



Digital Democracy
Initiative

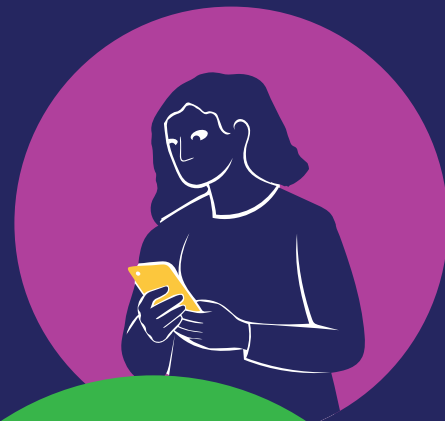
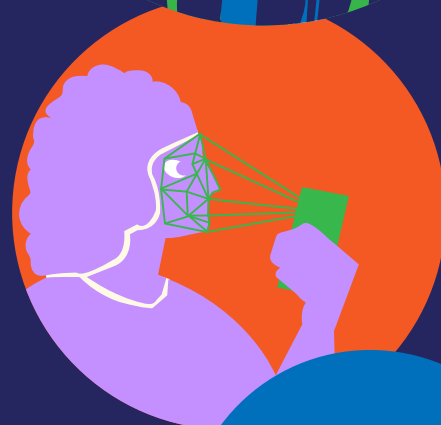
AI & CIVIL SOCIETY:

THREATS AND OBSTACLES TO DEPLOYMENT AND ADVOCACY

2025



CIVICUS



Acknowledgments

Author

Afef Abrougui (Fair Tech)

Design and layout

Monica Molano

We would like to acknowledge and thank all the civil society actors, researchers, and practitioners who generously shared their time and insights through interviews and expert consultations for this research



Digital Democracy Initiative

The Digital Democracy Initiative (DDI) is a programme to safeguard inclusive democracy and human rights in the digital age. It focuses on supporting local civil society in the Global South, particularly in countries undergoing democratic regression and where civic space is under pressure.

For more information visit:

digitaldemocracyinitiative.net



CIVICUS

CIVICUS is a global alliance of civil society organisations and activists working to strengthen citizen action and civil society throughout the world.

For more information visit:

civikus.org

KEY TERMS AND DEFINITIONS

A

- **Artificial Intelligence (AI):** In the field of computer science, AI is defined as machines and technologies performing tasks previously attributed to humans, such as language processing, decision-making, and identifying objects and analysing data.¹
- **AI accountability and advocacy:** For the purpose of this research, this refers to civil society efforts to advance protection of human rights in the deployment and regulation of AI as well as to hold deployers and developers accountable for rights violations.
- **AI hallucinations:** When a large language model (LLM) generates a response “that seems plausible but is actually inaccurate or misleading”.²
- **Algorithmic opacity:**

The lack of transparency about the factors influencing the decisions and outputs of AI systems.

- **AI sandbox:** An isolated testing environment where developers and researchers can experiment with AI and trial new models.

D

- **Data annotation:** The process of labelling raw data such as images, text, videos, and audio to make it understandable to a machine learning model.³
- **Deepfake:** A type of hyper-realistic synthetic media created using GenAI.
- **Deep Packet Inspection (DPI):** A method of examining and managing network traffic, whereby network packets (smaller segments of information sent over the internet such as an email, video or image) are evaluated.

Based on what the content of a packet is, traffic can be altered, blocked or filtered.⁴

F

- **Facial Recognition Technology (FRT):** A type of AI-powered biometric identification system that identifies or verifies someone's identity by analysing their facial features.
- **Fine-tuning:** Databricks defines fine-tuning as “the process of adapting or supplementing pretrained models by training them on smaller, task-specific datasets”.⁵

G

- **Generative AI (GenAI):** A type of AI that uses algorithms to create and generate synthetic content such as text, videos, photos and audio based on training data patterns.⁶

H

- **Human-in-the-loop (HITL):** A process or a system where humans are actively involved in AI workflows so that their feedback can help improve models “to ensure accuracy, safety, accountability or ethical decision-making” of AI systems.⁷

L

- **Large Language Models (LLMs):** In natural language processing, LLMs are AI models trained on large text datasets with the capacity to “understand and generate human-like text, recognize patterns in language, and perform a wide variety of language tasks without task-specific training. They represent a significant advancement in the field of natural language processing (NLP).”⁸

M

- **Machine Learning (ML):** A subfield of AI focused on training algorithms on large datasets to learn from data and make their own rules without being programmed.⁹
- **Misinformation and disinformation:** While misinformation is generally defined as the spread of inaccurate information, disinformation is the spread of false information with the intention to deceive or cause harm.¹⁰

N

- **Natural Language Processing: (NLP):** A subfield of computer science and linguistics concerned with understanding and analysing human language. Widely deployed in many of the technologies and software people use on a daily basis, NLP combines statistical models, machine learning, and linguistic rules to perform a wide range of tasks including sentiment analysis, machine translation, and text classification.¹¹

P

- **Phishing:** A tactic of crafting and spreading deceiving messages to trick recipients into clicking on a malicious link, downloading an infected file or inputting sensitive personal information such as passwords and financial information.¹²

S

- **Spyware:** A type of malicious software (malware) that infects devices to collect personal information and monitor user activities, usually on behalf of governments.¹³

T

- **Training datasets:** Collections of information in different formats—text, images, videos, audio— on which a machine learning model is trained “to make predictions, recognize patterns or generate content”.¹⁴

KEY FINDINGS

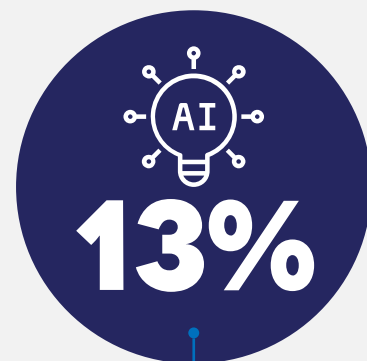
- Surveillance of public and digital spaces pose severe risks to civic space. FRT deployment facilitates mass surveillance in public spaces and severely undermines anonymity, freedom of movement, and freedom of peaceful assembly. AI tools also provide law enforcement agencies and government authorities with the capabilities to collect and analyse vast amounts of data on their citizens, and those they perceive to be a threat, for instance, through AI data mining, predictive policing, and risks analysis tools. Despite the risks and biases associated with these tools, states continue to acquire them from companies looking to profit and then deploy them in human rights violations. In contexts where robust data protection and AI regulations are lacking, civil society organisations (CSOs) are even more vulnerable to the chilling effect of surveillance. As a result of increased deployment of AI in surveillance, civil society actors feel monitored and unsafe. Fearing retaliation, they may resort to self-censorship and avoid protesting or engaging in civic life and public matters.

Surveillance of public and digital spaces pose severe risks to civic space.

- Content moderation systems used by social media platforms have long been shown to impact human rights and the civic space, as they can lead to over-moderation of the legitimate right to free speech, hindering access to information and civil society's ability to mobilise and reach audiences. This is due to a number of factors, especially: 1) embedded biases in these systems which tend to disproportionately affect Global South users; 2) while systems can categorise the type of content posted, they may still fail at determining the context in which it is posted, particularly with sensitive content, for example, health information or documentation of human rights abuses; and 3) the lack of human oversight in algorithmic content moderation makes it difficult for those affected to understand why their content was removed and then to address it.
- Under-moderation of harmful content could put the safety and wellbeing of civil society activists, journalists, and excluded groups at risk. Content recommendation systems deployed by social media platforms can exacerbate the spread of misinformation and disinformation, as well as controversial and sensationalist content such as hate speech and incitement. As a result, access to credible information and open public debate is undermined. Content recommendation systems also downrank the visibility and reach of civil society content.
- The use of AI to spread disinformation and target and discredit civil society is a serious concern. Many survey respondents said they were concerned how these tools could be used to distort narratives about their work and negatively influence public debate. Further, deepfake sexual abuse against

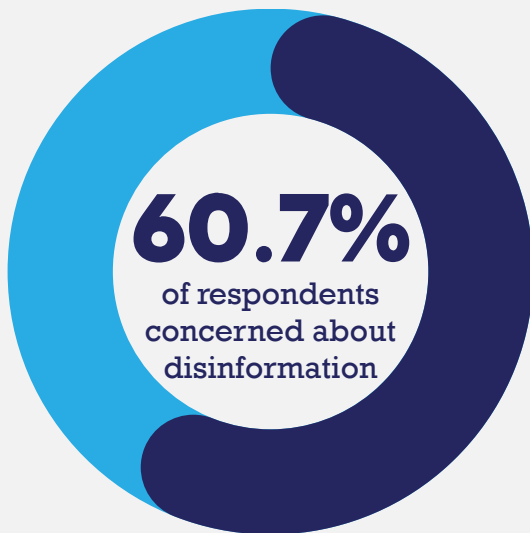
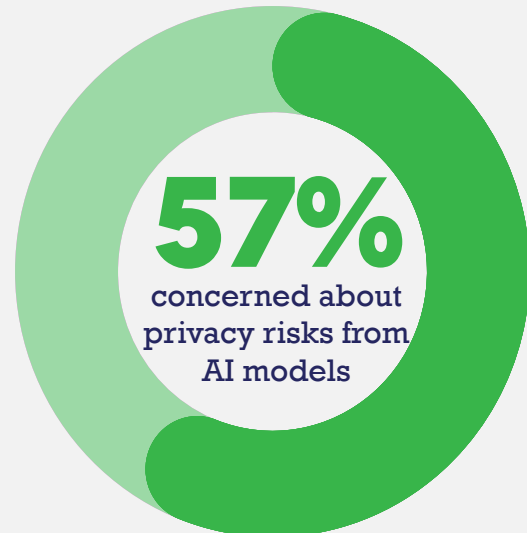
women is contributing to their exclusion from civic participation due to safety concerns.

- Most CSOs that responded to the survey mentioned that they use AI in very limited ways, mostly to assist with research tasks, and some do not use it at all. Only a small fraction of organisations (13 percent of respondents) reported that their organisations have experience with developing their own AI solutions, such as chatbots and AI models, for data analysis or training datasets.

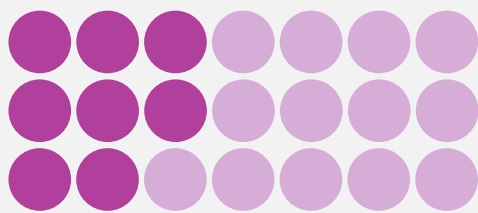


of CSOs have developed their own AI solutions

- Based on the survey results, a lack of financial resources and technical capacity are major obstacles to AI adoption by CSOs, followed by the lack of localised models that understand local contexts and languages.
- Nearly a third of survey respondents cited human rights and ethical concerns as one of the obstacles that prevented their organisations from effectively deploying AI or deploying it at all. Respondents were particularly concerned about the potential exacerbation of misinformation and disinformation (60.7% of all respondents) and risks to the right to privacy associated with the design and use of AI models and systems (57%).



- There is a lack of policies and risk mitigation measures to ensure that AI does not infringe on human rights. Of the 230 respondents who said their organisations deploy AI, only 38 percent stated that they have considered and attempted to address issues of ethics and rights in the use of AI. This can be explained by two factors. First, for those that deploy AI tools, its usage remains very limited, while only a small fraction has developed their own models. Second, CSOs lack the technical expertise and resources to address these risks.



38% of CSOs that deploy AI have addressed ethics and human rights risks

- Measures organisations take to protect human rights with AI include protecting personal and sensitive information, conducting assessments of human rights and data audits, ensuring transparency and rights protection in automated decisions, outputs, and stakeholder engagement.
- Several obstacles prevent CSOs, particularly in the Global South, from engaging in AI advocacy and accountability efforts. First, there is a lack of financial resources and the necessary knowledge and expertise. Second, there is, in some countries, an unfavourable regulatory and political environment, as well as algorithmic opacity and secrecy from both deployers and developers of AI systems that shape AI governance and accountability. Further, CSOs face exclusion from national and international discussions and forums on AI, and for those based in the Global South, expensive travel costs, strict visa rules, predominance of the English language and lack of remote participation alternatives present additional hurdles to engagement.

Several obstacles prevent CSOs, particularly in the Global South, from engaging in AI advocacy and accountability efforts

INTRODUCTION

Willingly or unwillingly, billions of people around the world have come to interact with or be affected by Artificial Intelligence.

Willingly or unwillingly, billions of people around the world have come to interact with or be affected by Artificial Intelligence (AI) in some way or another on a daily basis through their use of digital services and platforms.¹⁵ Content recommendation systems on social media platforms, streaming services, online shopping platforms and search engines shape the content people see and consume.¹⁶ Content moderation algorithms are used to detect undesirable content and remove it.¹⁷ Algorithms deployed by digital assistants recognise speech to help users perform an action, such as placing an online order or playing a song.¹⁸ Smartphones, messaging apps, and email use AI to predict what users are typing and help them write messages faster.

Rapid advancements and wider adoption in various industries, including banking, healthcare, security and defence, welfare allocation and human resources, has not only made AI more ubiquitous but has also given rise to far-reaching ethical, societal, economic and political implications.¹⁹ In particular, military deployments in autonomous weapons are a threat to the right to life and violate international humanitarian law.²⁰ Additionally, a few powerful tech giants are infringing on the right to privacy, stealing copyrighted work, and exploiting

data workers, mostly in the Global South, to train and develop their models.²¹ Further, the unchecked AI advancements are exacerbating the climate crisis due to the excessive use of water resources.²² Of concern, public sector deployment of AI systems, for instance, in border and migration management,²³ policing²⁴ and welfare allocation,²⁵ contribute to the discrimination of excluded groups and amplify inequalities.

Civil society's position in relation to AI deployers, developers, and regulators is unique and shaped by power asymmetries

Civil society, as any other sector, is not immune to AI's disruptive implications and impacts; however, its position in relation to AI deployers, developers, and regulators is unique and shaped by power asymmetries. First, civil society organisations (CSOs) face AI-enabled threats from state and non-state actors that negatively interfere with their human rights work. These include facial recognition systems used for surveillance of activists and protesters,²⁶ content moderation algorithms that suppress dissent, and AI-generated disinformation campaigns that erode information integrity and public trust.²⁷ Second, as CSOs and activists experiment with AI and increasingly adopt it in their work, such as in fact-checking, operations, interactions with audiences and in creating content, they must grapple with the ethical questions and potential risks to human rights.²⁸ These risks include amplifying biases and the spread of misinformation, over-reliance on corporate models from the Global

North, particularly the U.S., and compromising personal and sensitive information.²⁹ Without adequate safeguards, CSOs risk perpetuating the same harms they seek to address. Third, civil society is a key actor in advocating for the respect and protection of human rights in AI deployment and governance.³⁰ Yet, they face multiple challenges, including interference with their work, civic space under pressure, funding challenges³¹ and obstacles in participating in AI governance discussions, which are dominated by a few powerful governmental and private sector actors.³²

This research, therefore, aims to explore these three power asymmetries pertaining to AI and civil society. It aims to provide a better understanding of the risks of AI to CSOs around the world and the challenges they face in the deployment of the AI and in AI advocacy and accountability efforts.

The research is guided by the following questions:

1. What risks does the deployment of AI by state and non-state actors pose to civil society?
2. How do CSOs deploy AI in their work? What challenges and ethical questions arise from their deployment of AI and how do they attempt to address them?
3. What challenges does civil society face in engaging in AI accountability and advocacy efforts?

The first chapter examines the risks of AI for civil society groups and their work. The second chapter of the report sheds light on the extent of AI adoption by civil society, including how they deploy it and navigate risks to human rights and ethical questions. The third chapter looks at how CSOs have been advocating for human rights safeguards in the deployment and governance of AI systems and the different challenges they face in this regard. The report concludes with recommendations to improve support to civil society in their adoption of AI in ways that align with human rights standards and advance a rights-based approach to the adoption and regulation of AI.



METHODOLOGY

The methodology consisted of desk research, a survey of CIVICUS members, interviews and an online focus group.

About this research

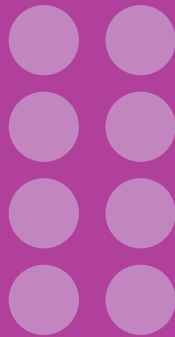
Literature review



293

survey responses

1 focus group
(8 CSO reps)



7



expert interviews



1.5

hours of discussion

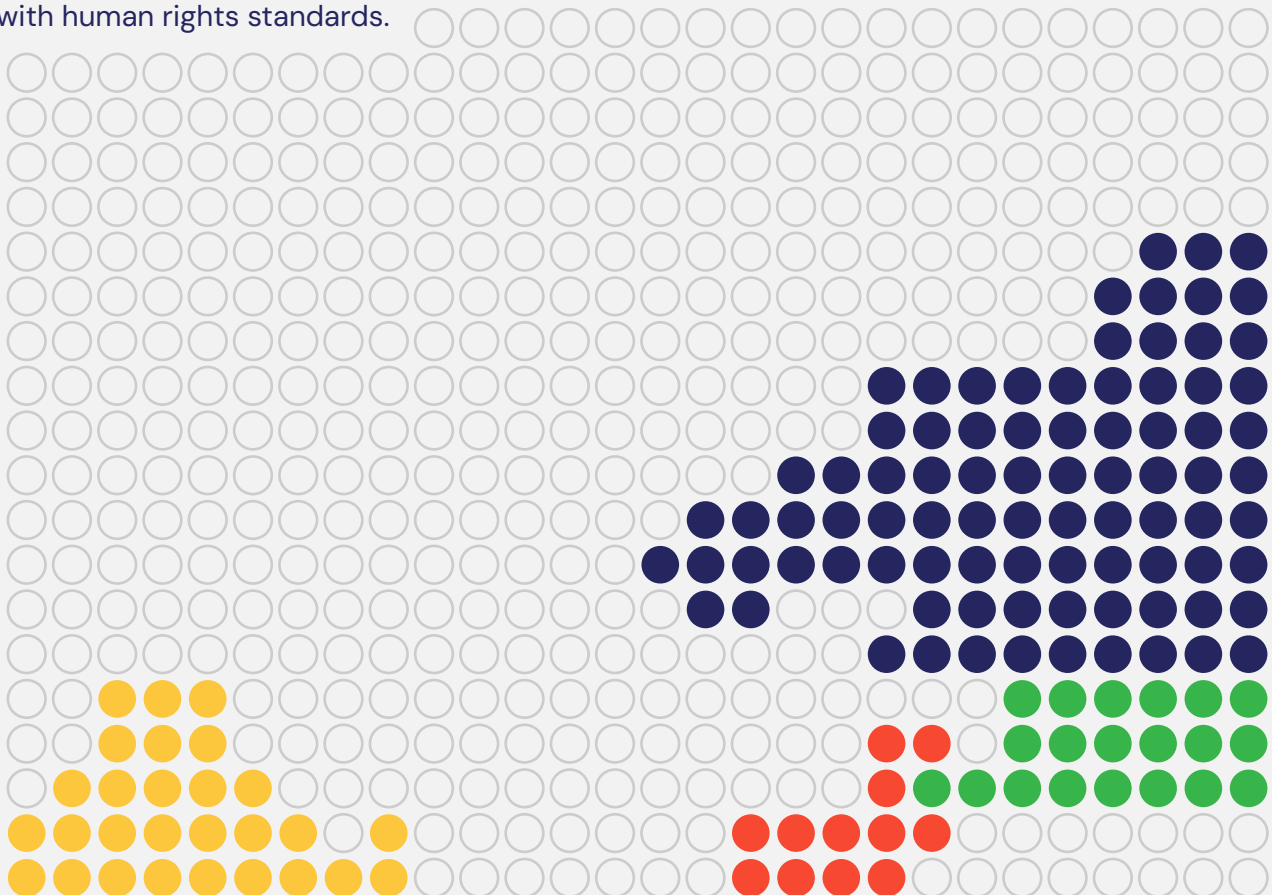
First, a comprehensive literature review was conducted to identify trends and key gaps in knowledge. The review of academic studies, AI policy documents, civil society reports, and credible media reports focused on identifying 1) risks posed by AI to civil society—such as state surveillance, algorithmic content moderation, and disinformation; 2) risks emerging from civil society's adoption of AI tools without adequate safeguards; 3) obstacles and challenges that prevent CSOs from deploying AI or engaging in the field of AI accountability; and 4) a review of key policy trends in AI regulation globally.

