



FRIEDRICH NAUMANN
FOUNDATION For Freedom.

AI for Human Rights Monitoring

How Europe can become a leader

ANALYSIS

Imprint

Publisher

Friedrich-Naumann-Stiftung für die Freiheit
Truman Haus
Karl-Marx-Straße 2
14482 Potsdam-Babelsberg
Germany



/freiheit.org



/FriedrichNaumannStiftungFreiheit



/FNFreiheit

Authors

Milica Mirković
Research Consultant, Geneva Academy for Human Rights

Editor

Patrick Reiland, Program Manager, Human Rights Hub
Friedrich Naumann Foundation for Freedom

Contact

Phone: +49 30 22 01 26 34
Fax: +49 30 69 08 81 02
email: service@freiheit.org

Date

12.2025

Notes on using this publication

This publication is an information offer of the Friedrich Naumann Foundation for Freedom. It is available free of charge and not intended for sale. It may not be used by parties or election workers for the purpose of election advertising during election campaigns (federal, state or local government elections, or European Parliament elections).

Licence

Creative Commons (CC BY-NC-ND 4.0)

Table of contents

1. Executive Summary	4
2. A critical juncture: What can AI do for human rights monitoring	5
3. Persisting challenges in using AI for human rights monitoring	6
4. Europe's opportunity to lead	7
5. Governance challenges	8
6. Policy priorities	10

1. Executive Summary

Artificial Intelligence (AI) is transforming how we gather and analyze information, offering powerful new tools to monitor human rights around the world. From scanning satellite images for evidence of atrocities to analyzing thousands of reports for emerging patterns, AI can help human rights defenders work more efficiently and effectively. Yet without proper safeguards, these same technologies pose risks that could undermine human rights protection efforts, particularly in monitoring contexts, through issues such as bias and discrimination, erosion of privacy, and lack of transparency.

At this critical juncture, the European Union (EU) has a strategic opportunity to lead global efforts in harnessing AI for good. The EU's strong human rights tradition, robust regulatory frameworks (such as the General Data Protection Regulation (GDPR) and the AI Act), and institutional strengths position it to set standards that ensure AI advances justice and human rights values.

This policy brief outlines how AI can enhance human rights monitoring, the persisting challenges in applying AI in human rights monitoring and Europe's opportunity to lead in AI for human rights monitoring. It concludes with actionable priorities for politicians and stakeholders to help secure a future where AI facilitates and strengthens, rather than threatens, the protection of human rights, with Europe at the forefront.

Some of the insights presented in this brief are drawn from the interdisciplinary discussions held at the AI for Human Rights: Smarter, Faster, Fairer Monitoring workshop, co-organized by the Geneva Human Rights Platform and the Friedrich Naumann Foundation for Freedom Human Rights Hub at the 2025 AI for Good Global Summit in Geneva. This event brought together practitioners from the human rights, technology, and policy fields to explore the potential and risks of AI in human rights and SDG monitoring. The depth of exchange underscored the urgent need for informed, ethical, and inclusive approaches to AI deployment, reinforcing the relevance and timeliness of the policy reflections that follow.

2. A critical juncture: What can AI do for human rights monitoring

AI technologies are rapidly becoming integral to how societies address complex challenges, and human rights monitoring is no exception. When applied responsibly, AI can make monitoring efforts more timely, efficient, and scalable, thus complementing, rather than replacing, human expertise.

Human rights monitoring has already been significantly transformed by the emergence of digital tracking tools and databases, which have enhanced data collection and analysis and enabled near real-time updates on the implementation of international human rights standards across national contexts. These systems manage a wide array of data, from national legislation and policy developments to records of violations and the implementation of recommendations issued by UN and regional bodies.¹

However, the proliferation of such tools, often developed in isolation and based on differing data formats or processing standards, poses challenges to interoperability and meaningful data exchange.² This fragmentation underscores the need for greater integration across platforms and actors.

