectives, making governance a critical factor in addressing risks such as surveillance, state control, and restrictions on individual freedoms.

As we explore these themes, several additional key factors come to light that are directly relevant to global development around AI:

- **Investment disparities and digital divides:** As investments in AI surge (the Stargate in the USA and the EU announced a **EUR 200 billion InvestAI initiative**), these efforts drastically upped the ante in the race to uncover the innovative potential of AI. However, resources and investments are unevenly distributed. The majority is concentrated in the U.S., China, and the EU, leading to stark inequalities that threaten democratic participation and access to technology. These inequalities extend beyond software, encompassing disparities in **computational infrastructures**, **hardware**, quality datasets, access to talent, access to quality data and AI literacy. Such gaps contribute to significant digital divides that undermine the ability of various communities to engage meaningfully in democratic processes.

- **Corporate influence:** Private sector dominance complicates this landscape further. Major tech companies often prioritise profit over public interest, **lobbying against regulations** that could protect democratic values. This unchecked growth allows them to develop cutting-edge technologies without democratic oversight or accountability. The exploitation of data access and labor by corporations, including the outsourcing of data processing and model training to low-wage labor markets in the Global South has been critiqued as a form of digital neocolonialism. Notably, US and Chinese corporations play a role in this exploitation, extracting and monetising data from the respective regions. Proposals like establishing a "**CERN for AI**" – supported by think tanks and most recent EU investment in AI, and ongoing efforts at the UN (through the Global Digital Compact) aim to redistribute power and knowledge, fostering a safer and more inclusive approach to technological development.

- **Power dynamics in AI research:** The private sector also features prominently in dominating the field of AI research, with its outputs far outnumbering outputs from academia and governments. Industry involvement within academic research is also increasing, potentially complicating and compromising the value of academic freedom and ethics in the pursuit of research. When research that shapes societal futures is largely controlled by corporations, it undermines the conditions necessary for democracy to thrive. While some companies are trying to democratise AI through initiatives like META's Community Forums, these efforts depend on corporate goodwill and can be easily abandoned.

- **Risk of ongoing developments in the USA and their global impact:** Since January 2025, the new Trump administration has been swiftly dismantling institutions such as USAID and reforming key US federal institutions. On the very first day in office, the administration revoked the October 2023 Executive Order on Safe, Secure and Trustworthy AI signed by the Biden administration, stating that it creates '**barriers to American AI innovation**.' Further shifts toward deregulation in U.S. politics are expected, alongside a renewed focus in propelling American supremacy in the AI race. The hand-in-glove involvement of the tech sector in politics foreshadows a drastic reduction in accountability for tech companies, allowing them to prioritize profit over public welfare and ethical considerations in AI development. Additionally, promoting private-sector innovation without adequate oversight may worsen inequalities in access to AI technologies, particularly for marginalized communities. The deregulatory stance is also pursued by the US on a global level and this could thwart efforts at the UN and other multilateral institutions that aim to pursue safe, secure, trustworthy and equitable access to the benefits of AI, with an outsized impact on the Global South.

- **Global South challenges:** The widening power gaps between states and between governments and private entities present significant challenges for democracy, particularly in the Global South. Countries in this region often find themselves as users rather than developers of AI technologies, lacking the necessary influence, capacity, and expertise. To tackle these disparities, governance frameworks must focus on reducing digital divides and promoting inclusive global cooperation. This is where development cooperation can play a vital role by facilitating partnerships that empower local stakeholders.

Key efforts **to address these inequalities** include governance initiatives aimed at promoting digital sovereignty, creating open models, and controlling the harmful effects stemming from unequal access to technology. Various global governance models are emerging, ranging from soft law ethical principles to more formal regulations. Notable existing efforts include the UNESCO Recommendations on the Ethics of AI, adopted in 2021, as well as initiatives at the UN level, such as the High-Level Advisory Body on AI and the UN Global Digital Compact (GDC).