Based on the literature review and the research questions, a survey was developed and disseminated among CSOs that are members of the CIVICUS network. The survey questions addressed the following:

- AI systems and applications that pose risks to CSOs and how they impact human rights and their work.
- To what extent CSOs are deploying AI in their work and the challenges they face in AI deployment, and what (if any safeguards) do they have in place to address and mitigate risks associated with such adoption such as bias, privacy, technological dependence on U.S. models, etc.
- To what extent CSOs aim to shape how AI is developed, deployed, and regulated and the obstacles they face in advocating for human-rights based deployment and governance of AI (such as lack of funding and resources, algorithmic opacity, lack of opportunities to develop required knowledge and skills, etc.).

The survey was disseminated among CIVICUS members and 293 responses were received. The interviews were conducted with seven experts working at the intersection of human rights and AI and technology to help gain more in-depth insights and address any gaps in data.

Subsequently, a 1.5-hour focus group discussion (FGD) was conducted remotely with eight representatives of CSOs that have experience with deploying AI to discuss their experience and how they approach ethical questions and govern AI use internally. This FGD helped inform the development of an introductory toolkit on how CSOs can deploy AI in a way that aligns with human rights standards.





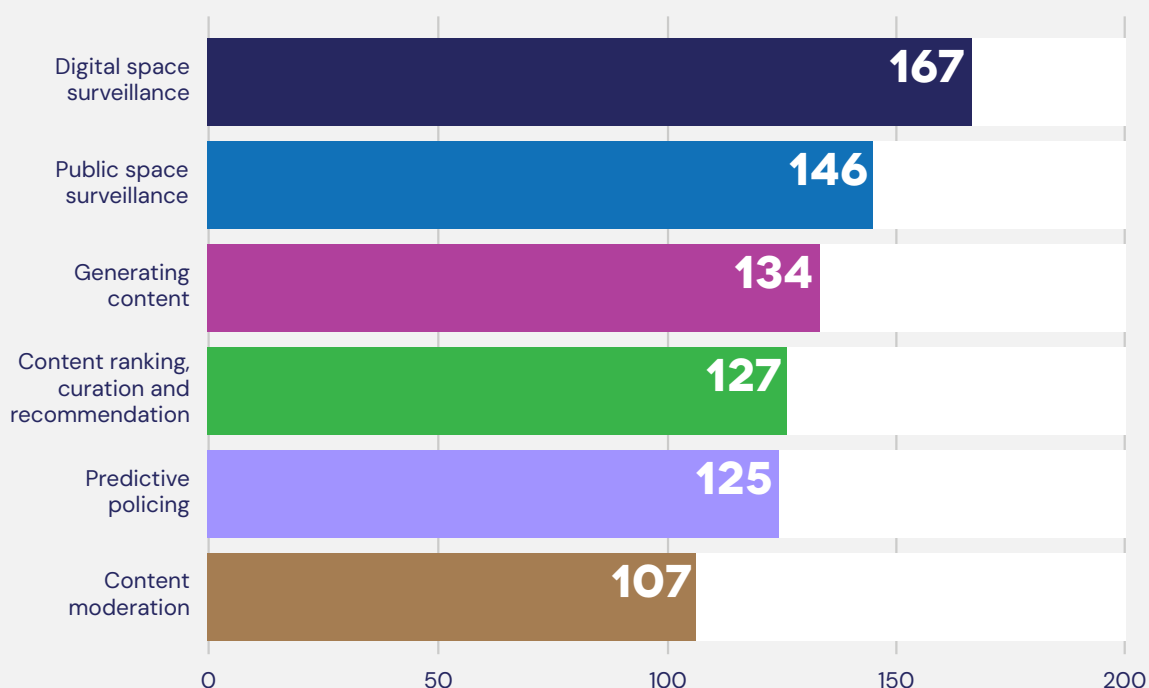
Chapter 1

AI-ENABLED HUMAN RIGHTS THREATS TO CIVIL SOCIETY

As a result of increased deployment of AI in surveillance, civil society actors feel monitored, unsafe, and more afraid to engage in civic and public matters

AI deployment by state and non-state actors and digital platforms pose several threats to civil society around the world. Survey respondents were most concerned about risks emanating from AI in surveillance, specifically within the digital space (57 percent) and in public space (50 percent). Predictive policing was a concern for 43 percent of respondents.

In what areas does AI deployment by the government or the private sector constitute a threat to your organization's work?



A little less than half of respondents stated that they were concerned about AI deployment in generating content (46 percent) and curating, ranking, and recommending content (43 percent). Slightly over a third of respondents were concerned about algorithmic content moderation systems.


57%
of respondents concerned about digital space surveillance


50%
of respondents concerned about surveillance in public spaces


46%
of respondents concerned about AI-generated content


43%
of respondents concerned about content curation and recommendation

This chapter provides an overview of AI systems that undermine civic space and pose a threat to civil society.

How AI is shaping and exacerbating surveillance of civil society around the world

The deployment of AI-powered surveillance and policing technologies pose a serious risk to civil society. Despite calls from civil society to strictly regulate and reign in the deployment of the most invasive AI systems, states around the world continue to acquire these systems from companies looking to profit from lucrative government contracts and deploy them in ways that violate human rights. These technologies include FRT, predictive policing, data analytics, and risk analysis tools and spyware.

Overview of AI-enabled public space and digital surveillance tools and tactics

This sub-section provides an overview of different AI-enabled surveillance technologies that threaten civil society.

Facial Recognition Technology (FRT)

FRT enables deployers to scan facial biometrics of people and compare them against databases of images, often captured illegally or without consent, such as by scraping people's images from the internet.³³ People can be tracked and identified in real time or retroactively, but regardless, FRT is considered a tool of mass surveillance that threatens the rights to privacy, association, and peaceful assembly, and exposes Black,

Indigenous, and People of Colour (BIPOC) to discrimination and bias.

FRT deployment has been increasing around the world despite the severe risks this technology poses to fundamental rights and freedoms and the work of civil society. In Iran, authorities have used it to identify and track protesters, dissidents, and women not adhering to the country's mandatory hijab rules.³⁴ In China, the "world leader" in FRT deployment,³⁵ these technologies are deployed as part of a national surveillance system that constantly tracks people and targets minority groups.³⁶ In Russia, which is second to China in FRT use, the authorities use the technology pre-emptively to stop those who joined protests in the past from joining other protests.³⁷ Other governments in Europe, including those of Hungary, the UK³⁸ and the Netherlands, deploy such technology at protests.³⁹

In India, where FRT use by law enforcement agencies has been proliferating,⁴⁰ data collected from facial recognition systems is connected to a "comprehensive and integrated database of criminal records and crime statistics across the country".⁴¹ In Phuket, Thailand, as part of a safe city project,⁴² the government contracted a UK company to provide "a smart city system with AI-powered video analytics connected to international databases for facial recognition and wanted persons".⁴³

Across Africa, governments are buying CCTV camera surveillance tech with facial and car number plate recognition, with Nigeria, Ghana, and Zambia having each spent over US\$350 million on 'safe cities' mass surveillance programmes from China, according to a 2025 report from the African Digital Rights Network.⁴⁴ Egypt has been building dozens of new smart cities, including a new capital that has a surveillance network of 6,000 facial recognition cameras.⁴⁵ This is viewed as an attempt by the ruling regime to control

the public space and prevent a repeat of the 2011 protests that toppled the Mubarak dictatorship.⁴⁶

In Latin America, the technology's adoption "has advanced quickly and quietly", where it is applied the most is in "public safety" and "surveillance of public spaces" as well as "applications in the framework of transport, social assistance and migration".⁴⁷ In 2024, it was reported that law enforcement agencies in several countries in Latin America and the Caribbean, including Brazil, Colombia, Chile, the Dominican Republic and Trinidad and Tobago were deploying a facial recognition tool developed by a controversial U.S. company, Clearview AI.⁴⁸

O Panoptico: Mapping FRT in Brazil

O Panoptico, a Brazilian civil society monitoring project, has documented 535 facial recognition technology projects across the country — affecting an estimated 97 million people.

The project reveals how FRT deployment has expanded rapidly with little regulatory oversight or public debate.

Predictive policing and border management

Predictive policing relies on AI and data analytics to predict potential crimes or undesirable activity (such as a protest) before they happen.⁴⁹ While it is not new for law enforcement agencies to track and monitor crimes, AI provides them with the capabilities to analyse vast amounts of data. In many cases, this data is collected in illegitimate ways.

Israel⁵⁰ has used algorithms to profile Palestinians under the pretext of preventing terror attacks by analysing vast amounts

of social media data such as a user's social media posts, profile photos, age and hometown to "identify attackers before they act".⁵¹ These systems are already inherently biased, given the Israeli government's criminalisation of Palestinians' peaceful calls for an end to the occupation or their resistance to it.⁵² Authorities in a number of Indian states deploy predictive policing.⁵³ "Delhi's Crime Mapping Analytics and Predictive System (CMAPS) collects data every three minutes from the Indian Space Research Organisation's satellites, historical crime data, and the 'Dial 100' helpline to identify 'crime hotspots' which are crime-prone areas".⁵⁴

Policing data, on which predictive policing software is built, is incomplete or biased, leading to a 'feedback loop', amplifying policing of communities that are already over-policed and face discrimination.⁵⁵ The technology was shown to perpetuate existing biases and forms of discrimination against Black people in the U.S.⁵⁶ These systems are also deployed to predict the likelihood that someone who was convicted of a crime is likely to reoffend.⁵⁷

Predictive policing, risk-assessment and data analytics tools are deployed in border and migration areas, which can not only hinder freedom of movement, but potentially endanger the migration status of people, including activists.

Authorities across the world deploy AI-powered profiling and risk-assessment tools to profile passengers and analyse their data with the purpose of anticipating their risk level.⁵⁸ Based on outputs from these tools, border and migration authorities can decide whether to allow entry to a traveller or subject them to additional security checks. Companies that sell such tools "claim their software can detect terrorists, human traffickers, drug dealers, serious criminals,

missing persons and increasingly, people migrating without papers". For instance, French company Idemia's BORDERGUARD suite combines AI and biometrics for border control. It includes BORDERGUARD Targeting, an AI-based risk-assessment software for governments to "anticipate risks by processing and analyzing passenger data".⁵⁹ In 2023, as Kenya implemented an Electronic Travel Authorisation (ETA) system to replace its previous e-Visa system, it acquired a border security system from a Swiss company called Travizory.⁶⁰ According to the company, its system combines "real-time carrier data, AI-powered profiling and risk-assessment and automated alerts" and "enables proactive intervention and informed decision-making to prevent threats from reaching the border".⁶¹ The government of Kenya reportedly replaced the Travizory system with a different one in 2025.⁶² However, the system has been operational in the Seychelles since 2021.⁶³

Spyware, DPI and phishing attacks

State and non-state actors deploy algorithms in different forms of online surveillance that infringe on civic actors' rights to privacy, freedom of expression and information, and freedoms of peaceful assembly and association. These include Deep Packet Inspection (DPI), phishing, and spyware attacks.

While DPI is a method that can be used for legitimate purposes, such as the detection and blocking of spam, viruses, and child exploitation content, governments use it to censor legitimate speech. For example, the Egyptian government uses it to block access to websites, including those of independent media and CSOs⁶⁴ as well as to virtual private networks (VPNs),⁶⁵ which can help users navigate the internet more privately and circumvent government censorship. Machine learning allows deployers of DPI to be more precise in monitoring and categorising

content to be blocked.⁶⁶

Machine learning is used in spyware attacks by governments to infect the devices of targeted individuals in order to access and steal sensitive information and monitor their behaviours.⁶⁷ Governments around the world deploy spyware to spy on opponents, dissidents, HRDs, and journalists.⁶⁸ Since the 2010s, spyware providers have been integrating machine learning with their software which "has broadened targeted surveillance capabilities".⁶⁹ Machine learning can be deployed at different stages of a spyware attack, making it easier for attackers to identify victims, discover software vulnerabilities to be exploited in zero-day attacks,⁷⁰ and bypass passwords and security tests.⁷¹

In addition, governments and cybercriminals can abuse Generative AI in phishing attacks⁷² to gain unauthorised access to others' personal information. Governments use phishing to target opponents, critics and civil society. Generative AI allows attackers to better research their targets and craft convincing messages. Attackers can also deploy AI to create deepfakes, –hyper-realistic audio, videos, and photos– aimed at impersonating someone the target knows in order to trick them into clicking on a link that contains malware or installing spyware onto their device.

Impacts of AI-powered surveillance on civil society

Civil society groups and activists around the world are concerned about the impact of AI-powered surveillance technologies on human rights and their work, especially given that robust, rights-based regulation of AI still significantly lags behind development and deployment. AI deployment in surveillance and invasive technologies such as FRT, spyware, and predictive policing infringe on privacy, freedom of movement, right to

association and freedom of expression. In the Occupied Palestinian Territories, Israel has for years deployed facial recognition systems to subject Palestinians to constant surveillance and policing and restrict their freedom of movement.⁷³ During the genocide in Gaza, Israel has accelerated its deployment of facial recognition technology, which it uses, along with automated decision making tools,⁷⁴ to identify, track, and select targets for killings.⁷⁵ The Israeli military has deployed machine learning tools, such as Lavender, which “assigns ratings to people in Gaza related to their suspected affiliation with Palestinian armed groups for purposes of labeling them as military targets” and Where’s Daddy?, which “purports to determine when a target is in a particular location so they can be attacked there”.⁷⁶ These tools, which rely on mass (biometric) surveillance, lack proportionality and human oversight, and have contributed to mass killings and the evisceration of civilian infrastructure in Gaza,⁷⁷ affecting everyone including journalists, humanitarian workers, and HRDs. According to the Committee to Protect Journalists (CPJ), at least 259 journalists and media workers have been killed in the conflict (through 30 March 2026),⁷⁸ making it the deadliest conflict for journalists since CPJ started collecting data more than 30 years ago.⁷⁹

“You are being constantly watched through various systems, and they are all collecting so much data because we know the Israeli surveillance regime is extensive. Our phone calls are always tapped; our communications are always going through Israeli ports...There are also the drones that are collecting so much information on Palestinians. There are also the checkpoints, where every checkpoint has a set of cameras. So someone like me as a Jerusalemite, I really worry about how much information they are collecting and how it could affect me in a harmful way. It is the way the occupation works. It is control and domination. So if anything is

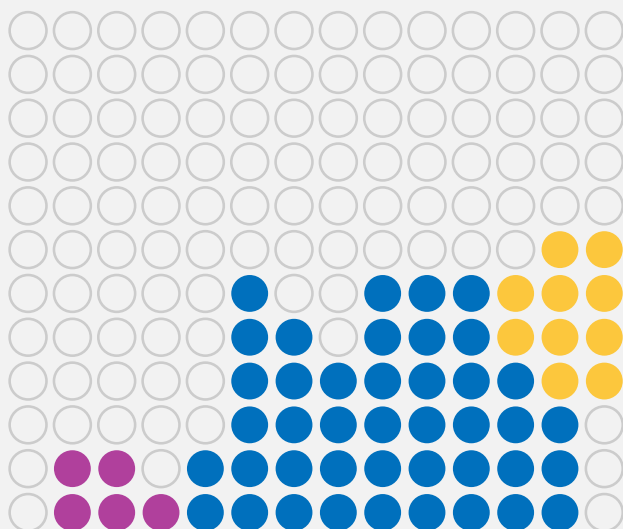
a bit off, according to them, which could be totally legal and normal for me, they use that information to harm us. We fear retribution if we practice our basic rights and civil liberties”, said Jalal Abukhater, Policy Manager with 7amleh, a Palestinian digital rights organisation. He added, “Because of the killing and the genocide in Gaza people are terrified of the idea of communicating over WhatsApp or communicating over certain channels because they know that any sort of information that Israel receives, they have AI systems that are created to just assume that you are guilty and assume that you and your entire family can be bombed”.

“

“Our phone calls are always tapped; our communications are always going through Israeli ports”

— Jalal Abukhater

These systems severely obstruct civic freedoms, and the impacts are disproportionate for excluded groups and individuals. FRTs are more likely to mis-identify people of colour and predictive policing systems are trained on historical racist policing data that are more likely to criminalise historically marginalised groups.



The Pakistan-based disability rights organisation, Spring NGO, reported that:

Predictive policing powered by AI can result in the over-policing or profiling of marginalised groups, including persons with disabilities, under the guise of maintaining security

“Predictive policing powered by AI can result in the over-policing or profiling of marginalised groups, including persons with disabilities, under the guise of maintaining security. Such practices may discourage community members and activists from participating freely in public advocacy events, thereby shrinking civic space and creating an atmosphere of fear. Public space surveillance technologies, such as facial recognition, also threaten privacy and freedom of assembly. These systems often fail to accurately identify persons with disabilities or those using assistive devices, leading to discrimination or misidentification. The knowledge that public gatherings or protests might be monitored using AI-driven tools can reduce public participation in advocacy activities supported by Spring NGO. Digital surveillance through spyware or phishing attacks directly undermines the security of NGOs. As Spring NGO engages with human rights defenders, government bodies, and international partners, AI-powered digital threats could expose sensitive information, jeopardize the safety of activists and erode trust among stakeholders.”