AI has the potential to address these issues and substantially enhance digital human rights tracking. By processing large volumes of data in real time and automating key elements of the monitoring process, AI can help overcome issues of fragmentation and inefficiency of monitoring efforts. In practical terms, it offers several critical capabilities to human rights professionals involved in monitoring, as well as to policymakers who rely on monitoring outcomes to inform decisions:

Data processing and pattern recognition: AI systems can analyze large volumes of content, ranging from satellite images to social media posts and official documents, far beyond human capacity.³ For instance, machine learning (ML) models have been used to identify the destruction of villages in Darfur and monitor attacks on the Rohingya in Myanmar through satellite imagery.⁴ Natural Language

Processing (NLP) tools can detect patterns and shifts in official narratives, flagging early signs of rights violations.⁵

Enhanced early warning: Predictive analytics powered by AI can support early action in crisis-prone contexts. By training models on historical data, organizations like the Danish Refugee Council have been able to forecast displacement patterns and prepare humanitarian responses more effectively demonstrating how algorithmic foresight can strengthen prevention.⁶

Granular insight into vulnerable groups: AI enables disaggregation of data to highlight how the implementation of human rights standards impacts specific populations, such as women, children, or ethnic minorities,⁷ supporting more targeted and equitable policy responses. For example, Plan International's Girls' Rights Platform uses the UWAZI tool, powered by ML, to automatically collect and categorize human rights documents by topic and affected group.⁸

Cross-verification and validation: AI can automatically cross-reference multiple sources of information, government reports, NGO documentation, and media coverage, to corroborate or flag discrepancies. This function is especially useful in contexts like state reporting to UN bodies, where official national narratives may need to be balanced against independent assessments often contained in shadow reports.⁹

Streamlined reporting and accessibility: AI can assist in drafting human rights reports by gathering data from different sources in real-time and processing complex textual information. A notable example is the integration of AI features into the National Recommendations Tracking Database (NRTD), developed by the UN Human Rights Office. These features enable government stakeholders to generate AI-suggested reports on follow-up to UN recommendations, based on the implementation data they input into the system.¹⁰

In short, AI offers exciting and complex possibilities for

¹ Zipoli, "Briefing N° 23."

² Ibid.

³ Beduschi, "Human Rights and the Governance of Artificial Intelligence."

⁴ Dulka, "The Use of Artificial Intelligence in International Human Rights Law."

⁵ Park et al., "How to Teach Machines to Read Human Rights Reports and Identify Judgments at Scale."

⁶ Dulka, "The Use of Artificial Intelligence in International Human Rights Law."

⁷ Mirković, Scurrall, "AI Decoded: Key Concepts and Applications of Artificial Intelligence for Human Rights and SDG Monitoring."

⁸ Finch, "How machine learning is helping Plan International put girls' rights at the forefront of the international agenda."

⁹ Mirković, Scurrall, "AI Decoded: Key Concepts and Applications of Artificial Intelligence for Human Rights and SDG Monitoring."

¹⁰ Geneva Academy, "Exploring the role of Artificial Intelligence in human rights monitoring."

human rights monitoring. It can make monitoring systems smarter, faster, and fairer, but realizing these benefits requires conscious effort. This is a critical juncture: the

technology to greatly improve human rights protections is there, but must be used with care and foresight.

3. Persisting challenges in using AI for human rights monitoring

While AI offers transformative potential for human rights monitoring, its use in such sensitive contexts raises critical risks. Without clear, enforceable safeguards grounded in international human rights standards, AI tools may inadvertently harm the very causes they aim to support, reinforcing or even worsening existing inequalities. The key concerns include:

Algorithmic bias and discrimination: AI systems are only as reliable as the data they are trained on. ML models often reflect the biases present in datasets, particularly when certain demographic groups are underrepresented or misrepresented.¹¹ As a result, these models can produce discriminatory outcomes. In human rights monitoring, this may lead to blind spots, such as failing to detect violations affecting specific communities, simply because the relevant data is missing or skewed.¹² Vigilance, along with the use of diverse and representative training data, is essential to prevent AI from perpetuating injustice under the guise of objectivity.