The EU has taken a pioneering step by introducing comprehensive legislation through its AI Act, which adopts a risk-based approach informed by fundamental rights protections. In contrast, the U.S., driven by a market-oriented approach to innovation, has yet to establish comprehensive AI legislation. Meanwhile, China's state-controlled model reflects a use-case-based legislative approach focused on issues such as algorithmic recommendations. These differing value-driven approaches highlight the challenges of reaching consensus in international governance, even as UN efforts progress. Some core issues, such as AI safety, may be addressed collectively, while others - like algorithmic bias - might require more contextualised domestic or regional regulatory responses.

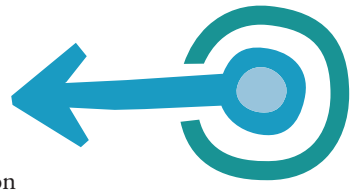
Amidst this landscape of dominance and unequal access to technological benefits, the concept of **digital sovereignty** has gained traction. Countries are encouraged to pursue their own technological pathways without relying on external powers or private interests. For nations in the Global South, integrating local realities into AI development is crucial for ensuring that these systems are applicable and accepted within their contexts. It can also serve to benefit communities on site, rather than corporations in the Global North. Key strategies include investing in local talent, developing context-specific AI models, enhancing AI literacy, investments in local data and compute infrastructures, thus creating an enabling environment for further progress. However, it is essential to balance these efforts with environmental considerations.

Furthermore, despite the dominance of closed source models such as OpenAI's ChatGPT, the pursuit of open-source models has been encouraged, including by key figures within the tech community. Open-source models can bring immense potential for development, encouraging reusability, innovation, accessibility, transparency and the reduction of market dominance. The recent release of the latest model of Deepseek – an open-source model – has reignited focus towards how open source can ignite innovative potential and change the dynamics of the AI race. At the same time, there are risks that open-source models can be misused for malicious purposes such as cyberattacks or disinformation campaigns as these models are accessible to everyone and not typically subjected to centralised control measures. There is a strong focus on AI safety pursued by companies such as Anthropic, Google and OpenAI who argue that centralised controls – including through evaluations and red-teaming – can ensure that models are not misused for malicious purposes. Thus, while acknowledging the immense benefits of open source, requisite governance measures may need to be adopted, alongside the need for further research in this area. New efforts such as **ROOST**, a resource providing open-source safety infrastructure, appear promising in addressing these challenges.

A critical takeaway from this analysis is the necessity of including voices from the Global South in emerging global governance frameworks for AI. Governance cannot be shaped solely by the experiences of the Global North; it must also consider how AI harms disproportionately affect communities in the Global South.

4 Recommendations for German development cooperation

These recommendations are aimed at enhancing initiatives of German development cooperation to promote democracy, ensuring alignment with global standards while addressing local needs.



4.1 General policy recommendations

- **Support Global South engagement:** Facilitate the Global South's participation in international AI dialogues, including those with the UN High-Level Advisory Body on AI, and through mechanisms proposed under the Global Digital Compact such as the Scientific Panel on AI and the Global Dialogue on AI Governance. This ensures that local perspectives and needs are integrated into global discussions, promoting equitable representation.

- **Encourage inclusive multi-stakeholder processes:** Advocate for inclusive processes in drafting National AI strategies that involve diverse stakeholders, including marginalized communities. This can be achieved through citizen assemblies and collaborations with civil society organisations to ensure broad public engagement.

- **Promote a critical approach to AI adoption:** Emphasise the need for a critical adoption of AI technologies that recognizes them as sociotechnical systems; advocate for a risk-aware and rights-respecting framework that prioritises human welfare over technological determinism.

- **Be cautious about adoption of EU regulations:** While acknowledging the strengths of the EU's regulatory approach to AI, advise against uncritical adoption in other jurisdictions. The EU AI Act is informed by a focus on the protection of health, safety and fundamental rights and these are basic values that can and should be replicated. However, the EU regulation is part of a broader legal framework that includes various institutions and regulations that are there to ensure the protection of these rights. Promoting of its adoption must take into account the state of rule of law, whether there are existing local institutional protections and at the same time, also encourage the development of regional solutions and contextualising political debates around locally identified needs.