“

“I maintain that, in Chile, AI still makes mistakes in recognition that could endanger people or in the interpretation of attitudes, which could lead to a person being wrongfully detained due to a camera error”

— Survey respondent, Chile

“I maintain that, in Chile, AI still makes mistakes in recognition that could endanger people or in the interpretation of attitudes, which could lead to a person being wrongfully detained due to a camera error”, a survey respondent stated.

The repercussions for civil society activists with a migration background could include a ban on entry to a country or the revoking of their visa or their right of residence. This has been a concern for many Palestinian and non-Palestinian students involved in the Palestine solidarity movement in the U.S. In March 2025, the U.S. State Department announced the deployment of an AI-powered programme called Catch and Revoke, where “algorithms would collect data from social media profiles, news outlets, and doxing sites” to revoke visas of students for their pro-Palestine activism and positions.⁸⁰

“I have lots of public information that is available, if someone wants to look into my name, they can see that Jalal writes about Jerusalem, about life in Palestine and works for 7amleh. This is easily identifiable information, if you go and ask the system, is this person anti-Israel, very obviously, the answer will be yes, but that is not a criminal act, that is not something illegal. The problem with the AI tools that are being used on those borders is that their standard is just anti-Israel, as if it is a crime of thought, as if it is a crime that you are suspected of.

Your freedom of expression to be anti-Israel is unacceptable to them. You are denying people entry based on who or what they could stand for or the views they espouse. So, in a way, as a precaution, you are denying people entry just because of the views they hold, which are not illegal, they are harmless, and they are literally principled expressions, in a sense”, said Abukhater.

Surveillance and policing AI tools infringe on the right to privacy, leading to self-censorship and creating a chilling effect on freedom of assembly and freedom of expression.⁸¹ Israel’s use of predictive policing to allegedly prevent attacks severely infringes on Palestinians’ freedom of expression. According to Abukhater:

“AI does get things wrong many times and using it for predictive policing methods usually puts innocent people in the crosshairs and does not really effectively stop people who actually incite. It is just that people fear that they are all being targeted, and that your speech is being policed, and that you do not have any freedom of expression or freedom of association because AI tools or the assumption that you could potentially pose a threat means that you are already a criminal in the eyes of the occupation. Of course, it is not the way due process is supposed to be, but this is how it is for Palestinians”.

Governments do not have to acquire the most sophisticated and expensive AI tools to conduct surveillance. Readily available data mining tools allow them to easily analyse and observe what people are saying online, as one interviewee who requested anonymity said. For example, through automated scraping of vast amounts of data, law enforcement agencies can analyse and monitor what people are posting online and talking about, thereby obstructing users’ privacy and freedom of expression⁸² Social media monitoring is also deployed to monitor (potential) protests and profile protesters in

order “to learn the identities and affiliations of the organisers, the location and timing of a planned action, and other related information”.⁸³ Use of social media monitoring has been documented in Mexico, Colombia, and the UK, among other countries.⁸⁴

These different AI surveillance systems are frequently deployed in contexts where privacy and data protection laws are weak, and state surveillance is not restricted or subject to independent oversight. In Libya, “the lack of transparency and regulation allows biometric and facial recognition systems to be misused for tracking activists and restricting freedom of expression and assembly”, survey respondent said, adding that, “AI-driven surveillance tools, supported by open-source intelligence (OSINT) practices and spyware, have been linked to monitoring of online activities by law enforcement and armed groups”. One civil society respondent who requested anonymity mentioned that AI deployment in predictive policing and surveillance threatens the organisation’s “ability to organise and mobilise young people freely”, further explaining that: “These technologies can be used to monitor gatherings, discourage civic participation, or unfairly target youth groups. Digital surveillance, including spyware and phishing, creates security risks for our team, volunteers, and community partners—undermining trust and open communication”.

In Africa, “a big problem is that there is no adequate legal framework in place. So, the technologies are being introduced without that framework, without adequate oversight, and even stakeholder engagement. Many are just waking up and seeing things being installed without them being explained... In some countries, there are still no data protection laws and in some countries, and if those laws are there, the institutions are not robust enough to oversee the security agencies”, Victor Kapiyo, a senior researcher

at the Collaboration on International ICT Policy for East and Southern Africa (CIPESA), said in an interview.

“

“A big problem is that there is no adequate legal framework in place. The technologies are being introduced without that framework, without adequate oversight”

— Victor Kapiyo, CIPESA

Similarly, in Latin America, technologies are deployed without proper regulation. “Many countries do not have proper privacy or data protection policies beyond having data protection norms. For example, in Bolivia, there is no personal data protection law. In Argentina, the data protection law is very old, it is 20 years old”, said Juan Manuel Garcia, an independent consultant working at the intersection of technology and human rights, in an interview. In Brazil, where project O Panoptico has so far documented 535 FRT projects in public spaces and potentially more than 97 million people monitored by these technologies across the country (as of 4 April 2026),⁸⁵ the data protection law does not apply to processing “done for purposes of public safety, national defense, state security, or activities or investigation and prosecution of criminal offenses”.⁸⁶ Such a loophole gives authorities broad powers to infringe on people’s privacy rights. Yamin Curzi, a professor at the Center for Technology and Society with the Fundação Getúlio Vargas Law School in Rio de Janeiro, explained in an interview:

“We do not have anything right now in Brazil that actually addresses how data should be treated when collected by police forces, by authorities, by judges, etc. So, with this blank space, the city councils, the mayors, etc. are

signing contracts with private enterprises to deploy CCTV and other facial recognition technologies without any transparency, without public consultations, without accountability. And people are vulnerable”. She added that there have been several cases of people who were misidentified by FRT systems in Brazil, especially people of colour. “But we also know that the systems, they cannot detect properly, for example, trans people and also children, they are also vulnerable”.

“

“We do not have anything in Brazil that addresses how data should be treated when collected by police forces or authorities”

— Yamin Curzi, Brasil

As a result of increased deployment of AI in surveillance, civil society actors feel monitored, unsafe, and more afraid to engage in civic and public matters

One Nigeria-based respondent stated: “AI deployment, especially through state surveillance systems, predictive policing tools and data-driven profiling, interferes with our advocacy work by creating fear and mistrust among communities and activists we engage with. Civil society actors often feel monitored, which limits their freedom to organise, speak and mobilise around sensitive issues such as governance, human rights, and youth inclusion”.

A survey respondent working with the Centre Stage Media Arts Foundation (CSMA) in Zimbabwe, expressed concerns about the abuse of CCTV cameras to restrict the right to protest: “The deployment of cameras in public spaces is not only being used to maintain safety and public order. There is

a real risk that as civic organisations, if we organise demonstrations, cameras are likely to be used for surveillance and violating our rights to freedoms of assembly and expression. This will likely result in further restrictions on civic space”.

Unauthorised access to an organisation’s systems or employees’ devices, for instance, through the deployment of AI-powered spyware and phishing attacks, can expose their information to governments or cybercriminals seeking to commit fraud.

“Hackers are well positioned to exploit your vulnerability in use of AI to sabotage an organisation’s work through hacking into its systems”, noted one respondent. Another, stated: “we collect racism stories from asylum seekers, and participants request for anonymity; we are always afraid AI can expose our collection”.

Further, information extracted in an unauthorised manner can be later used by authorities to prosecute civic society actors or by state and non-state actors to discredit them in disinformation or propaganda campaigns. A Uganda-based respondent, said that: “AI-driven phishing attacks can compromise the security of organisational systems and data, potentially leading to the exposure of sensitive information about human rights defenders, beneficiaries, or partners. This could put individuals at risk of persecution or harm”. An interviewee who requested anonymity highlighted the sensitivity of financial information, when leaked and accessed by unauthorised parties, thereby putting civic activists at risk of prosecution on allegations of “illicit funds” and fraud. For example, Tunisian authorities have used foreign funding as a pretext to crack down on CSOs and independent media in recent years.⁸⁷

An environmental organisation mentioned

that mining companies “are always on the lookout for bad PR against them”. Unauthorised access to sensitive information can be abused to facilitate Strategic Lawsuits Against Public Participation (SLAAPs), legal actions initiated by private companies and individuals to silence critics.⁸⁸

Algorithms used by digital platforms and their implications for civic space

Social media platforms deploy content moderation algorithms at scale to detect and remove content that violates their policies.⁸⁹ Algorithmic techniques in content moderation include hash matching “the process of transforming a known example of a piece of content into a ‘hash’ – a string of data meant to uniquely identify the underlying content” and classification, which “assesses newly uploaded content that has no corresponding previous version in a database” to categorise it and predict whether such content may or may not violate a platform’s policies.⁹⁰

These algorithms can have consequences for human rights and civic space when they result in over-moderation of legitimate speech, infringing on freedom of expression and access to information as well as hindering civil society’s ability to mobilise and reach audiences. In other cases, these algorithms may fail to detect content that is abusive, harassing, hateful or inciting, putting the safety and wellbeing of civil society activists and journalists at risk, and in turn, preventing them from doing their work.

Throughout the years, researchers, journalists, and activists have documented several challenges with algorithmic content moderation. First, there is the influence of socio-political factors and the Global

North—particularly the U.S.— standards in the design and implementation of these systems, resulting in biases that disproportionately impact users in the Global South.⁹¹ For instance, Meta’s dangerous organisations and individuals (DOI) policy, based on which the company instructs its algorithms to remove content by “terrorist” organisations or in support of them, is highly influenced by U.S. foreign policy interests, and those affected by DOI-related removals are disproportionately of Middle Eastern, South Asian, and Muslim backgrounds.⁹²

Second, while algorithms can categorise content, they still struggle with the context in which content is posted.⁹³ For example, algorithms can accurately detect when content is graphic or should not be shown to younger users with high accuracy; however, it may still fail to recognise when violent or graphic content is posted for awareness-raising and educational purposes.⁹⁴ As one respondent noted: “AI-driven content moderation and ranking systems sometimes suppress nutrition advocacy posts that mention hunger or child malnutrition, misclassifying them as ‘graphic’ or ‘sensitive’.” In the aftermath of the 7 October 2023 attacks by Hamas on northern Israel and Israel’s ensuing war on Gaza, Meta’s algorithmic content moderation systems erroneously removed videos for depicting graphic violence.⁹⁵ In another example, in 2020, Facebook and Instagram’s systems accidentally flagged posts about a big protest movement in Nigeria denouncing the controversial Special Anti-Robbery Squad (SARS) as “fake news”,⁹⁶ associating the acronym SARS with misinformation about the COVID-19 virus.⁹⁷

Third, the lack of “human in the loop” in algorithmic content moderation decisions can present challenges in human oversight, transparency, and access to effective remedy.⁹⁸ This makes it difficult for people to

understand why their content was removed and appeal such decisions at a time when social media platforms have come to rely on AI in content moderation.

Another challenge is the lack of high-quality training data to train content moderation systems in low-resourced languages, many of which are spoken in the Global South. In fact, English dominates the field of AI due to the availability of digitised text data in this language to train models.⁹⁹ To address this challenge, platforms train their systems on multilingual, large language models which “infer connections between languages, allowing them to uncover patterns in higher-resourced languages and apply them to lower-resourced languages”.¹⁰⁰ These systems, however, have shortcomings, as concluded in a 2023 study by the Centre for Democracy and Technology (CDT) which found that they often rely on machine-translated text that contains errors and does not reflect local languages. They do not work well in all languages and fail to consider and reflect the contexts of local language speakers; and when problems arise, they are hard to identify and fix.¹⁰¹

The biases, mistakes, and inaccuracies in content moderation are compounded by content recommendation systems, which activists and civil society actors say negatively affect the reach and visibility of their content and expose them to harmful content. “The right-wing movement in the United States is using AI and algorithms (primarily in social media) to their advantage to spread disinformation and radicalise people as they promote hate, bigotry, misogyny, xenophobia, etc”, said a U.S.-based survey respondent who requested anonymity. Another respondent wrote that “social media algorithms often downrank grassroots voices from regions like South Sudan, limiting the visibility of our campaigns”.

Over the past decade, research and investigative reporting have shed light on the way the algorithmic systems used by social media platforms to rank, curate, and recommend content are exacerbating the spread of hate speech, disinformation, and other harmful content. These algorithms are deployed as a part of business models aimed at generating profits from targeted advertising to the platforms and the companies that own them, determining the content users see, including ads, organic posts, recommended pages and accounts to follow, etc.¹⁰² These recommendations are made based on the data platforms collect and infer about users, including their location, language(s) they speak, age, profession, previous content they engaged with, interests, likes and dislikes, etc. This means that users are more likely to be recommended content that they would engage with so that they stay longer on the platforms, which can thus generate even more information for platforms to tailor further recommendations and targeted ads.

By prioritising engagement over quality of content, these algorithms often end up making sensationalist content,¹⁰³ including sexist, racist, and hateful rhetoric go viral. This can lead to content that undermines civic space becoming more visible for users such as misogynistic influencers and the anti-rights movement, both of which spread disinformation and hate that impacts women, feminist groups, and LGBTQIA+ people, among others. In addition, those subjected to online attacks may retreat from civic space and public life, self-censor or adopt lower profiles.¹⁰⁴ In a 2023 report exploring the negative impacts of gendered disinformation, a form of Tech-Facilitated Gender-Based Violence that is particularly deployed to target and silence women and -non-binary activists, HRDs, protesters and journalists, the UN Special Rapporteur on the promotion and protection of the right to freedom of

opinion and expression highlighted the role of algorithms in spreading this type of content: “Algorithmic recommender systems play a major role in exacerbating the problem. Adversarial narratives exploit the human tendency towards negative content and disproportionately drive engagement on platforms. Algorithmic newsfeeds craft automatically generated, highly personalized adversarial content streams that keep users engaged on the platform, and monetized, and that in the end corrupt the entire global information ecosystem”.¹⁰⁵

The weaponisation of Generative AI to spread disinformation and target and discredit civil society is another threat to civic space. A Taxonomy of Human Rights Risks Connected to Generative AI, by the UN B-Tech Project, identified a number of risks that are relevant to civic space, including its use to create and spread disinformation, “inciting targeted physical violence against specific individuals or groups and non-consensual sexualised content”, and “derogatory or otherwise harmful outputs pertaining to people with marginalised identities”.¹⁰⁶

The authors also noted: “Generative AI may be leveraged to create false but convincing content presented as authentic for the purpose of misleading people. Generative AI’s capacity to create this content at scale can make online abuse campaigns easier to carry out, by both private and public actors. Such campaigns may successfully harass targets—including journalists, activists, and political actors—into self-censorship. Female public figures are especially at risk of targeted online harassment”.

One form of AI-generated content that has been particularly harmful is the spread of deepfake sexual abuse, which disproportionately targets women. This is exemplified when the Grok AI chatbot, developed by Elon Musk’s xAI, was abused on a wide scale to generate fake and

nonconsensual nudes of women and children and spread them on X, the social media platform he owns.¹⁰⁷ A form of gendered disinformation,¹⁰⁸ deepfake sexual abuse, particularly targets women who are visible online, such as journalists, HRDs, feminist activists and politicians, to discredit and silence them and exclude them from participating in political and civic life.

“For women human rights defenders in Tanzania, in Uganda, the women candidates, vying for elections in Uganda, image-based abuse is the tactic that is now draining their swamp in terms of political participation”, said Irene Mwendwa, a Nairobi-based consultant on technology, AI, and law. She noted that, while moderation of text-based content moderation in some African languages has improved, “non-consensual image sharing is still very, very high, and there is very limited knowledge or understanding about pulling some of this content down”. She added, “the implications are higher, so that is what is really happening, and women human rights defenders have really tried to hold training opportunities, but also to hold some of these tech companies accountable. But, from the Big Tech side, the content moderation and public policy and human rights departments were all slashed down”.

“

“For women human rights defenders in Tanzania, in Uganda, the women candidates vying for elections, image-based abuse is the tactic that is now draining their swamp in terms of political participation”

— Irene Mwendwa, Nairobi

During protests and conflicts, where the role of civil society is even more crucial in disseminating factual information, including

life-saving information, documenting violations and holding powerful actors accountable, abuse of GenAI and platforms by state and non-state actors disrupts information flow and access and thereby threatens civic space.