Lack of transparency and explainability: AI systems, especially those based on deep learning, often function as “black boxes”, with decision-making processes that are difficult to interpret. This opacity poses serious challenges for accountability, particularly in sensitive fields like state reporting. Indeed, human rights findings based on AI-generated data risk being dismissed if they are not explainable or clearly attributable.¹³ To maintain credibility and trust, stakeholders must be able to understand and challenge AI outputs. Transparency and explainability are therefore essential features of any AI system used in human rights monitoring.

Privacy and misuse risks: AI-driven human rights monitoring often involves processing large volumes of sensitive personal data, raising significant privacy concerns. Without strict safeguards, these tools can become overly intrusive and risk being misused for mass surveillance or repression, particularly by non-democratic actors.¹⁴ For example, facial recognition technology used to locate missing persons could also be deployed to track

activists or minority groups. Adherence to data protection principles, along with the use of Privacy-Enhancing Technologies (PETs),¹⁵ is essential to prevent such abuses and uphold individual rights.

Data gaps and unreliable inputs: In many regions, particularly in conflict zones or low-resource settings, data is often sparse and inconsistent, or even politically manipulated.¹⁶ AI models trained on such inputs can produce misleading conclusions by underreporting violations or generating false narratives, thereby deepening existing marginalization and discrimination. Addressing this challenge requires sustained investment in ground-level data collection and human oversight of AI-generated outputs to ensure contextual accuracy and fairness.

Global inequality and capacity gaps: Access to advanced AI tools remains uneven. Well-funded entities can adopt and deploy AI more easily, while smaller civil society organizations (CSOs), particularly in the Global South, often face significant resource and technical constraints. This reflects a broader structural imbalance: AI development and use are largely dominated by industrialized countries, while many Global South nations lack the infrastructure and support needed to access or shape these technologies.¹⁷ Without targeted investment and capacity-building, AI risks deepening existing power imbalances and limiting inclusive participation in human rights monitoring.

Reduced human oversight and ethical judgment: Over-reliance on AI risks sidelining the ethical and contextual judgment that only humans can provide. Human rights monitoring requires more than pattern recognition: it demands critical reflection and moral discernment. As such, monitoring cannot be fully outsourced to AI, but should instead be used to assist human

¹¹Dulka, “The Use of Artificial Intelligence in International Human Rights Law.”

¹²Musizvingoza, “AI as a Catalyst for Sustainable Progress: Ensuring Information Integrity.”

¹³Zipoli, “Briefing N° 23.”

¹⁴Mirković, Scurrall, “AI Decoded: Key Concepts and Applications of Artificial Intelligence for Human Rights and

SDG Monitoring.”

¹⁵OECD, “Emerging Privacy-Enhancing Technologies.”

¹⁶Miller et al., “An agenda for addressing bias in conflict data”.

¹⁷Scurrall, Mirković, “Who writes the code? AI integration in a post-aid world.”

analysts.¹⁸ Maintaining human-in-the-loop systems and clear protocols for review and correction is crucial.

These challenges are compounded by a fragmented global regulatory landscape. In early 2025, the United States (U.S.) reversed course on its previous oversight-oriented approach by issuing an Executive Order that dismantled federal safeguards and emphasized deregulation. By framing AI as a tool for national competitiveness, the U.S. strategy removes many of the equity, transparency, and safety provisions embedded in earlier policies.¹⁹ This deregulatory shift contrasts sharply with the EU's rights-based model that "reflects a precautionary principle that prioritizes societal safeguards"²⁰. As the EU AI Act sets strict compliance obligations for high-risk AI systems, companies operating across jurisdictions may find it increasingly difficult to navigate conflicting standards. The absence of shared global rules risks regulatory fragmentation and weakens efforts to build trustworthy AI systems that uphold human rights.