- **Implement ethical guidelines:** Urge states to adopt and implement frameworks like the UNESCO Recommendation on the Ethics of AI. This includes utilising tools such as the Readiness Assessment Methodology (RAM) and Ethical Impact Assessment (EIA) to ensure ethical practices in AI deployment.

- **Promote the adoption of a risk-based approach to AI governance** among partners, emphasising the identification, measurement, management, and mitigation of risks associated with AI technologies. This approach should align with frameworks such as the Artificial Intelligence Risk Management Framework by the National Institute of Standards and Technology (NIST Risk Management Framework) or the ISO standard 42001 on AI management systems. These tools help to guide risk management, impact assessments and in the adoption of a lifecycle approach, leading to more trustworthy AI. This is particularly important where standards are linked to regulatory compliance, such as in the EU (which has its own upcoming standards for compliance), ensuring that AI development respects and upholds democratic values and human rights.

- **Support digital sovereignty initiatives:** Tailor support for digital sovereignty efforts based on the specific needs of each country. This includes enhancing local capabilities in computing resources, data (e.g. building up a data ecosystem), model development, fostering local technical talent development, and AI literacy to empower communities.

- **Foster private sector transparency:** Encourage transparency within the private sector regarding AI research and development; advocate for robust accountability frameworks that ensure AI technologies are developed responsibly and ethically.

- **Ensure that any AI-related initiative includes clear provisions for transparency, accountability, and access to remedies.** These provisions should not make it unnecessarily difficult for individuals to obtain information about AI systems or create overly complicated processes (for example, requiring proof of causation). Encourage open-source development as a way to increase transparency. Additionally, where applicable, ensure that the responsibility for proving compliance with these standards lies with the designers or operators of AI systems, not with individuals affected by them.

4.2 Operational recommendations

To ensure that strategic policy recommendations lead to tangible outcomes, a series of operational measures is proposed to support the overarching objectives.

Within (semi-)democratic contexts:

- **Support AI governance and accountability:**
  - **Support state-level AI governance:** Provide technical assistance to governments in developing inclusive AI governance strategies that respect human rights and democratic values while accounting for local realities. This includes conducting feasibility studies, enhancing AI literacy within various institutions and creating implementation roadmaps.
  - **Support enhancing data governance practices:** Collaborate with governments to create robust data governance frameworks and ensure public datasets are representative, secure, and up-to-date, incorporating best practices in cybersecurity and data protection. This should also include support to creating accessible **public AI registers** listing all government uses of AI and include information on how citizens can seek accountability regarding these technologies.
  - **Encourage voluntary industry standards:** In regions lacking regulatory frameworks, motivate industry stakeholders to adopt voluntary AI standards based on universal values, such as those outlined by UNESCO.

- **Increase capacity building and AI literacy:**
  - **Support building AI literacy in civil society and human rights institutions:** Equip national civil society and human rights bodies with training on AI literacy, focus on identifying and addressing democracy and human rights impacts of AI through workshops, staff exchanges, and regional collaborations.
  - **Support local capacity building and community-led AI initiatives:** Strengthen local technical expertise in AI through targeted training programs and workshops. This can include support to collaboration with educational institutions to develop curricula that focus on ethical AI practices and governance, or initiatives aimed at improving tech and AI literacy among diverse groups, including children, youth, and marginalised populations. Further, support training for journalists and civil society on using AI tools to verify the authenticity of various content types and developing tools for content provenance and verification to combat mis- and disinformation across various media formats. Additionally, fund grassroots projects that leverage AI for social justice and democratic engagement, such as citizen assemblies using innovative technologies.
- **Support civil society empowerment and oversight:**
  - **Empower civil society for AI oversight:** Support civil society in mapping AI applications in public services and documenting related harms. Promote transparency initiatives and strategic litigation to hold accountable those responsible for AI misuse.
  - **Strengthen independent media:** Provide independent media organizations with tools to trace content provenance, identify mis- and disinformation, and access training on AI technologies to improve the information ecosystem.
- **Support research and innovation:**
  - **Increase public funding for AI research:** Advocate for increased funding for public AI research initiatives to counterbalance the dominance of privately funded research, ensuring that public interest remains a priority. Specifically, support interdisciplinary research with focus on Global South that examines the intersection of technology and democracy, ensuring that tech-driven solutions do not overshadow democratic processes.
  - **Facilitate scientific exchanges:** Promote responsible technology sharing and scientific collaboration between parties, including between the Global North and South, as well as within the Global South, ensuring that knowledge transfer respects local contexts, needs and do not perpetuate a neo-colonialist agenda.