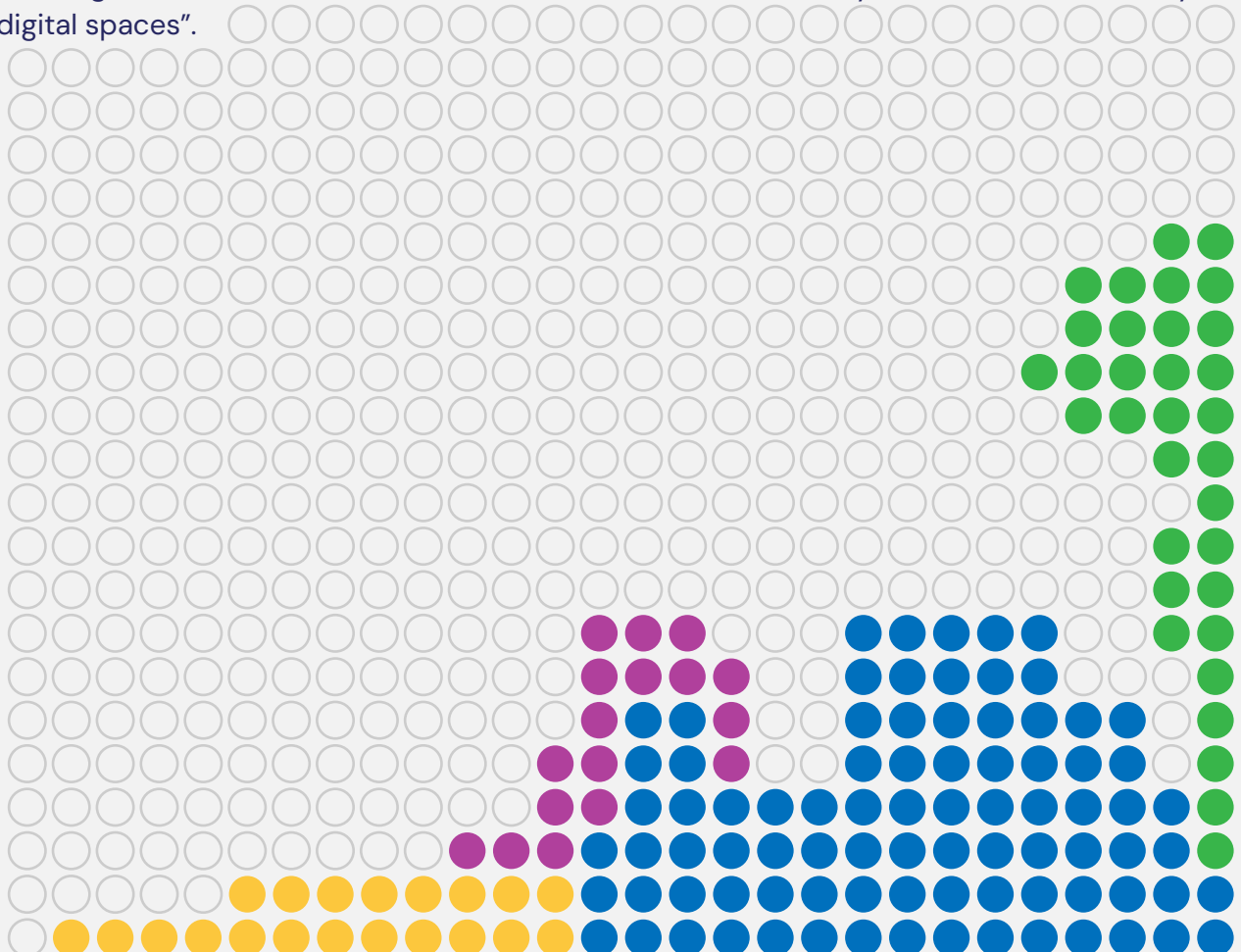
Victor Kapiyo, senior researcher at the Collaboration on International ICT Policy for East and Southern Africa (CIPESA), acknowledged in an interview that there have been positive uses of Generative AI by civil society, which “has made it easier to edit things, to advance its message around and do its work. But when we have groups that are organised to amplify wrong or misleading narratives, then it becomes a problem”. He explained, “AI-generated images, algorithms have been leveraged by different actors to target and ridicule or diminish the work of civil society. So, in some countries, we saw campaigns that targeted civil society to discredit their work. Also, from a civic space perspective, the resulting threat to information integrity, in the sense that spreading misinformation about key topics during elections and conflicts have become rampant. [There are] organised groups that are paid by politicians or other political actors to interfere with information”. He specifically referenced the spread of AI-generated images during anti-finance bill protests in Kenya in 2024. Some of these images were spread to manipulate public opinion. According to researchers at the Digital Forensic Lab (DFRLab), AI tools, including GenAI, were deployed by a “network of X accounts [that] appeared to operate in coordination, amplifying several hashtags that attacked protestors and boosted support for President William Ruto and the government”.¹⁰⁹

Many survey respondents said they were concerned about how these GenAI tools could be used to distort narratives about their work and manipulate public debate.

“Synthetic media and deepfakes increase misinformation, which can distort our narratives, harm our credibility, and weaken public trust in civic actors like us”, one respondent said. Another, based in the United Kingdom wrote: “AI systems can generate inaccurate or misleading information, which may distort the narrative about our organisation’s activities or goals, potentially undermining our credibility with the public and policymakers”.

By suppressing civil society content on platforms and exacerbating the spread of harmful content, these algorithmic systems are shrinking opportunities for peaceful expression, exchange of ideas, and access to information in the digital space. This is disruptive to the work of civil society around the world.

A Nigeria-based respondent recounted that “algorithmic bias in digital platforms can reduce the visibility of advocacy campaigns or misclassify human rights content as “sensitive” or “political”, thereby suppressing civic voices online. AI-generated misinformation and deepfakes further complicate advocacy efforts by spreading false narratives that can discredit organisations like ours or undermine genuine social justice campaigns”. A Ghana-based organisation responded: “Since we are a small startup and have not deployed AI tools ourselves, our main concern lies in how external AI systems—like social media algorithms and content recommendation tools—affect our visibility and outreach. These systems can sometimes limit the reach of our posts or misclassify our content, reducing engagement with our target audience. Additionally, the rise of AI-generated content and deepfakes poses a risk of misinformation, which can undermine public trust in genuine initiatives like ours. These challenges make it harder to communicate our mission effectively and maintain credibility in digital spaces”.





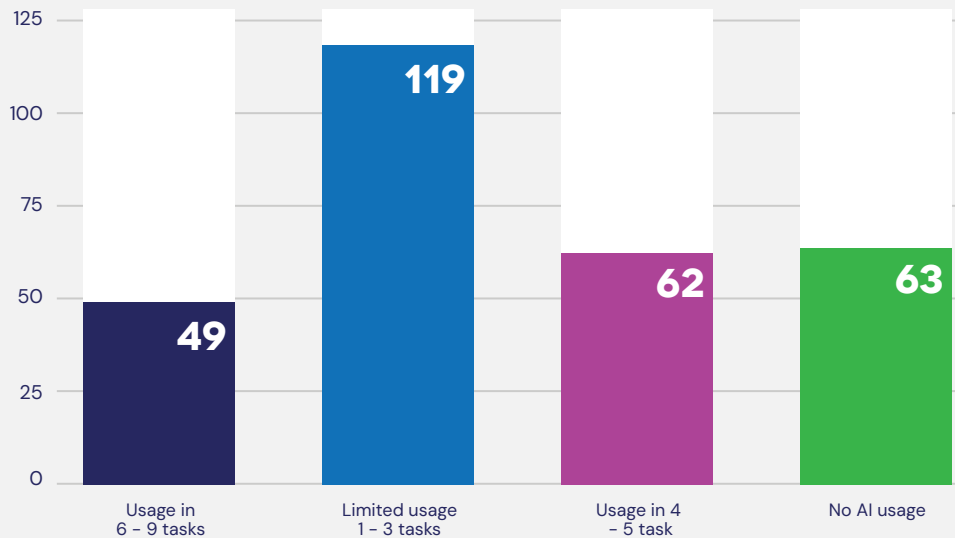
Chapter 2

CIVIL SOCIETY DEPLOYMENT OF AI:

Hurdles to adoption and
navigating ethical questions

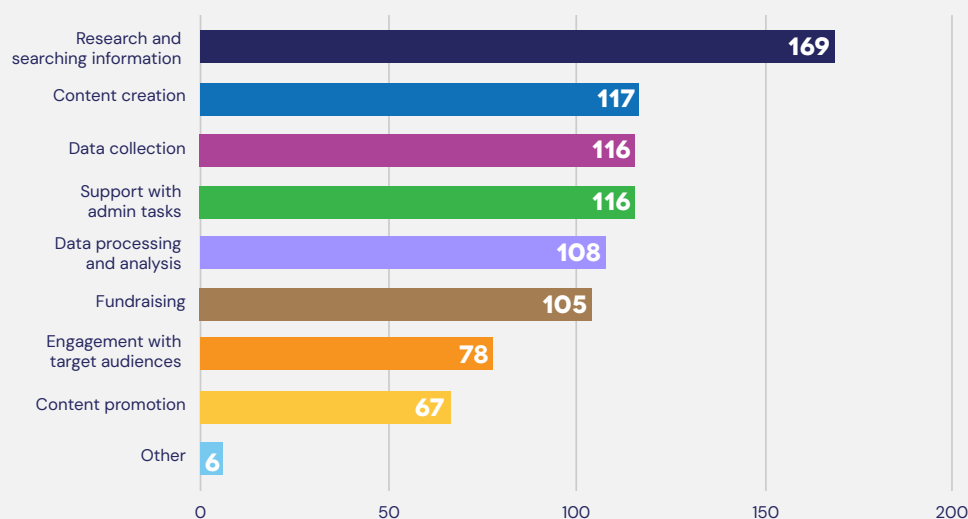
Out of 293 surveys received, 63 respondents (21.5 percent of all responses) said that their organisations do not deploy AI at all. Of those that said they deploy AI tools, over half use them in very limited ways, mostly to assist with research tasks. Only 49 respondents reported that their organisations deploy AI tools more broadly, namely to assist them with a variety of tasks (upwards of six to nine tasks) such as research, content creation, data collection, data processing and analysis, and fundraising.

Scope of AI usage



Of the organisations that deploy AI, three quarters use it for research. Just over half use AI tools for content creation, data collection, and administrative tasks such as note-taking during meetings, calendar management, and drafting emails. Slightly over 45 percent of respondents said AI tools provide their organisations with support in data processing and analysis and fundraising tasks such as writing proposals.

How CSOs use AI



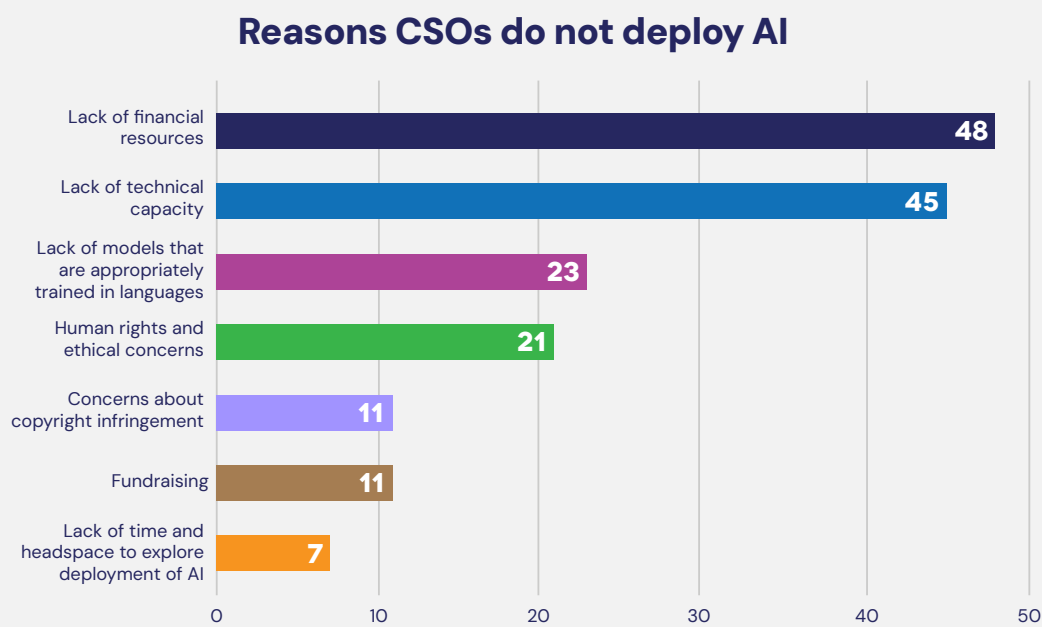
While deployment of AI is still limited, its development is even more limited. Only 37 respondents said their organisations have experience developing their own AI solutions such as chatbots, AI models for data analysis or building training datasets.

Examples of development based on survey responses, desk research, and the focus group include:

- An internal AI-powered digital engagement and research tool built by the Roberto Save Dreams Foundation in Angola to support policy research and data analysis on climate change, gender equality, and rural development, as well as communication for awareness campaigns and stakeholder mapping.
- In Türkiye, the Centre for Democracy Research (CDR) developed Policyful,¹¹⁰ an AI-powered platform that allows its users to analyse and interpret policy content interactively and access a regulatory archive covering the Turkish Parliament, ministries, regulatory agencies and municipal councils.
- In Jordan, the Jordan Open Source Association (JOSA) built Nuha,¹¹¹ an open-source model for detecting and classifying online gender-based hate speech in different Arabic dialects.
- Independent Tunisian media outlet *Nawaat* launched *Nawaat Chronicles*,¹¹² an AI-enhanced archive platform allowing readers, researchers, students and journalists to easily access and search an archive of over 22 years of independent journalism on Tunisia in Arabic, French, and English.
- Independent media outlet *Raseef22* launched Ask Aunty, an AI-powered chatbot that enables readers to access accurate information on sexual and reproductive health and rights (SRHR) in Arabic.¹¹³
- In Nigeria, an organisation mentioned that it developed Value for Money AI, a tool that “automatically analyses audit reports, detects corruption indicators, and drafts legally sound petitions for submission to anti-corruption agencies”. The organisation said that: “Our motivation for developing the solution is to address the inefficiency and complexity of drafting and submitting corruption-related petitions in Nigeria. Many citizens, journalists and activists struggle to transform audit data into legally sound petitions. By automating this process using generative AI, we enhance accessibility, accuracy and efficiency, empowering stakeholders to fight public sector corruption effectively”.
- Ekwere Akpan Foundation stated that it developed EAF Insight, “an AI-powered data dashboard designed to analyse community needs and monitor project outcomes in areas such as education, poverty reduction, and women’s empowerment” to “enhance data-driven decision-making, improve transparency in project reporting, and strengthen collaboration with partners such as UN agencies and local governments”.
- South African organisation GRIT built Zuzi, an “African and trauma-informed” AI chatbot to provide information on the rights and legal services available to survivors in addition to SRHR issues.¹¹⁴

Obstacles to AI deployment

A combination of factors prevents organisations from deploying AI effectively or deploying it at all. Of the 63 surveyed respondents that do not use AI at all in their work, a lack of financial resources and technical capacity were the most commonly cited factors. Slightly over a third of respondents cited challenges such as the lack of models in languages and contexts related to their work as well as human rights and ethics concerns.



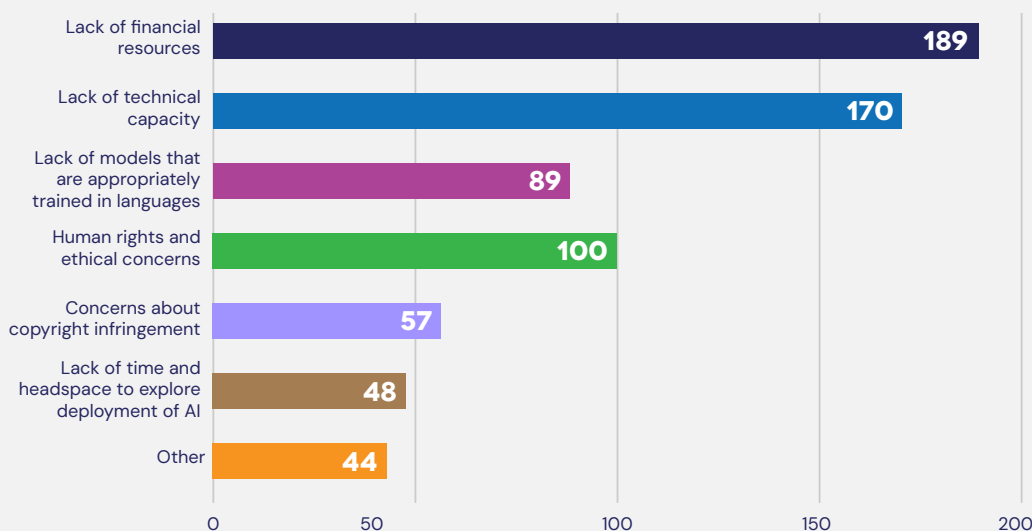
Regardless of whether organisations deploy AI or not, the lack financial resources and technical capacity are major obstacles to deployment.

One respondent working with a South Sudanese organisation that deploys AI in a variety of tasks—data processing and analysis, content creation and promotion, engagement with target audiences, support with administrative tasks and fundraising— wrote that AI “is literally not familiar to us”, particularly when it comes to development of chatbots and apps for engagement with target audiences. The respondent highlighted the “need [for] complete package training” to be able to deploy AI tools “without difficulties”.

Another respondent working with an organisation that deploys AI in very limited ways in research wrote: “The major obstacles preventing our organisation from developing and deploying Artificial Intelligence (AI) solutions are primarily financial constraints and limited technical capacity. As a nonprofit organisation, most of our resources are directed toward programme implementation rather than advanced technological investments. We lack the funding required to procure AI infrastructure, software and specialised expertise. Additionally, our staff have limited exposure to AI tools and applications, making it difficult to integrate

AI into our operations without targeted training and technical support. Overcoming these barriers would require both capacity-building and financial investment to harness AI for social impact effectively”.

Obstacles to AI deployment



The constraints mentioned above can be interlinked. CSOs often lack financial resources to hire professionals who can support them with exploring the possibilities offered by AI tools, as well as integrate such tools into their workflows with policies and measures to ensure AI adoption does not result in bias, inaccuracies, privacy violations and other ethical and human rights infringements. This reality has been compounded in recent years by large-scale foreign aid budget cuts that have forced many organisations around the world to scale back or shut down operations.¹¹⁵ This has further limited the resources available for employees, making it a challenge to “balance daily programme demands with learning and experimenting with AI...due to time constraints and competing priorities” as one respondent noted. Another respondent, whose Guatemala-based organisation does not deploy AI at all, stated: “Human rights organisations like GAM are facing a big crisis related to funding, and this generates a big gap in getting more information about how to use AI tools to strengthen staff knowledge”.

In some cases, finding people with necessary technical skills is a challenge. One respondent, who works in Ethiopia, noted that “the most significant challenge is limited financial resources, which makes it difficult to invest in advanced AI infrastructure and high-quality computing tools. In addition, there is a shortage of local expertise in AI programming and data science, making it necessary to rely on external consultants for technical support”.

“As a non-profit organisation primarily focused on civic and climate work, we face constraints in terms of in-house AI expertise. Developing and managing AI models requires specialised knowledge in data science, machine learning, and software engineering. These skills are not always readily available within the civic sector,” a respondent from Kenya said.

Financial resources are also needed to acquire the necessary hardware for organisations to develop their own tools.Spring NGO in Pakistan said: “Developing or adapting AI solutions

requires investment in software, hardware, and skilled professionals, which has not yet been feasible within the organisation's existing funding structure. Most available grants focus on direct programme delivery rather than technological innovation". Some respondents mentioned bad overall infrastructure, such as poor internet connectivity and lack of cloud computing and data storage and management infrastructure, which prevent them from deploying AI in their work. One respondent working in South Africa, noted that "reliable internet connectivity, high-performance computing resources and stable power supply are not always available—especially in remote or developing regions where our work is primarily focused".

Respondents also noted the lack of models that reflect the contexts in which they work as another obstacle. "Many AI models are not trained to reflect local languages, cultural contexts, or the lived realities of marginalised groups such as LGBTQI+ persons in Nigeria, making them less useful or even biased in our setting. We also have ethical concerns around data privacy, misinformation and the possibility of algorithmic bias, which could harm the very communities we seek to protect", one respondent wrote.

One respondent, whose organisation "is in the early stages of developing AI-driven advocacy and research systems" said that "we are cautious regarding ethical implications and language representation. Most AI tools lack training in Arabic and MENA cultural contexts, limiting fair accessibility and representativeness".