Within this European context, Germany is positioning itself as a frontrunner in ethical AI governance. Closely aligned with the EU AI Act, German policy emphasizes transparency, accountability, and data protection, aiming

to strike a careful balance between innovation and regulation. The country is fostering a robust governance ecosystem by prioritizing the adoption of trustworthy AI and reinforcing its role as a global leader in responsible AI development. Through its Strategy for International Digital Policy, the government promotes a human-centered approach to AI, while national regulations seek to ensure that systems are explainable, non-discriminatory, and aligned with fundamental rights.²¹ Germany's proactive stance makes it a key reference point for broader EU ambitions to shape international AI norms.

As EU member states pursue their own approaches to AI governance, there is a growing momentum within the Union to define common regulatory frameworks. This is particularly urgent in high-stakes contexts such as human rights monitoring, where trust, accuracy, and legitimacy are paramount. To ensure that AI supports rather than threatens these principles, responsible and transparent governance rooted in international human rights standards must become the norm. Ensuring that AI operates within a clear, transparent, and rights-compliant framework will be essential to maintaining public confidence.

4. Europe's opportunity to lead

Europe today stands at the forefront of defining a digital future grounded in ethics and human rights.

A tradition of rights and rules

The EU has long been a champion of human dignity and the rule of law, values codified in instruments like the EU Charter of Fundamental Rights and the European Convention on Human Rights. This ethical backbone is reflected in Europe's approach to digital governance. The EU General Data Protection Regulation (GDPR), for example, set a global benchmark for privacy protection in 2018, asserting that individuals' data rights must be respected by corporations and governments alike.²² Now, the EU is poised to do the same for AI. In late 2023, EU legislators reached an agreement on the world's first comprehensive AI law, the AI Act, aimed at ensuring AI systems are aligned with European values of human rights and democracy. This risk-based legislation bans the most harmful AI practices, such as social scoring and indiscriminate surveillance, while imposing strict regulations on high-risk uses and requiring transparency and human oversight, making it one of the most

comprehensive AI frameworks globally.²³ To safeguard fundamental rights, the AI Act mandates that deployers and operators of high-risk AI systems conduct Fundamental Rights Impact Assessments (FRIAs) to identify and mitigate potential harms to individuals' rights and freedoms.²⁴ By taking this bold step, the EU has made clear that AI's social impact will not be left to chance, or to the discretion of Big Tech.

Institutional capacity and inclusive policy making

Europe's multi-layered governance system, including the European Commission, Parliament, Council, national regulators, courts, and agencies, provides a strong foundation for implementing and enforcing AI regulation. The EU draws on the expertise of bodies such as the European Union Agency for Fundamental Rights (FRA) and national Data Protection Authorities to assess AI systems against human rights standards. Under the AI Act, this framework is bolstered by a dedicated enforcement structure: the EU AI Office oversees general-purpose AI models, while National Competent Authorities and market surveillance bodies supervise national implementation of

¹⁸Geneva Academy, "Exploring the role of Artificial Intelligence in human rights monitoring."

¹⁹Mackowski et al., "Key Insights on President Trump's New AI Executive Order and Policy & Regulatory Implications."

²⁰Ibid.

²¹MarketsandMarkets, "Germany's AI Governance

Revolution: Leading with Ethics and Innovation."

²²EU Parliament, "The impact of the General Data Protection Regulation (GDPR) on artificial intelligence."

²³Molina Obedman, "EU rights watchdog reveals 'significant gaps' in oversight of major tech platforms."