Additionally, within authoritarian contexts:

- **Empower civil society monitoring**, including accountability of private sector: Fund civil society organisations to document human rights abuses related to AI, disinformation and state surveillance. Provide technological resources (e.g. databases, software) to enhance their monitoring capabilities. Support initiatives that enable civil society to also track the private sector's role in human rights violations and advocate for accountability regarding AI-related harms.
- **Facilitate advocacy against repressive environments:** Assist civil society groups in engaging in regional and international advocacy to amplify their voices in oppressive contexts.
- **Promote collaborative coalitions:** Encourage the formation of coalitions among local civil society actors focused on AI, digital democracy, and technological harms to strengthen collective advocacy efforts.
- **Support capacity building for democratic engagement:** Facilitate training programs that empower civil society and local actors to engage effectively in democratic processes, ensuring their participation in policy-making related to AI governance.
- **Encourage companies to adopt an ethical AI stance:** This includes accessing the impacts of the potential misuse of technology by authoritarian regimes, including for human rights violations. An ethical AI stance would entail that companies should not be complicit in facilitating abuses, violence or repression.

As a federally owned enterprise, GIZ supports the German Government in achieving its objectives in the field of international cooperation for sustainable development.

Published by:
Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn

Division Global Policy, Governance
Section Governance, Human Rights
Sector Programme Governance

Friedrich-Ebert-Allee 36
53113 Bonn, Germany
T +49 228 44 60-0
F +49 228 44 60-17 66

Dag-Hammarskjöld-Weg 1-5
65760 Eschborn, Germany
T +49 61 96 79-0
F +49 61 96 79-11 15

E info@giz.de
I www.giz.de

Responsible:
Sector Programme Governance

Authors:
Dr. Sue Anne Teo, Associate Research Fellow, Center for European Policy Studies;
Researcher, Raoul Wallenberg Institute of Human Rights

Barbora Bukovska, Senior Director for Law and Policy at ARTICLE 19,
Global Campaign for Free Expression

Contact:
sv-governance@giz.de

Design und Layout:
Barbara Reuter | Oberursel | Germany | barbarareuter-grafik@web.de

Photo credits:
Photo on the cover and backpage © AdobeStock, Jittapon – AI
Photo on page 4 © AdobeStock, leszekglasner – AI | on page 6 © AdobeStock, bird_saranyoo– AI | on page 8 © AdobeStock, visoot - AI | on page 13 © AdobeStock, Vadym – AI | on page 14 © AdobeStock, metamorworks | on page 16 © AdobeStock, Mikki Orso – AI | on page 20 © AdobeStock, Miumzlik – AI

On behalf of
German Federal Ministry for Economic Cooperation and Development (BMZ)
Division G12: Governance
Bonn, Germany

GIZ is responsible for the content of this publication.

Eschborn/Bonn April 2025

Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn

Friedrich-Ebert-Allee 36 + 40
53113 Bonn, Germany
T +49 228 44 60-0
F +49 228 44 60-17 66

Dag-Hammarskjöld-Weg 1 – 5
65760 Eschborn, Germany
T +49 61 96 79-0
F +49 61 96 79-11 15

E info@giz.de
I www.giz.de

On behalf of

Federal Ministry
for Economic Cooperation
and Development