"Our main limitation is access to locally contextualised models trained on African nutrition and behavioural datasets. Current tools often reflect Global North biases and lack sensitivity to local nutrition, and language and gender dynamics", Nigeria-based respondent said.

Centre Stage Media Arts Foundation (CSMA) in Zimbabwe explained that in addition to being constrained by limited human, technical, and financial resources, language is an obstacle: "We are unable to deploy free AI tools since they use English mostly, yet our audiences are more conversant in local languages".

In addition, the lack of high-quality data availability on which models can be trained is a major hurdle for civil society actors to develop their own models instead of relying on dominant models. This lack of data becomes particularly scarce for low-resource languages and for excluded groups. "For AI to function effectively, it requires large, high-quality datasets. In our case, linguistic and cultural data (especially for under-documented or oral languages) are limited or scattered, which makes model training difficult", said the same respondent.

This is a sentiment echoed by disability rights organisation Spring NGO in Pakistan: "access to reliable data remains a barrier. Effective AI deployment depends on large, high-quality and well-structured datasets — something that is often unavailable in the disability and inclusion sectors in Pakistan. The lack of standardised, disaggregated data makes it difficult to build or train AI systems that could meaningfully support advocacy work".

"In many of the communities we serve, especially in rural or marginalised areas, there is limited access to reliable data. Additionally, existing data may be fragmented, outdated, or even not collected in formats suitable for AI processing", said a respondent who works in the DRC.

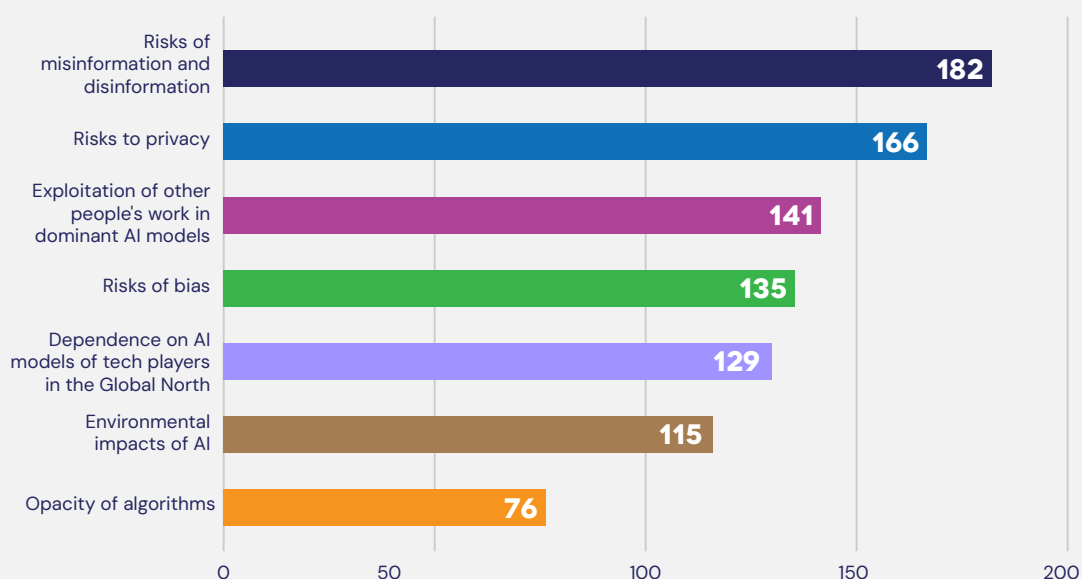
Finally, sometimes AI use can be perceived negatively, which may deter deployment. One respondent from Sri Lanka wrote: "Normally when we are writing proposals we use AI like ChatGPT to develop our proposals, once we

give our own input, it refines our English. But, the issue is, once it is submitted the funders find it as AI generated as the wordings are from AI though the idea is from our side. So, though we like to use AI, the tendency of rejecting our proposal is high due to recognition of AI use". Another respondent from Nepal added that due to the multiple risks they fear public backlash: "Currently there are too many opaqueness about AI models and their algorithms that prevents us from providing the models [with] our data which we do not know how [it] will be used and stored. Also we do not want to misuse other's copyright materials in our works without proper attribution, and we definitely do not want to lose public trust in our institution due to AI".

Human rights and ethical concerns associated with AI deployment by civil society

Civil society globally has expressed concerns over human rights and ethical questions related to AI use. Nearly a third of respondents cited human rights and ethical concerns as one of the obstacles that has prevented their organisations from effectively deploying AI or deploying it at all. Respondents were particularly concerned about exacerbating the spread of misinformation and disinformation (62 percent of all respondents) and the risks to privacy associated with the design and use of AI models and systems (56 percent).

Ethical concerns about civil society deployment of AI



When it comes to misinformation and disinformation, concerns include the use of Generative AI models, which can generate outputs that may contain inaccuracies, biased information or hallucinations. Inaccuracies and biases in Generative AI outputs can thus contribute to undermining information integrity and public debate.

"Generative AI models can produce convincing but false or misleading content, which can easily be weaponised to spread propaganda, manipulate public opinion, or discredit CSOs. This is especially concerning in contexts where information ecosystems are already fragile. The speed and scale at which AI-generated misinformation can circulate threaten both democratic discourse and public trust in legitimate information sources", one respondent noted. "AI-generated content may contain inaccuracies or outdated information, which could

be misleading or harmful if disseminated to the public or used in advocacy efforts”, another respondent said.

An online environment of information disorder makes it challenging for civil society actors to spread their messages. As one anonymous respondent noted, “the growing problem of misinformation generated by AI could weaken public trust in advocacy messages and make it harder for people to distinguish fact from manipulation, a risk that can undermine human rights work and advocacy credibility”.

Respondents expressed concern about the way existing and dominant models were developed—through massive data collection including personal data without informed consent—and thereby potentially exposing personal information, in inputs into AI models. According to one respondent, “many AI systems rely on large amounts of data, including personal or sensitive information. This raises serious concerns about how such data is collected, stored and used, especially in environments where digital surveillance is common and data protection frameworks are weak. There is a real danger that AI systems could unintentionally expose or misuse information belonging to activists, journalists, or human rights defenders, putting their safety and privacy at risk. Ensuring data security, consent and transparency in AI operations is therefore a top ethical priority for us”.

“

“Our main worry is privacy — most AI tools collect and process large amounts of data, and in a country like Nigeria where LGBTQI+ people face discrimination, any misuse or exposure of personal data could put lives at risk”

— Anonymous respondent, Nigeria

Nearly half of the respondents were concerned about exploitation of other people’s work in dominant AI models (for example, the training of models using copyright protected work without consent of authors or compensation) and risks of bias in the design and outputs of AI models and systems. Some organisations also cited the risk of dependence on AI models designed and controlled by private tech players in the Global North (44 percent).

The risk of bias is attributed to dependence on Global North models that do not reflect realities and contexts in which respondents and their organisations work. It is also attributed to existing systems being trained on biased data and reflecting existing inequalities. These risks can particularly affect excluded groups such as LGBTQI+ people and persons living with disabilities.

“The organisation’s foremost concern is that AI technologies, if not developed and implemented responsibly, may unintentionally reinforce existing inequalities and biases—particularly those affecting persons with disabilities. Many AI systems are trained on datasets that under-represent or misrepresent marginalized groups”, said Spring NGO in Pakistan.

“Many AI models are trained using data that does not reflect the realities or voices of African or queer communities. This can lead to distorted or harmful representations that reinforce stigma instead of supporting equality. Another ethical issue is the dependence on AI tools developed and controlled by big tech companies in the global north. These systems often lack transparency (“black box” algorithms) and are not accountable to local civil society needs”, another respondent said.

“We are concerned that most AI tools are trained on data from contexts that do not

reflect African realities, which can perpetuate bias in advocacy analysis and misrepresent local voices. Additionally, privacy and data protection remain critical challenges in South Sudan where regulatory oversight is weak”, stated a respondent.

For some, the dependence on dominant models is seen as an ‘inequality issue’ that only advances the interests of very few powerful companies. For one respondent based in Honduras, “one of the main ethical concerns about the use of Artificial Intelligence relates to the displacement of human interests and rationalities by an algorithm that ultimately responds to the interests and priorities of a small group that often does not even know the priorities of the majority”.

“Exploitation occurs when creators’ copyrighted work is used to train AI models without consent or fair compensation. The high cost of developing these systems is an inequality issue, concentrating power and access within a few wealthy organizations, which in turn reinforces biases and compromises safety”, noted another respondent.

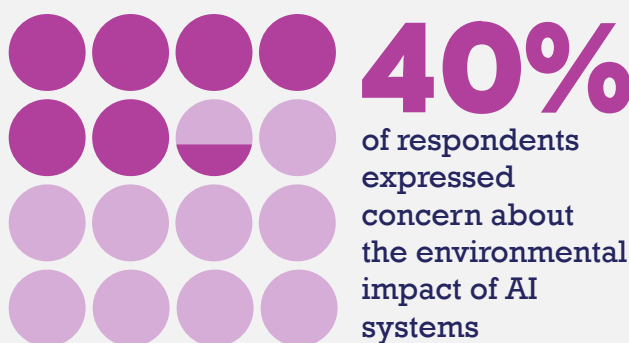
“

“The high cost of developing AI systems is an inequality issue, concentrating power and access within a few wealthy organizations”
— Survey respondent, Honduras.

The climate and environmental impacts were also a concern for 115 survey respondents, nearly 40 percent of all respondents.

One respondent said: “Artificial Intelligence (AI) systems have significant environmental impacts throughout their lifecycle. Training

large models requires massive energy, often thousands of megawatt-hours, leading to high carbon emissions—sometimes exceeding the lifetime emissions of several cars. Continuous operation for [AI] inference¹¹⁶ also consumes large amounts of electricity. Data centres use vast quantities of water for cooling, contributing to resource stress, especially in arid regions. The production and disposal of AI hardware creates electronic waste and demand for rare minerals like lithium and cobalt, which have their own environmental and social costs. Additionally, data storage and transmission add to energy consumption. These impacts contribute to climate change, resource depletion, and environmental inequality.”



Another respondent working with the World Diplomacy Organisation (WDO), wrote in response to the survey that “as an institution that promotes sustainable development, responsible governance, and ethical innovation, WDO believes that the rapid expansion of AI technologies must be balanced with environmental responsibility. The organization advocates for greener, energy-efficient AI models and greater transparency from technology providers regarding the carbon footprint and resource consumption of AI infrastructure”.

There are also concerns about the lack of human oversight and accountability in decisions around AI systems and their outputs. Increased reliance on AI to make decisions and find information could erode

human judgement,¹¹⁷ agency, empathy and creativity and other human qualities, which are necessary to the work of civil society. One survey respondent who chose to remain anonymous wrote: “Civil society is often the last line of defense for marginalised groups. Its strength lies in human empathy, contextual understanding, and the ability to challenge unjust systems. Our deep concern is that an over-reliance on AI could lead to the ‘automation of compliance’, where human workers simply follow the AI’s recommendation without question. This erodes the essential human judgment and moral reasoning that is the bedrock of effective civil society work. When a flawed decision is made, who is accountable? AI? The developer? The CSO that deployed it? This accountability vacuum is a serious ethical problem. An AI does not ‘know’ it is replicating past redlining or gender discrimination; it just finds patterns in the data it was given. The ethical failure occurs when this output is accepted as fact without critical scrutiny”.

“There are cases where people blindly trust the answers provided by AI without analysis or comparison with reality”, said a respondent from Bolivia. One respondent stated that “our organisation has not yet deployed AI primarily because of concerns about its limited ability to authentically connect with human emotions and relationships”, adding that: “We place high value on genuine human interaction, empathy and contextual understanding—qualities that current AI systems often struggle to replicate with originality and depth”.

Whether organisations deploy AI or not, respondents who took part in the survey generally showed an awareness of the risks. They highlighted the need to deploy AI responsibly and with sufficient safeguards to prevent any harm.

One respondent from Kenya noted the following: “We believe that civil society has a responsibility not only to adopt new technologies but also to advocate for safeguards, policies, and practices that ensure AI serves the public good, especially in areas like climate justice, civic participation, and social equity. We are particularly interested in collaborative approaches that centre community voices in the design and implementation of AI tools, ensuring that technology amplifies, not replaces, human agency and democratic processes”.

In some cases, organisations decide not to deploy AI tools due to ethical and human rights concerns.

A respondent who works with an Ethiopia-based organisation that does not deploy AI said that while the organisation STEM Plus “recognises the transformative potential of Artificial Intelligence, deployment is hindered by unresolved ethical, legal, and human rights concerns”. The respondent further added that: “we actively consider potential ethical and human rights implications in all our digital initiatives. Our approach is precautionary and principled, ensuring that when AI becomes part of our programs, it is introduced responsibly, transparently and inclusively. We believe that sustainable technology must not only advance innovation but also protect and uplift humanity”.

These human rights and ethical concerns do not necessarily equate to disinterest from CSOs in deploying AI but rather show that integration should be done responsibly. A Colombia-based respondent stated the following: “We are interested in exploring artificial intelligence, but we are also concerned about how it might reinforce existing inequalities. Most models reflect perspectives and languages foreign to our contexts, which limits their value for working with local communities. There

are also concerns about data protection, the transparency of algorithms, and the environmental cost of these technologies.



“The question is not just whether to use AI, but how to do so without losing autonomy or ethical purpose.”

The next section explores how organisations address ethical issues and human rights risks in their deployment of AI.

How CSOs attempt to navigate human rights risks and ethical questions in their deployment of AI

Despite the concerns noted above and general awareness among respondents about the risks of AI deployment, not enough actions are taken by CSOs to implement policies and measures to ensure that AI use does not infringe on human rights. Of the 230 respondents who said that their organisations deploy AI, only 38 percent clearly stated that they consider and attempt to address ethical concerns and human rights risks.

There are two factors that could explain this trend. First, for those organisations that deploy AI tools, its usage remains very limited, while only a small fraction of organisations have developed their own models. Another factor is the lack of technical knowledge and resources CSOs need to address the identified risks and adhere to ethical guidelines when deploying AI.

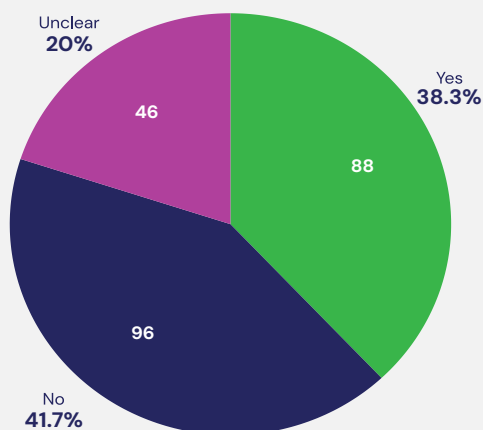
One respondent stated that: “AI is a relatively new technology, about which we have little

information regarding its nature and risks; however, the issue of privacy security is one of the biggest fears.” A respondent whose organisation does not deploy AI explained, “What is preventing our organisation from using AI is mainly the lack of knowledge on the subject and the necessary funding. We want responsible AI that does not violate human rights, an AI that is useful for our work and our community. The biggest challenge for us is the various risks and impacts associated with using AI. We won’t know how to prevent them because we have very limited knowledge”.



of CSOs that deploy AI have considered and attempted to address ethics and human rights risks

Respondents who said their organisations consider and attempt to address potential ethical and human rights risks in AI deployment



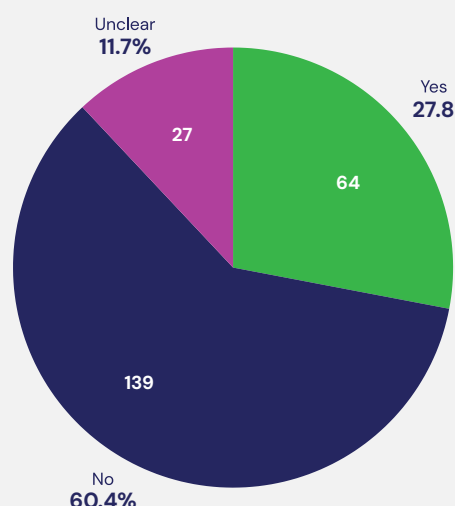
Overarching policies and principles guide deployment of AI for some of the organisations that took part in the survey. In particular, these emphasise respect for data privacy, protection from bias and non-discrimination, fairness, inclusivity and inclusion, transparency and accountability, informed consent, human dignity, protection of vulnerable populations and marginalised communities, gender-sensitivity and respecting copyrights.

A respondent, based in Bangladesh, said: “We approach all forms of AI use through a strict ethical lens. Our primary focus is to ensure data privacy, avoid algorithmic bias and maintain respect for human dignity. Even in early stages of AI integration, we consult relevant international human rights frameworks and prioritize inclusive, safe and accountable practices. If we adopt or develop AI solutions in the future, a dedicated ethics oversight mechanism will be established”.

Several organisations mentioned that while they still do not deploy AI, if they would in the future, they plan to ensure adherence to ethical principles and human rights standards. “As the organisation explores integrating AI into program design and monitoring, it

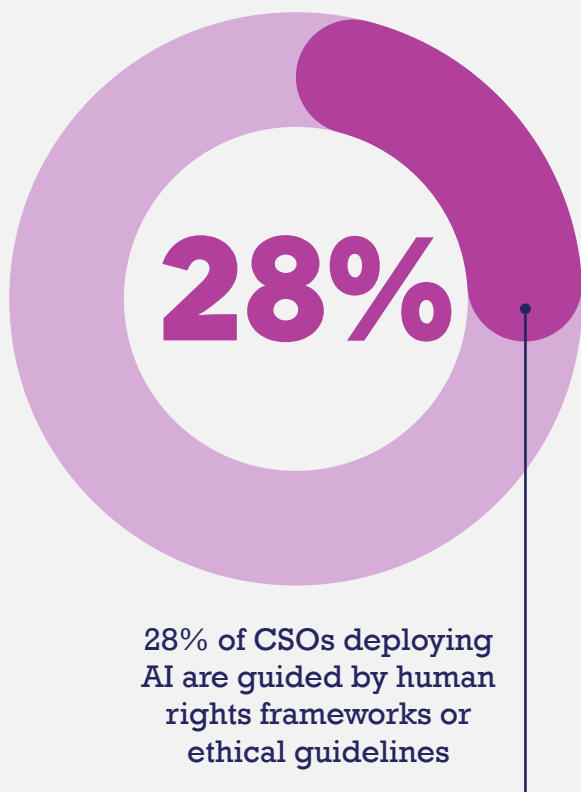
remains committed to aligning with ethical and human rights standards. This includes ensuring data privacy, obtaining informed consent, promoting gender-sensitive algorithms and avoiding bias or harm to vulnerable populations. WGHEN intends to collaborate with ethical AI experts and adopt global best practices to ensure that any future AI solutions uphold transparency, accountability, and inclusivity in line with human rights principles”, stated a respondent.