²⁴Waem et al., "Fundamental Rights Impact Assessments under the EU AI Act: Who, what and how?"

the AI Act and oversee high-risk AI systems.²⁵ The AI Office holds significant powers, including the ability to request technical documentation from providers of AI models, conduct evaluations and investigations, and mandate corrective actions or market withdrawal of AI systems when necessary.²⁶

Europe also benefits from a vibrant civil society and research community (e.g. organizations like AlgorithmWatch and academic centers on AI ethics) that contribute to governance efforts and help hold stakeholders accountable. Crucially, the EU's approach to AI governance is generally inclusive and evidence based. For example, the AI Office has invited input from a broad range of actors, from AI developers and deployers to CSOs and academia, for the drafting of the first "general-purpose AI Code of Practice", which aims to guide the proper application of the AI Act to general-purpose AI models.²⁷ This participatory approach extends to Europe's external digital diplomacy: the EU's international digital strategy emphasizes partnerships that advance value-based digital governance grounded in human rights.²⁸ Through its engagement in forums such as the Council of Europe, G7, and OECD, the EU can help shape global standards, particularly in areas where other jurisdictions have yet to establish comparable frameworks.

A shifting global landscape

The timing of the EU's regulatory efforts is critical. AI technologies are evolving rapidly, and divergent approaches to their governance by global actors are reshaping the digital ecosystem. As previously noted, in 2025 the U.S. shifted from an oversight-oriented model to a deregulation-focused strategy through an Executive Order that prioritizes AI innovation and global competitiveness over ethical safeguards.²⁹ The resulting deregulatory model now stands in sharp contrast to Europe's rights-based approach, creating potential compliance challenges for global companies and contributing to an increasingly fragmented international regulatory landscape. Meanwhile, actors such as China are advancing state-driven models of AI governance that lack democratic accountability and emphasize national security and economic development.³⁰ This approach has facilitated the use of AI technologies for extensive surveillance and social control.³¹

In the absence of a global consensus on AI regulation, the EU has a strategic opportunity to help shape international standards by promoting its human-centric approach. With its emphasis on transparency, accountability, and the protection of fundamental rights, the EU is well placed to guide AI governance toward trustworthy and responsible outcomes.

5. Governance challenges

The EU has the chance to lead efforts in the application of AI in human rights monitoring as laid out in chapter 2 of this document. However, it faces underlying issues related to its global position in AI governance as well as internal implementation of AI regulation. These challenges must be addressed and understood in order to properly apply AI technology in the field of human rights.

Chief among these internal challenges is the uneven policy landscape across Member States, marked by varying levels of readiness and institutional capacity for AI governance.³² These disparities risk undermining the coherent implementation of the AI Act and hinder efforts toward a harmonized approach to AI governance generally and a thus also the application of AI for human rights monitoring on the EU level.

Another strategic challenge is the EU's continued dependence on foreign digital infrastructure and AI technology providers, primarily U.S.-based.³³ In light of the U.S.'s recent shift toward a deregulatory AI strategy that prioritizes innovation over ethical safeguards, this reliance has become a strategic vulnerability for the EU, reinforcing the urgency of advancing technological sovereignty on its agenda.

A proposal published by the EU Commission in November 2025, called the digital omnibus, seeks to address this strategic vulnerability. The Commission proposes simplification measures 'to ensure timely, smooth and proportionate implementation of certain of the AI Act's

²⁵Stephenson Harwood, "The EU AI Act: enforcement overview."

²⁶Impact International, "EU AI Act Enforcement: Impact on Business Transparency & Human Rights in 2025."

²⁷EU Commission, "AI Act: Participate in the drawing-up of the first General-Purpose AI Code of Practice"

²⁸EU Commission, "An International Digital Strategy for the European Union."

²⁹Mackowski et al., 19.

³⁰Sajduk, Dziwisz, "Comparative analysis of AI development strategies: a study of China's ambitions and

the EU regulatory framework."

³¹Directorate-General for External Policies of the Union, "Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights."

³²Costa, Mendonça, "Assessing European Union Member States' Implementation of the Artificial Intelligence Act."