Is your organisation’s development or deployment of AI guided by any specific human rights frameworks or ethical guidelines?



Of respondents whose organisations deploy AI, only 28 percent clearly stated that such deployment is guided by specific human rights frameworks or ethical guidelines. These include the African Union Artificial Intelligence Continental Strategy,¹¹⁸ OECD AI Principles,¹¹⁹ organisations’ internal policies, the UNESCO Recommendation on the Ethics of Artificial Intelligence,¹²⁰ and to a lesser extent the UN Guiding Principles on Business and Human Rights.¹²¹ Some organisations also listed local laws and policies such as national AI strategies and data protection laws. To help implement overarching policies and principles, organisations mentioned putting in place practical measures. The following

sub-section provides an overview of these measures based on survey responses and insights from the focus group discussion.



Practical measures for responsible AI deployment

Personal data protection measures

Protection of privacy and personal data is of utmost importance for organisations deploying AI. To protect personal data, some CSOs said they refrain from inputting identifiable and sensitive information into AI models, whether public systems or internally developed ones, or they anonymise such data.

A survey respondent who requested to be anonymous wrote: “We try to be mindful of ethical and human rights risks whenever we use AI tools, even though we have not developed our own AI model yet. Our organisation works closely with vulnerable LGBTQI+ individuals, so we prioritize data privacy, confidentiality and consent in everything we do. For instance, we avoid

uploading sensitive or identifiable information about beneficiaries into AI platforms and only use trusted tools for research, learning and communication”.

For its Ask Aunty chatbot, *Raseef22* minimised personal data collection from the start. “We do not require any accounts or people to log in [to use the chatbot], and we do not ask them to give us any information. The only information that we store is the country – just for analytical purposes. We do not collect any personal identifiers, and we do not share the information with any third parties. So, our team is the only one who can see the queries and the questions on the chatbot”, a *Raseef22* employee explained during the FGD.

Raseef22’s Ask Aunty: privacy-first design for sensitive health information

For its Ask Aunty chatbot, *Raseef22* minimised personal data collection from the start. “We do not require any accounts or people to log in, and we do not ask them to give us any information. The only information that we store is the country — just for analytical purposes. We do not collect any personal identifiers, and we do not share the information with any third parties.”

One FGD participant whose organisation deploys AI internally to gather and organise data mentioned that they limit the tool’s access to the organisation’s systems and private information. “The thought we used there is, would we give a junior staff person the same level of permissions? If we would, we can think about it with AI. If we would not, [then] we should not give AI that many permissions in our system to view information and work with information”.

To build Nuha, an AI model that can classify hate speech against women in different Arabic dialects, JOSA collected data from social media. “We went through a lot of processes on how we should handle this data, what kind of data we should include when we are monitoring social media and how we can ensure privacy through social media accounts we collected data from, and it was very important that no personal information was to be involved”, said a JOSA employee during the FGD. She further explained: “we do not store any personal information, and in the data collection process, the data collectors remove the data manually of any personal information”.

Informed and prior consent of data holders, ensuring they understand how their information is used and that they can refuse processing are essential. One organisation stated: “We will avoid deploying AI where consent, safety and data sovereignty cannot be guaranteed”. Another respondent based in South Africa, stated that “when working with communities — especially in language documentation, translation, or storytelling — we make sure participants understand how their voices, stories, or recordings will be used. We never collect or share data without permission”.

Measures to protect personally identifiable information include secure data management and storage practices. “We prioritise the principles of data minimisation, informed consent, and confidentiality. Personal or sensitive information is handled with strict access controls and encryption. Where possible, we use anonymised or aggregated data to protect individuals and communities we serve”, said Tanzania-based.

Ethical reviews and data audits

In their deployment of AI tools, some respondents mentioned that their organisations assess risks to identify

potential harms associated with bias, privacy, and misinformation. Assessments were emphasised not only by those that developed their own models but also organisations that use existing models. One respondent in Bangladesh wrote: “We consider and address potential ethical and human rights risks by implementing rigorous safety evaluations, fairness assessments and ongoing monitoring to ensure responsible AI deployment”.

A number of organisations mentioned that they conducted these assessments prior to AI adoption or they make these assessments on a regular basis. “Before implementing any AI tool or system, we conduct an internal review to identify possible risks to privacy, security and fairness. This includes assessing how data is collected, stored and processed to ensure compliance with human rights and data protection standards”, stated a respondent based in Tanzania.

Auditing datasets for biases was specifically highlighted for those that created or are interested in creating their own tools during the design and development stage.

“We carefully review datasets and AI outputs to identify and minimize biases that could disproportionately affect marginalized groups, especially women and rural communities”, said a survey respondent whose organisation developed their own model. A respondent who requested anonymity stated the following: “If and when we do deploy AI tools, our ethical and rights-based approach would include... reviewing datasets for potential bias and ensuring that AI models do not reinforce harmful stereotypes as well as exclude underserved populations. Where necessary, we will work with technical partners to audit data and algorithms for fairness”.

During the data annotation process to develop Nuha, women annotators working in the human rights field were given guidelines to ensure that the model would be free from

biases and would not infringe on freedom of expression. “The thing that we had challenges with is that there is a very thin line between hate speech and freedom of speech; how do the annotators know which is which, and how to differentiate them both because we do not want model that classifies everything as hate speech”, explained a JOSA representative in the FGD discussion.

In some cases, respondents acknowledged that the capacity of their organisations is limited when it comes to conducting these types of assessments. “We also discuss ethical implications within our team when exploring AI-assisted work such as possible bias, misinformation and data security risks. However, we recognize that our current capacity to conduct deeper ethical assessments is limited. We are interested in building stronger digital literacy and AI ethics skills to ensure that as we integrate technology into our advocacy, we do so in a way that protects human rights and maintains community trust”, said a survey respondent.

Transparency and accountability

Respondents noted the following as important factors, namely transparency about AI tools through the documentation of their development and purpose; informing users of potential risks; and making it clear to users what the AI tool does. “If and when we do deploy AI tools, our ethical and rights-based approach would include ...making AI tools understandable to users and stakeholders, with clear explanations of how decisions are made”, stated one organisation.

AI outputs and decisions also need human oversight. Some of the mechanisms put in place to ensure this oversight include Human-in-the-loop (HITL) validation, feedback mechanisms, and human review and verification of any AI-generated content “for accuracy, fairness and sensitivity before publication”, as one respondent stated.

CDR in Türkiye, which developed Policyful, said it ensures transparency of its AI-assisted outputs and human oversight. The organisation explained in a survey response: “Our organisation develops AI tools to strengthen civic engagement and transparency in policymaking. We prioritise the ethical use of AI by ensuring data integrity, minimising hallucinations through human-in-the-loop validation and maintaining accountability in automated analyses....We integrate ethical review and human rights considerations into every stage of our AI projects. This includes careful data selection to avoid bias, transparency about AI-assisted outputs and continuous human oversight in model training and deployment. We also provide public explainers and open datasets to ensure that our tools contribute to democratic accountability rather than replace human judgment”.



“Our organisation develops AI tools to strengthen civic engagement and transparency in policymaking, prioritising ethical use through human-in-the-loop validation”

— Centre for Democracy Research (CDR), Türkiye

One Tanzania-based organisation, stated that: “We are committed to transparency in how AI systems operate and how decisions are made. We provide clear communication to stakeholders about the purpose and limitations of AI tools, and we maintain accountability through regular audits and feedback mechanisms”. Another organisation based in South Africa, wrote: “We openly communicate when AI tools are involved in our processes, ensuring that human reviewers, translators and leaders maintain oversight and final responsibility for all content produced”.

When AI is deployed in research, one respondent mentioned that they ensure outputs are verified with scientific sources and adequately cited. Another based in Kenya, said that “we do not rely solely on AI outputs for public advocacy without verification”.

Emphasis was put on using AI to assist humans and not replace them, particularly in important decisions and when working with excluded groups. According to the same Tanzania-based organisation: “AI tools are used to support, not replace, human judgment. Critical decisions, especially those affecting rights or reputations, remain under human review to ensure ethical alignment with our values and mission”.

Spring NGO said: “If civil society groups adopt AI tools without understanding how algorithms function or how decisions are generated, it can lead to a loss of human oversight. Spring NGO highlighted that AI should support, not replace, human judgment — especially in matters affecting vulnerable populations”.

“

“AI should support, not replace, human judgment — especially when working with marginalised groups”

— Spring NGO, Pakistan

Stakeholder engagement

Some respondents underlined the need to consult with and involve key stakeholders, and particularly experts on AI ethics and risks, as well as affected communities and local authorities in relation to civil society’s deployment of AI.

CSMA said: “If CSMA is to deploy AI models,

ethical and human rights considerations and risks will be of utmost importance. Adhering to ethical issues will ensure that we respect local cultures, involve local communities in our deployment of AI models and also ensure that our people, the primary beneficiaries of our programmes, harvest the fruits of our AI inspired work.”

JOSA conducted consultations with feminist organizations and human rights defenders in Jordan to build Nuha. “Feminist organizations in Jordan had a really big role in the model to be inclusive because it was important for us to build something that is localized that is similar to what we think violence looks like in our own countries, and so it was very important for these feminist organizations to decide on the terminologies on the types of violence”, said an FGD participant from JOSA.

Prioritising accessibility and contextual adaptation

When AI tools are deployed, it is essential that they are accessible to people and communities they are supposed to serve and that such deployment do not further result in inequality and marginalisation, for instance, due to disability, language obstacles, or lack of internet access.

“If and when we do deploy AI tools, our ethical and rights-based approach would include...ensuring that the benefits of AI are distributed equitably and do not widen the digital divide. This includes building capacity and digital literacy among our beneficiaries and partners”, a respondent who requested anonymity said.

Spring NGO said: “accessibility and inclusion remain at the core of Spring NGO’s ethical framework. AI tools must be designed to be accessible to persons with disabilities in terms of both interface and content. Without inclusive design principles, technological innovation risks deepening the digital divide

rather than closing it. In summary, Spring NGO advocates for an ethical, inclusive, and rights-based approach to AI, where technology amplifies human dignity, promotes equality, and ensures that no one — especially persons with disabilities — is left behind”.

Further, some respondents said in their deployment they seek to adapt AI to their local contexts to prevent risks of bias and exclusion.

CDR mentioned that it has a “multidisciplinary team of AI developers, researchers and legal engineers [that] ensures that model outputs remain contextually accurate and rights-respecting”.

“We recognize that most AI systems are trained on data that reflect Western or dominant cultural worldviews. Therefore, we work carefully to contextualize AI-generated outputs, especially when translating Scripture or creating educational content, ensuring cultural and linguistic accuracy”, said a South African organisation.

Awareness raising among staff

To help implement their AI policies and measures, organisations also take steps at raising awareness among their staff, partners and beneficiaries on how to use AI ethically and responsibly and about the human rights risks and implications of such technologies.

Wardil Organization in Iraq stated that it: “considers the ethical and human rights risks related to the use of AI. Although we have not developed our own AI model, we use AI tools in limited ways—mainly for communication, translation, and data organisation. We address potential risks by promoting transparency, privacy protection and fairness, and by raising awareness among our staff about responsible and ethical AI use”. While another respondent wrote: “We regularly train our staff and partners on responsible AI use,

data ethics and human rights implications, ensuring ongoing awareness and adaptation to emerging standards and best practices”.

Limiting scope of AI use or size of training datasets

Further, to minimize human rights risks, some organisations deliberately choose to deploy AI in limited ways, for instance, in administrative tasks and data processing.

“We take potential ethical risks into account, which is why we only implement artificial intelligence for administrative processes and data processing to reduce manual work”, said a respondent from Ukraine. Another, working in Ethiopia stated: “We take precautions not to impact the rights of others. This may be through not completely relying on AI as it may create misinformation. Particularly we apply caution while using AI for research and content creation”.

An FGD participant from Uganda said his organisation is “not actually jumping onto the [AI] train is not driving core programs at his organisation due to the sensitivity of the topic it is working on—gender-based violence— and the lack of context sensitive tools. “The biggest concern is that AI does not really reflect our realities. Tools do not understand our languages, our culture, our context, which can really lead to harm, especially for organizations like us that are dealing with issues that are related to gender-based violence”, he explained.

Others choose to develop models based solely on small datasets or data that they control to minimize risks associated with large language models, which are built on massive amounts of data scraped from the internet. This is the case for several AI tools developed by CSOs and highlighted above in this report.



Chapter 3

ADVOCATING FOR AI ACCOUNTABILITY AND GOVERNANCE:

Challenges faced by civil society

Around the world, civil society organisations, coalitions, and activists have been advocating for stronger AI governance, in line with human rights standards, and for mechanisms that hold developers and deployers of harmful and invasive AI technologies accountable.¹²² For example, a coalition of 150 civil society groups in Europe worked together to safeguard human rights through the development and application of the EU AI Act.¹²³ Following the adoption of this Act, civil society had to further monitor its implementation, as the European Commission and EU Parliament took steps to weaken its safeguards.^{124 125} In Brazil, civil society advocacy has been successful in addressing a weak and self-regulatory AI governance approach lobbied for by the private sector.¹²⁶ In the Occupied Palestinian Territory, collaborative media investigations revealed the extent of Israel's lethal deployment of AI and biometric surveillance systems in connection to the genocide in Gaza¹²⁷ and the complicity of specific tech companies in these abuses and war crimes.¹²⁸ In India, the Internet Freedom Foundation has been mapping FRT projects in the country¹²⁹ and pushing for transparency in the deployment of AI in public spaces.¹³⁰

Most survey respondents (nearly half) stated that their organisations do not conduct any advocacy work to influence how AI systems are developed, deployed or regulated. About a third of respondents stated that they would like to influence how AI systems are developed, deployed or regulated. Only 17 percent of respondents stated that their organisations seek to influence the ways AI systems are developed, deployed, and regulated.

There are obstacles that stand in the way of broader civil society participation, particularly in the Global South, and contribute to the debate around AI governance and accountability. The main obstacles are a lack of resources, a hindering political

and regulatory environment, and a lack of transparency surrounding the development and deployment of AI systems.

First, in terms of resources, there is the lack of funding and staff with needed knowledge and skills in their organisations. These two are interrelated as the lack of funds affects organisations' ability to hire employees and consultants who can support them with AI accountability and governance activities.

A Ugandan organisation, said that it "lacks the necessary expertise in AI governance, policy and technical development, making it challenging to effectively advocate for responsible AI practices". It further explained in the survey response: "Limited financial resources hinder the organization's ability to participate in key forums, workshops and conferences on AI governance and regulation, restricting their ability to engage with policymakers, experts and other stakeholders. The organization requires training and capacity-building programs to enhance its understanding of AI technologies, their applications, and implications for human rights and social justice".

"...Internally, we face a lack of technical knowledge and skills needed to fully understand how AI systems function, how they affect human rights and how to design strong advocacy strategies around them. This gap makes it difficult to analyze or respond to the growing use of AI in surveillance, information control and public communication", said a respondent from Tanzania.

In other cases, some respondents stated that there is a general shortage of people with the needed knowledge and technical skills in the contexts in which they are working. "The main obstacle is severe talent scarcity in Nigeria. We lack local experts to technically audit proprietary AI systems for bias and non-

compliance. This knowledge gap means we cannot credibly challenge powerful deployers, undermining our advocacy for fairness and accountability in critical areas like facial recognition”, one respondent who chose to remain anonymous said.

In some cases, available funding tends to exclude those conducting critical AI work, particularly when that funding is coming from the tech industry. Disha Verma, Program Manager at the Global Tech Institute asserted: “I will say that there is a glaring lack of funds right now to fund critical work on AI because there is way too much money to do favourable work on AI. If you want to work closely with an AI provider, and build an AI safety institute, or have red lines and guardrails for AI, work with the government on a regulation, the positive reinforcements are many. You will get a lot of money from it, from all kinds of funders coming in from outside India, or within India. But the minute you start to say that maybe we do not need AI right now, maybe we do not need AI in this sector, or maybe we do not need AI in policing, we do not need AI in welfare services, people cut you out. There is no money for that research. So how do we continue making that knowledge? How do we continue creating knowledge and research and outputs and advocacy material when we are not getting paid as researchers and activists? So that is really concerning to me. Also to note that AI has become a tool of soft power for India. We take a lot of nationalistic pride in our AI systems and our AI mission. So, the minute you start to question it, it becomes like an anti-national sentiment and a statement that can be jarring for the government, so they try to shut you off and basically attack you”.

Another major hurdle for civil society participation in shaping how AI is developed, deployed, and regulated is an unfavourable regulatory and political environment. Respondents and interviewees noted, in

particular, the lack of robust regulations on AI that align with international human rights standards in addition to a climate that hinders civic participation, further complicating advocacy efforts. Despite rapid and disruptive advancements in the AI field and increased adoption across different sectors and industries, robust human-rights based regulation of AI is still moving at a slow pace. In 2024, the EU adopted the AI Act, a comprehensive framework for regulating AI that takes a risk-based approach.¹³¹ The Act bans applications with unacceptable risk and imposes strict requirements for high-risk applications such as FRTs and emotion recognition systems. It only imposes transparency requirements for lower-risk systems. The Act, however, has major loopholes, including broad exceptions for use of high-risk AI systems for the purposes of defence, law enforcement, and migration.¹³² Furthermore, under pressure from the tech industry and the administration of Donald Trump in the U.S., the EU Commission proposed to delay the implementation of some provisions in the AI Act and weaken some of its safeguards.

In other regions, Brazil has been in the process of drafting a similar law that also adopts a risk-based approach but it has yet to be adopted. Other countries, like Japan¹³³ and South Korea,¹³⁴ have adopted a ‘soft approach’ in their AI legislation, prioritising innovation and imposing limited obligations on deployers and developers compared to the EU AI Act. Most recently, India adopted issue-specific rules on deepfakes which mandated the labelling of deepfakes and banned “deceptive impersonations, non-consensual intimate imagery and synthetic content tied to serious crimes”.¹³⁵ It also mandated platforms to take down infringing content within three hours.¹³⁶ This short deadline was criticised by digital rights groups which argue this will only result in more automated removal decisions and could have a chilling effect on freedom of

expression.¹³⁷

Many countries are currently opting to adopt non-binding principles or codes such as India,¹³⁸ Saudi Arabia,¹³⁹ and Singapore.¹⁴⁰ There are also guidelines adopted at the international or regional level such as the OECD's Recommendation of the Council on Artificial Intelligence (the "Recommendation")¹⁴¹ and the Continental African Intelligence Strategy.¹⁴²

There are "weak and sometimes non-existent regulatory frameworks in Kenya and much of Africa; there is a lack of comprehensive AI governance laws as well as enforcement mechanisms that protect communities from misuse of AI. Existing data protection laws (like Kenya's Data Protection Act) may not be fully enforced as well as adapted to the complexities of AI. There are often regulatory capture and lack of independence in oversight bodies", said one respondent, adding that as a result, "CSOs have limited legal and institutional pathways to challenge unethical and harmful AI deployments".

Lack of political willingness and an unstable political environment were also cited as factors that hinder adoption.^{143 144} Civil society in many countries is working in restrictive and repressive operating environments. As a result, they cannot effectively question powerful actors' deployment of AI to hold them to account. Freedom of expression and access to information, freedom of peaceful assembly and the right to privacy are often tightly controlled. According to the CIVICUS Monitor's 2025 data, "Only 7.2 percent of the global population now lives in countries where civic space is open or narrowed, 7.5 percentage points less than in 2024, indicating a further deterioration of civic space conditions globally".¹⁴⁵

There is a "fear of retaliation or repression raising concerns about government-

deployed AI systems, especially those related to surveillance, censorship, and policing, which can be politically sensitive. Civil society organisations may fear retaliation, de-registration or intimidation when calling out unethical practices. This inhibits open advocacy and transparency, particularly in restrictive civic spaces", one anonymous respondent stated. Another respondent working in Afghanistan wrote: "Our organisation does not currently engage in direct AI advocacy due to operating restrictions and security concerns in Afghanistan. The lack of resources, expertise, and safe civic space limits our ability to influence AI policies or participate in related forums. However, we are aware of the human rights implications of AI and hope to engage more once a safer environment becomes available".

Additionally, algorithmic opacity makes it difficult for civil society to understand how AI systems are deployed, hindering their ability to hold deployers and developers to account. In fact, 45 percent of respondents listed the lack of transparency from deployers on when and how they deploy these systems and 38 percent listed the lack of algorithmic transparency from developers on how their systems work as obstacles to exercising oversight and calling for accountability.

According to one respondent, "Externally, there is a lack of transparency from government and private sector actors regarding when, where, and how AI systems are deployed. This secrecy prevents civil society from monitoring potential abuses, assessing compliance with human rights standards, or holding deployers accountable. Together, these challenges limit our ability to meaningfully influence AI governance and advocate for responsible and ethical use of AI technologies". Another respondent stated that "many AI systems, especially those developed by private companies or

deployed by governments, operate as ‘black boxes’ with little to no public visibility into: 1. How they are built 2. What data they use 3. What assumptions or biases they encode. It becomes nearly impossible to assess and even challenge harmful impacts, especially when those affected lack the technical knowledge as well as legal means to demand redress”.

In other cases, governments’ own communication about when AI is deployed and when it is not can be confusing to advocates, as Verma noted of the Indian government’s approach in an interview. “The government is not honest about what is AI and what is not AI. They call everything AI. We do not know if it is just machine automation, if it is just simple maths. But when you ask, when you seek further questions, it turns out to be just a thermal imaging system, for instance, not AI. It is called AI for the hype train to add to the nationalistic fervour and so on and so forth”.

According to Juan Manuel García, an Argentinian consultant at the intersection of technology and human rights, some organisations in Latin America had previously submitted Freedom of Information Requests (FOIAs) to government agencies to understand how they deploy AI, but those requests have not always resulted in positive responses. “One of the main problems that have been very common for many organisations in Latin America is access to information regarding the use of AI. The Argentinian government in mid-2024 created this new agency that is called the Unit for AI for Security that has a really broad spectrum of activities and some of the activities are based on the use of AI surveillance of social media. Different organisations have presented FOIA requests, but they did not answer them, so we do not know how they are using it...Sadly, we do not know the specificities, because they do not answer the requests. It is

a problem”.

Lacking resources, access to information, and facing retaliation for their work, civil society is disadvantaged compared to powerful and well-resourced government and private sector actors in shaping how AI is deployed and governed. Civil society also faces exclusion from national and international AI discussions.

“Civil society organisations like ours often face closed-door policymaking and technical barriers in understanding AI systems. Without funding and open access to information, it is hard to hold large tech actors accountable or influence regulation effectively”, said a Kenya-based organisation.

“There are few formal mechanisms for civil society participation in policymaking, and many discussions on AI governance take place behind closed doors, limiting NGO influence and representation”, said disability rights organisation Spring NGO.

This often results in a power imbalance, where the private sector with a few large players and governments have more power to influence these discussions and spaces and their outcomes. The lack of representation of civil society means human rights and impacts on excluded groups are neglected.

“We are quite involved in AI governance; we talk to our industries whenever we can, in terms of how you govern AI, how do you set up guardrails, etc. AI safety is a big thing in India. But the problem is that it is really noisy, the space. There are so many conversations happening. There is a bustling and booming AI industry in India, which wants in on every room, which wants to say that innovation versus regulation, we need innovation, we do not need regulation. And then there are the government quarters and the tech conservatives that say that, no, regulate

everything, put everything to the ground... The problem is that our government is authoritarian, so giving them powers through legislation and giving them powers through regulation is dangerous because they will surveil, they will intercept, they will do all of that and curb free speech and right to privacy whenever they can. But on the other hand, letting the private sector run amok and do whatever they like is also dangerous, because these then become very powerful, and as consumers, it tips the scale of balance of power quite a bit. And we are not equals, we are not able to hold our platforms accountable. So either way, it is a lose-lose situation for the citizens", said Verma of the Tech Global Institute.

While civil society in Brazil has been successful in pushing back against an initial AI draft bill that "was really favored by the industry"¹⁴⁶ and had "a very light touch and minimalistic approach"¹⁴⁷ to regulation, their resources still do not match those of the tech industry. As a more comprehensive draft bill currently in consideration, Bill No. 2338/2023, has yet to be finalised and enacted, the private sector has increased pressure on lawmakers.¹⁴⁸

Civil society from the Global South faces additional hurdles participating in international discussions and forums on AI governance such as the global AI summits¹⁴⁹ and the UN Global Dialogue on AI Governance¹⁵⁰ due to expensive travel costs and strict visa rules.

"Civil society voices from the Global South are rarely present in major AI policy spaces. Even when invited, discussions are often dominated by private sector actors. Without resources to participate, small organizations like ours are left to adapt to systems we had no say in shaping", said an organisation working in Nigeria. Another pointed out the power imbalance that advantages

companies and governments: "Large tech companies and government institutions often have asymmetric access to resources, expertise and policy influence. Civil society organisations, especially those working at grassroots level, may lack the technical capacity to audit systems and the political capital to influence AI policy. This creates a knowledge and power gap that prevents meaningful accountability and engagement".

The dominance of the English language in these spaces is another hurdle for voices from the Global South. "Civil society voices from the Arab region are rarely included in global AI policy debates. We aim to represent these voices but face linguistic and logistical limitations", said an organisation based in Jordan. Furthermore, there is the lack of digital remote participation for those facing visa restrictions or cannot afford travel. In some cases, where remote participation is available, lack of connectivity or expensive data bundles hinder online participation. "Civil society in South Sudan is often excluded from AI policy discussions due to limited connectivity and resources. We need access to regional and international policy dialogues to ensure local realities are considered in global AI ethics debates", said one respondent.

According to one respondent, "most AI discussions are led by actors in the Global North. Local CSOs like EAO rarely have access to consultations where African civic perspectives can influence AI ethics or governance frameworks".

CONCLUSION AND RECOMMENDATIONS

Despite facing threats of surveillance, censorship, disinformation, and a civic space under attack around the world, CSOs have been fighting against harmful trends and patterns in AI adoption by governments and the private sector. On the global stage, they have been successful in advocating for a ban on fully autonomous weapons at the UN Convention on Certain Conventional Weapons¹⁵¹ and campaigns led by tech workers have educated the public on the complicity of Big Tech in Israel's AI-assisted genocide in Gaza and occupation of Palestinian Territory.¹⁵²

In Brazil, civil society efforts have been successful in advocating for a more robust and human rights-based alternative law to a weaker and self-regulatory bill that was favoured by the private sector. Across the world, civil society organisations have been building their own small language models and their own datasets to advance AI uses that align with their values and respect of human rights, including the right to privacy, free expression, and freedom from discrimination.

Despite the limited resources, civic space actors demonstrate that an alternative approach exists in contrast to the extractive nature of the AI models proposed and implemented by Big Tech and then used by governments. However, for civic space actors to leverage AI's inclusive, non-extractive, and beneficial aspects for all, and to prevent its harmful applications, the sector needs resources and capacity building.

- With financial resources, CSOs can hire technical expertise to help integrate AI in their work, build their own models and datasets, and invest in the

needed infrastructure, both software and hardware.

- Networks and initiatives that support CSOs can also strengthen the capacity of CSOs, including starting with basic knowledge and understanding such as demystifying AI, the benefits and drawbacks – to deploy it ethically and in alignment with human rights.
- There is a need for multilingual toolkits, resources, and templates for frameworks and policies that can help guide CSOs in their deployment of AI. These resources can cover, for example, policy templates on AI use so that CSOs are transparent with different stakeholders and detailed technical guides on conducting human rights impact assessments when deploying and developing AI. Such resources should be available in as many languages as possible to reach as many organisations as possible around the world.
- In addition to guides and templates, many CSOs lacking technical expertise and resources will need mentorship and guidance from more technical coalitions and organisations to integrate AI in their work responsibly and assess and address human rights risks.
- It is essential to invest in tools, resources, capacity building and legal protection to allow CSOs to resist and challenge the dominance of Global North commercial models that do not reflect the realities in which they work. For instance, investing in independent initiatives that support AI sandboxes, the building of open, diverse and ethical datasets, and small language models could help CSOs access and deploy AI in ways that do align with their values and do not cause harm.
- In the face of AI-powered surveillance, censorship, and disinformation eroding civic space, it is worth exploring building and strengthening broad and diverse coalitions to exercise public pressure on governments and tech companies.
- It is also essential to continue raising awareness and building digital security skills of civil society actors to reduce risk of unauthorised access to data and mitigate such their impacts, particularly with emerging threats associated with AI.
- Given the far-reaching social, economic, and political implications of AI, the coalitions of CSOs pushing for robust AI tech global governance need to include more voices from the Global South and from other sectors such as unions, artists, and environmental and anti-racist movements.



REFERENCES

- 1 Notre Dame Learning. "AI Overview and Definitions". Undated. <https://learning.nd.edu/resource-library/ai-overview-and-definitions/> (accessed 30 April 2026).
- 2 Choi, A. & M, K.X. "What are AI hallucinations? Why AIs sometimes make things up". The Conversation. 21 March 2025. <https://theconversation.com/what-are-ai-hallucinations-why-ais-sometimes-make-things-up-242896> (accessed 30 April 2026).
- 3 Coursera. "What Is Data Annotation in Machine Learning"? 19 December 2025 <https://www.coursera.org/articles/data-annotation-in-machine-learning> (30 April 2026).
- 4 Abrougui, A. (2025). Phishing, spyware, and smart city tech: Surveillance in Sisi's Egypt. In T. Roberts & A. Mare (Ed.). Digital Surveillance in Africa: Power, Agency, and Rights (pp. 57–84). London,: Zed Books. Retrieved April 30, 2026, from <http://dx.doi.org/10.5040/9781350422117.ch-3>
- 5 Databricks. "What is Fine Tuning"? Undated. <https://www.databricks.com/blog/what-is-fine-tuning> (30 April 2026).
- 6 Notre Dame Learning. "AI Overview and Definitions". Undated. <https://learning.nd.edu/resource-library/ai-overview-and-definitions/> (accessed 30 April 21
- 7 IBM. "What is human-in-the-loop"? Undated. <https://www.ibm.com/think/topics/human-in-the-loop> (30 April 2026).
- 8 Hugging Face. "Natural Language Processing and Large Language Models". Undated. <https://huggingface.co/learn/llm-course/en/chapter1/2> (accessed 30 April 2026).
- 9 AI Myths."The term AI has a clear meaning". Undated. <https://www.aimyths.org/the-term-ai-has-a-clear-meaning> (accessed 30 April 2026)
- 10 United Nations. "Countering Disinformation". Undated. <https://www.un.org/en/countering-disinformation> (accessed 30 April 2026).
- 11 ISO. "Unravelling the secrets of natural language processing". <https://www.iso.org/artificial-intelligence/natural-language-processing> (accessed 30 April 2026).
- 12 Abrougui, A. "Global Trends in Digital Security: Civil Society & Media". October 2023. Internews. <https://internews.org/wp-content/uploads/2023/11/Global-Trends-in-Digital-Security.pdf> (accessed 30 April
- 13 Abrougui, A. "Global Trends in Digital Security: Civil Society & Media". October 2023. Internews. <https://internews.org/wp-content/uploads/2023/11/Global-Trends-in-Digital-Security.pdf> (accessed 30 April 2026).
- 14 IBM. "What is training data". Undated. <https://www.ibm.com/think/topics/training-data> (accessed 30 April 2026).
- 15 Adib Bin Rashid, MD Ashfakul Karim Kausik, AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications, Hybrid Advances, Volume 7, 2024, 100277, ISSN 2773–207X, <https://doi.org/10.1016/j.hybadv.2024.100277>. (<https://www.sciencedirect.com/science/article/pii/S2773207X24001386>).
- 16 Ibid.
- 17 Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation:

- Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951719897945>
- 18 ISO. "Unravelling the secrets of natural language processing". <https://www.iso.org/artificial-intelligence/natural-language-processing> (accessed 30 April 2026).
- 19 Adib Bin Rashid, MD Ashfakul Karim Kausik, AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications, *Hybrid Advances*, Volume 7, 2024, 100277, ISSN 2773-207X, <https://doi.org/10.1016/j.hybadv.2024.100277>. (<https://www.sciencedirect.com/science/article/pii/S2773207X24001386>).
- 20 CIVICUS. "'The lesson from Gaza is clear: when AI-powered machines control who lives, human rights die'". 16 July 2025. <https://lens.civicus.org/interview/the-lesson-from-gaza-is-clear-when-ai-powered-machines-control-who-lives-human-rights-die/> (accessed 3 May 2026).
- 21 Hao, K. (2025). *Empire of AI: Inside the reckless race for total domination*. Penguin.
- 22 Ibid.
- 23 Amnesty International. "Global: New technology and AI used at borders increases inequalities and undermines human rights of migrants". 21 May 2024. <https://www.amnesty.org/en/latest/news/2024/05/global-new-technology-and-ai-used-at-borders-increases-inequalities-and-undermines-human-rights-of-migrants/> (accessed 3 May 2026).
- 24 Lungu, M. "Predictive policing AI is on the rise – making it accountable to the public could curb its harmful effects". *The Conversation*. 6 May 2025. <https://theconversation.com/predictive-policing-ai-is-on-the-rise-making-it-accountable-to-the-public-could-curb-its-harmful-effects-254185> (accessed 30 April 2026).
- 25 European Artificial Intelligence & Society Fund. "AI & Social Welfare Funders' Briefing". September 2025. https://europeanaifund.org/wp-content/uploads/2025/09/EAISF-AI-Social-welfare_Funders-Briefing.pdf (accessed 3 May 2026).
- 26 CIVICUS. "Facial recognition: the latest weapon against civil society". 23 May 2025. <https://lens.civicus.org/facial-recognition-the-latest-weapon-against-civil-society/> (accessed 3 May 2026).
- 27 CIVICUS. "CHINA: 'The state is using generative AI to engineer reality through informational gaslighting'". 12 March 2026. <https://lens.civicus.org/interview/china-the-state-is-using-generative-ai-to-engineer-reality-through-informational-gaslighting/> (accessed 3 May 2026).
- 28 Asia Indigenous Peoples Pact Foundation. "Civil Society Manifesto for Ethical AI". 2023. <https://aippnet.org/wp-content/uploads/2023/12/civil-society-manifesto-for-ethical-ai.pdf> (accessed 3 May 2026).
- 29 Digital Justice Network. "Generative AI guide for civil society". Association for Progressive Communications. 6 January 2026. <https://www.apc.org/en/pubs/generative-ai-guide-civil-society> (accessed 3 May 2026).
- 30 CIVICUS. "'Civil society is pushing for transparent, rights-centred AI governance'". 18 March 2025. <https://lens.civicus.org/interview/civil-society-is-pushing-for-transparent-rights-centred-ai-governance/> (accessed 1 May 2026)
- 31 Human Rights Funders Network. "Funding at a Crossroads: Foreign Aid Cuts And Implications For Global

- Human Rights." September 2025. <https://www.hrfn.org/wp-content/uploads/2025/09/Funding-at-a-Crossroads-HRFN-Sept-2025.pdf>
- 32 Hickok, E., Mohan, S., Pielemeier, J. and Kakkar, J.M. "Multistakeholder Promises and Power Gaps in Global AI Summits". 17 March 2026. Tech Policy Press. <https://www.techpolicy.press/multistakeholder-promises-and-power-gaps-in-global-ai-summits/> (accessed 3 May 2026).
- 33 Prysiachniuk, V. & Ford, C. "How AI is accelerating the global surveillance state, and how you can defend yourself". NymVPN, 18 September 2025. <https://nym.com/blog/clearview-ai-and-palantir> (accessed 30 April 2026).
- 34 Mahoozi, S. "Mahsa Amini death: facial recognition to hunt hijab rebels in Iran". Thomson Reuters Foundation. 21 September 2022. <https://news.trust.org/item/20220921162019-v8vk/> (accessed 30 April 2026).
- 35 The Economist. "China: facial recognition and state control". Youtube. 24 October 2018. <https://www.youtube.com/watch?v=IH2gMNRUuEY> (accessed 30 April 2026).
- 36 Zomorodi, M., Monteleone, K. & Meshkinpour, S. "How facial recognition allowed the Chinese government to target minority groups". NPR. 9 December 2022. <https://www.npr.org/2022/12/09/1141627539/how-facial-recognition-allowed-the-chinese-government-to-target-minority-groups> (30 April 2026).
- 37 CIVICUS. "Facial recognition: the latest weapon against civil society". 23 May 2025. <https://lens.civicus.org/facial-recognition-the-latest-weapon-against-civil-society/> (accessed 30 April 2026).
- 38 Ibid.
- 39 Amnesty International. "Netherlands: Mass police surveillance of protests part of 'growing control culture' – new report". 15 October 2024. <https://www.amnesty.org.uk/latest/netherlands-mass-police-surveillance-protests-part-growing-control-culture-new/> (30 April 2026).
- 40 Sinha, A. "The Landscape of Facial Recognition Technologies in India". Tech Policy Press. 13 March 2024. <https://www.techpolicy.press/the-landscape-of-facial-recognition-technologies-in-india/> (accessed 30 April 2026).
- 41 SFLC. "Deployment of Facial Recognition Technology for State Surveillance and Monitoring". 16 January 2024. <https://sflc.in/deployment-of-facial-recognition-technology-for-state-surveillance-and-monitoring/> (accessed 30 April 2026).
- 42 Herta Security. "Safe City project with facial recognition in Phuket". 14 December 2017. <https://hertasecurity.com/news/safe-city-project-with-facial-recognition/> (accessed 30 April 2026).
- 43 Borak, M. "Thailand rolling out AI surveillance system after high-profile kidnapping of Chinese actor". Biometric Update. 28 January 2025. <https://www.biometricupdate.com/202501/thailand-rolling-out-ai-surveillance-system-after-high-profile-kidnapping-of-chinese-actor> (accessed 30 April 2026).
- 44 Roberts, T. et al. (2023) Mapping the Supply of Surveillance Technologies to Africa: Case Studies from Nigeria, Ghana, Morocco, Malawi, and Zambia, Brighton: Institute of Development Studies, DOI: 10.19088/IDS.2023.027
- 45 Abrougui, A. (2025). Phishing, spyware, and smart city tech: Surveillance in Sisi's Egypt. In T. Roberts & A. Mare (Ed.). Digital Surveillance in Africa: Power, Agency, and Rights (pp. 57–84). London: Zed Books. Retrieved