³³Tech Policy Press, "Europe Says it's Not Set Up to Succeed Globally on Tech"

provisions³⁴. The aim of the proposal is to reduce the compliance burden on EU tech companies. The proposed measures include a delayed implementation of the rules related to high-risk AI applications. The proposal is controversial as it is seen to weaken the protection of fundamental rights and at the same time strengthening the competitiveness of EU companies. At the time of writing, the digital omnibus is but a proposal and has yet to be voted into effect.

In global AI governance, a key limitation of the EU lies in the territorial scope of its regulation. While the AI Act establishes protections within the EU, it does not restrict companies from exporting AI systems, potentially including those banned domestically, without equivalent safeguards abroad.³⁵ This creates a risk of undermining human rights in jurisdictions with weaker regulatory frameworks. Despite this, EU regulations can still serve as influential reference points for other countries shaping their own AI policies.

The EU's rights-based legal framework, along with its strong institutions and regulatory experience, positions it to shape the global trajectory of AI in ways that reinforce human rights. Both in the lifecycle of the technology as well as its implementation for human rights monitoring. However, sustaining this leadership requires coherent internal implementation and strategic international engagement, both grounded in a commitment to responsible AI governance. Seizing this opportunity will demand consistent effort. European policymakers must remain vigilant against backsliding or external pressures that could undermine human rights standards. They must ensure that the EU's internal actions, such as regulation, enforcement, AI deployment, and innovation funding, continue to reflect its core values. In doing so, Europe can credibly present itself as a leader in responsible AI governance and a partner that other global stakeholders can cooperate with and emulate. The following section outlines concrete policy recommendations to advance the application of AI for human rights monitoring.

³⁴EU Commission, "COM(2025) 836 final"

³⁵Impact International, 27.

6. Policy priorities

To capitalize on this strategic moment, the European Union should champion a forward-looking agenda that harnesses AI for human rights monitoring while instituting robust safeguards. Below are key policy priorities and recommendations for politicians and stakeholders:

1. Put human rights safeguards at the center of AI legislation

Ensure that all AI-related laws and delegated acts, including under the AI Act, are grounded in international human rights standards. Support mandatory FRIAs for high-risk AI systems and embed human rights expertise in oversight bodies such as the EU AI Office.

2. Promote public-sector investment in trustworthy AI for monitoring

Encourage investment in responsible development of AI-powered tools for human rights monitoring, especially benefiting smaller CSOs that engage in digital human rights tracking and lack resources to enhance their existing tools and databases. This includes funding AI applications for pattern detection, translation, early warning, and documentation, especially through EU programs like Horizon Europe and NDICI.

3. Support capacity-building for responsible AI development and use

Champion structured training and capacity-building initiatives for civil society, public institutions, and human rights defenders, particularly those in sensitive contexts and under-resourced countries, to equip them with the skills and tools needed to responsibly apply AI technologies in their work.

4. Facilitate equitable data and infrastructure sharing with the Global South

Promote data- and resource-sharing partnerships that bridge the global digital divide. Support technical assistance, open datasets, and access to AI infrastructure for actors in the Global South, ensuring inclusive participation in AI development and governance.

5. Ensure legal requirements for human oversight and transparency

Mandate that AI systems used in high-risk and sensitive contexts, such as human rights monitoring, are subject to meaningful human oversight and transparent, explainable decision-making. Support audits and mechanisms that empower individuals to challenge AI-driven decisions.

6. Champion inclusive, multilateral AI governance

Engage actively in international forums (e.g. Council of Europe, UN, OECD) to promote a global AI framework rooted in transparency, accountability, and human rights. Ensure that global standard-setting reflects diverse perspectives, and advocate for the inclusion of stakeholders from underrepresented regions and communities in standardization and regulatory processes.

7. Engage in Multistakeholder processes of the UN

Engage in the multistakeholder processes surrounding new and emerging digital technologies such as the World Summit on the Information Society (WSIS), the International Telecommunication Union (ITU) and the Internet Governance Forum (IGF). Multistakeholder collaboration within these international bodies will advance innovation, regulatory certainty and allow for the responsible implementation of AI technology for human rights monitoring.