- April 30, 2026, from <http://dx.doi.org/10.5040/9781350422117.ch-3>
- 46 Ibid.
- 47 AI Sur Consortium. "Facial Recognition in Latin America". <https://estudio.reconocimientofacial.info/en/> (30 April 2026).
- 48 Bergengruen, V. "A Controversial Facial-Recognition Company Quietly Expands Into Latin America". Time. 18 June 2024. <https://time.com/6988684/clearview-ai-latin-america/> (accessed 30 April 2026).
- 49 Lungu, M. "Predictive policing AI is on the rise – making it accountable to the public could curb its harmful effects". The Conversation. 6 May 2025. <https://theconversation.com/predictive-policing-ai-is-on-the-rise-making-it-accountable-to-the-public-could-curb-its-harmful-effects-254185> (accessed 30 April 2026).
- 50 CIVICUS Monitor. "Israel downgraded in global ratings report on civic freedoms". 9 December 2025. https://monitor.civicus.org/press_release/2025/israel/ (accessed 30 April 2026).
- 51 Ibid.
- 52 Aljazeera. "Dareen Tatour sentenced to five months in prison over poem". 31 July 2018. <https://www.aljazeera.com/news/2018/7/31/dareen-tatour-sentenced-to-five-months-in-prison-over-poem> (accessed 30 April 2026).
- 53 Rani M (2024). "Impacts and ethics of using Artificial Intelligence (AI) by the Indian Police". Public Administration and Policy, Vol. 27 No. 2 pp. 182–192, doi: <https://doi.org/10.1108/PAP-06-2023-0081>
- 54 Ibid.
- 55 Privacy International. "How predictive policing can be used at protests". 5 May 2021. <https://privacyinternational.org/explainer/4501/how-predictive-policing-can-be-used-protests> (accessed 30 April 2026).
- 56 Majid, K. "Predictive Algorithms: Help or Hindrance"? Center of AI and Digital Ethics, University of Melbourne. 30 March 2023. <https://www.unimelb.edu.au/caide/news-media-and-events-archive/caide-community-blog/predictive-algorithms-help-or-hindrance> (accessed 30 April 2026).
- 57 O'Neill, Cathy. (2017). Weapons of Maths Destruction: How Big Data Increases Inequality and Threatens Democracy. Penguin Press.
- 58 Chandler, C. "Inside the Black Box of Predictive Travel Surveillance". Wired. 13 January 2025. https://www.wired.com/story/inside-the-black-box-of-predictive-travel-surveillance/?utm_source=chatgpt.com (accessed 30 April 2026).
- 59 IDEMIA Group. "BORDERGUARD™ Targeting". Undated. <https://www.idemia.com/borderguard-targeting> (accessed 1 May 2026).
- 60 Chandler, C. "Inside the Black Box of Predictive Travel Surveillance". Wired. 13 January 2025. https://www.wired.com/story/inside-the-black-box-of-predictive-travel-surveillance/?utm_source=chatgpt.com (30 April 2026).
- 61 Travizory. "Travizory API-PNR Targeting System". Undated. <https://www.travizory.com/what-we-do/advance-passenger-information> (accessed 1 May 2026).
- 62 CapitalFM. "Kenya's Controversial ETA System Switch Triggers Investor, Tourism Backlash". AllAfrica. 22 April 2025. <https://allafrica.com/stories/202504220208.html> (accessed 1 May 2026).
- 63 Lloyd, G. "African Border Security Week 2025". Travizory. <https://www.travizory.com/african-border-security-week-2025> (accessed 1 May 2026).

- 64 Access Now. "Egypt: rights groups condemn latest blocking of news websites. Press release". 22 June 2023. <https://www.accessnow.org/press-release/egypt-rights-groups-condemn-latest-blocking-of-news-websites> (accessed 1 May 2026).
- 65 Abrougui, A. (2025). Phishing, spyware, and smart city tech: Surveillance in Sisi's Egypt. In T. Roberts & A. Mare (Ed.). *Digital Surveillance in Africa: Power, Agency, and Rights* (pp. 57–84). London,: Zed Books. Retrieved April 30, 2026, from <http://dx.doi.org/10.5040/9781350422117.ch-3>
- 66 Švedkauskas, Zilvinas. 2022. "Digital surveillance, master key for MENA autocrats." In: *Liberty doom? Artificial intelligence in Middle Eastern Security*. EuroMeSCo, Volume 27. <https://www.euromesco.net/wp-content/uploads/2022/07/Policy-Study27.pdf> (1 May 2026).
- 67 Ibid.
- 68 Abrougui, A. "Global Trends in Digital Security: Civil Society & Media". October 2023. Internews. <https://internews.org/wp-content/uploads/2023/11/Global-Trends-in-Digital-Security.pdf> (accessed 30 April 2026).
- 69 Švedkauskas, Zilvinas. 2022. "Digital surveillance, master key for MENA autocrats." In: *Liberty doom? Artificial intelligence in Middle Eastern Security*. EuroMeSCo, Volume 27. <https://www.euromesco.net/wp-content/uploads/2022/07/Policy-Study27.pdf> (accessed 1 May 2026).
- 70 In a zero-day attack, attackers exploit previously unknown and unaddressed software vulnerability.
- 71 Švedkauskas, Zilvinas. 2022. "Digital surveillance, master key for MENA autocrats." In: *Liberty doom? Artificial intelligence in Middle Eastern Security*. EuroMeSCo, Volume 27. <https://www.euromesco.net/wp-content/uploads/2022/07/Policy-Study27.pdf> (accessed 1 May 2026).
- 72 D'Andrea, A., Palak, K. & Guccione, D. "How AI Is Making Phishing Attacks More Dangerous". Keeper. 13 September 2024. <https://www.keepersecurity.com/blog/2024/09/13/how-ai-is-making-phishing-attacks-more-dangerous/> (accessed 1 May 2026).
- 73 Amnesty International. "Automated Apartheid: Facial recognition entrenches the oppression of Palestinians". Undated. <https://banthescan.amnesty.org/opt/index.html> (accessed 1 May 2026).
- 74 Yuval, A. "'Lavender': The AI machine directing Israel's bombing spree in Gaza". +972 Magazine. 3 April 2024. <https://www.972mag.com/lavender-ai-israeli-army-gaza/> (accessed 1 May 2026).
- 75 Article19. "Israel: Stop using biometric mass surveillance against Palestinians". 9 April 2024. <https://www.article19.org/resources/israel-stop-using-biometric-mass-surveillance-against-palestinians/> (1 May 2026).
- 76 Human Rights Watch. "Questions and Answers: Israeli Military's Use of Digital Tools in Gaza". 10 September 2014. https://www.hrw.org/news/2024/09/10/questions-and-answers-israeli-militarys-use-of-digital-tools-in-gaza#_What_are_some (accessed 1 May 2026).
- 77 OHCHR. "Gaza: UN experts deplore use of purported AI to commit 'domicide' in Gaza, call for reparative approach to rebuilding". 15 April 2024. <https://www.ohchr.org/en/press-releases/2024/04/gaza-un-experts-deplore-use-purported-ai-commit-domicide-gaza-call> (accessed 1 May 2026).
- 78 Hasan, M. & Sabbagh, W. "How

- Israel's Unprecedented Killing of Palestinian Journalists in Gaza Makes Accountability Reporting Almost Impossible ". Global Investigative Journalism Network. 6 April 2026. <https://gijn.org/stories/unprecedented-killing-palestinian-journalists-gaza-press-freedom/> (accessed 1 May 2026).
- 79 Committee to Protect Journalists. "Record 129 press members killed in 2025; Israel responsible for 2/3 of deaths". 25 February 2026. <https://cpj.org/special-reports/record-129-press-members-killed-in-2025-israel-responsible-for-2-of-3-of-deaths/> (accessed 1 May 2026).
- 80 Goodfriend, S. "The AI Dragnet". Dissent. 30 April 2025. https://dissentmagazine.org/online_articles/the-ai-dragnet/#:~:text=The%20effects%20were%20nicely%20distilled,at%20the%20Harvard%20Kennedy%20School. (accessed 1 May 2026).
- 81 Privacy International "How predictive policing can be used at protests". 5 May 2021. <https://privacyinternational.org/explainer/4501/how-predictive-policing-can-be-used-protests> (accessed 1 May 2026).
- 82 Privacy International. "Social media monitoring in the UK: the invisible surveillance tool increasingly deployed by government". 17 July 2024. <https://privacyinternational.org/long-read/5337/social-media-monitoring-uk-invisible-surveillance-tool-increasingly-deployed> (1 May 2026).
- 83 Privacy International. "How social media monitoring can be used at a protest". 6 May 2021. <https://privacyinternational.org/explainer/4509/how-social-media-monitoring-can-be-used-protest> (1 May 2026).
- 84 Privacy International. "Social media monitoring in the UK: the invisible surveillance tool increasingly deployed by government". 17 July 2024. <https://privacyinternational.org/long-read/5337/social-media-monitoring-uk-invisible-surveillance-tool-increasingly-deployed> (1 May 2026).
- 85 O Panoptico. "Monitor de Novas Tecnologias na Segurança Pública do Brasil". <https://www.opanoptico.com.br/#publicacoes> (accessed 3 May 2026).
- 86 Isaza, J.J. & Katshir, H. "Brazil Passes Landmark Privacy Law: The General Law for the Protection of Privacy". American Bar Association. 24 April 2020. https://www.americanbar.org/groups/business_law/resources/business-law-today/2020-may/brazil-passes-landmark-privacy-law/ (accessed 1 May 2026).
- 87 CIVICUS. "Migrant support, women and LGBTQI+ rights organisations face suspension". 17 March 2026. <https://monitor.civicus.org/explore/migrant-support-women-and-lgbtqi-rights-organisations-face-suspension/> (accessed 3 May 2026).
- 88 Bychawska-Siniarska, D. and Nowicka, Z. "From Zero to Hero: Once grappling with SLAPPs, Poland could now be a model for how to fight them". Verfassungsblog . <https://verfassungsblog.de/anti-slapp-legislation-poland/> 4 September 2024 (accessed 3 May 2026).
- 89 Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. Big Data & Society, 7(1). <https://doi.org/10.1177/2053951719897945>
- 90 Ibid.
- 91 Ibid.
- 92 Alimardani, M. & Elswah, M. "Digital Orientalism: #SaveSheikhJarrah and Arabic Content Moderation". In POMEPS Studies 43: Digital Activism and Authoritarian Adaptation in the Middle East. August 2021. <https://pomeps.org/>

- [digital-orientalism-savesheikhjarrah-and-arabic-content-moderation](#) (accessed 1 May 2026).
- 93 Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951719897945>
- 94 Kayyali, D. & Altthabani, R. "Vital Human Rights Evidence in Syria is Disappearing from YouTube". Witness. 30 August 2017. <https://blog.witness.org/2017/08/vital-human-rights-evidence-syria-disappearing-youtube/> (accessed 1 May 2026).
- 95 Montgomery, B. & agencies. "Meta wrong to remove graphic Israel-Gaza videos, oversight board says". 19 December 2023. <https://www.theguardian.com/technology/2023/dec/19/facebook-moderation-israel-hamas-videos> (accessed 1 May 2026).
- 96 Olasoji, T. "Facebook, Instagram indiscriminately flag #EndSars posts as fake news". Mail&Guardian. 23 October 2020. <https://mg.co.za/africa/2020-10-23-facebook-instagram-indiscriminately-flag-endsars-posts-as-fake-news/> (accessed 1 May 2026).
- 97 Ricks, B. Surman, M. and Contributors. "Creating Trustworthy AI". Mozilla Foundation. 15 December 2020. <https://www.mozillafoundation.org/en/insights/trustworthy-ai-whitepaper/> (accessed 1 May 2026).
- 98 Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951719897945>
- 99 Nicholas, G. & Bhatia, A. "The Dire Defect of 'Multilingual' AI Content Moderation". *Wired*. 23 May 2023. <https://www.wired.com/story/content-moderation-language-artificial-intelligence/> (accessed 1 May 2026).
- 100 Nicholas, G. & Bhatia, A. "Lost in Translation: Large Language Models in Non-English Content Analysis". CDT. May 2023. <https://cdt.org/wp-content/uploads/2023/05/non-en-content-analysis-primer-051223-1203.pdf> (accessed 1 May 2026).
- 101 Ibid.
- 102 Maréchal, N. & Biddle Roberts, E. "It's Not Just the Content, It's the Business Model: Democracy's Online Speech Challenge". New America Foundation. 17 March 2020. <https://www.newamerica.org/insights/its-not-just-content-its-business-model/> (1 May 2026).
- 103 Diepeveen, S. (2022) Hidden in plain sight: how the infrastructure of social media shapes gender norms. ALIGN Report. London: ODI (<https://www.alignplatform.org/resources/report-hidden-in-plain-sight>).
- 104 <https://docs.un.org/en/A/78/288>
- 105 Ibid.
- 106 United Nations Human Rights Office of the High Commissioner. "Taxonomy of Human Rights Risks Connected to Generative AI: Supplement to B-Tech's Foundational Paper on the Responsible Development and Deployment of Generative AI". <https://www.ohchr.org/sites/default/files/documents/issues/business/b-tech/taxonomy-GenAI-Human-Rights-Harms.pdf> (accessed 1 May 2026).
- 107 Wislon, J. "Hundreds of nonconsensual AI images being created by Grok on X, data shows". *The Guardian*. 8 January 2026. <https://www.theguardian.com/technology/2026/jan/08/grok-x-nonconsensual-images> (1 May 2026).
- 108 <https://docs.un.org/en/A/78/288>
- 109 Chenrose, A. "AI tools used in Kenya to

- discredit protesters and allege Russian connections". DFRLab. 20 December 2024. <https://dfrlab.org/2024/12/20/ai-tools-used-in-kenya-to-discredit-protesters-and-allege-russian-connections/> (accessed 4 May 2026).
- 110 <https://policyful.ai>
- 111 <https://nuharesearch.com/>
- 112 <https://chronicles.nawaat.org/>
- 113 Kamel, R. "Ask Aunty bridges "taboo" conversations in the Middle East". JournalismAI. 2 June 2025. <https://www.journalismai.info/blog/ask-aunty-bridgesnbsp-middle-east-health-info-gap> (accessed 1 May 2026).
- 114 GRIT. "Our Tech". Undated. <https://www.grit-gbv.org/technology> (accessed 1 May 2026).
- 115 Human Rights Funders Network. "Funding at a Crossroads: Foreign Aid Cuts And Implications For Global Human Rights." September 2025. <https://www.hrfn.org/wp-content/uploads/2025/09/Funding-at-a-Crossroads-HRFN-Sept-2025.pdf>
- 116 The process "of using a trained AI model to make predictions on new data". See more at: <https://www.ibm.com/think/topics/ai-inference>
- 117 "A nuanced process that draws on context, trust, and human insight" to make decisions or form opinions.
- 118 African Union. Continental African Intelligence Strategy. July 2024. <https://au.int/sites/default/files/d>
- 119 OECD AI Principles. <https://www.oecd.org/en/topics/ai-principles.html> (accessed 3 May 2026).
- 120 UNESCO. Recommendation on the Ethics of Artificial Intelligence. 16 May 2023. <https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence> (accessed 3 May 2026).
- 121 OHCHR. UN Guiding Principles on Business and Human Rights. 2011. https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinessshr_en.pdf (accessed 3 May 2026).
- 122 CIVICUS. "'Civil society is pushing for transparent, rights-centred AI governance'". 18 March 2025. <https://lens.civicus.org/interview/civil-society-is-pushing-for-transparent-rights-centred-ai-governance/> (accessed 1 May 2026)
- 123 CIVICUS. "Human rights take a backseat in AI regulation". 16 January 2024. <https://lens.civicus.org/human-rights-take-a-backseat-in-ai-regulation/> (accessed 1 May 2026)
- 124 European Center for Not-for-profit Law (ECNL). "The Commission must uphold the AI Act and fundamental freedoms in Hungary". EDRI. 16 October 2025. <https://edri.org/our-work/the-commission-must-uphold-the-ai-act-and-fundamental-freedoms-in-hungary/> (1 May 2026).
- 125 Danes je nov. "A blueprint for success: How Danes je nov dan's advocacy led to a commitment for a Public AI Registry in Slovenia". EDRI. 16 October 2025. <https://edri.org/our-work/a-blueprint-for-success-how-danes-je-nov-dans-advocacy-led-to-a-commitment-for-a-public-ai-registry-in-slovenia/> (accessed 1 May 2026).
- 126 Coalizão Direitos na Rede. "Regulação da IA protetiva de direitos, já! Urgência de transparência e maior participação da sociedade civil nos debates do PL 2338". 10 December 2025. <https://direitosnarede.org.br/2025/12/10/regulacao-da-ia-protetiva-de-direitos-ja-e-maior-participacao-sociedade-civil-pl-2338/> (accessed 1 May 2026).
- 127 Abraham, Y. "'A mass assassination factory': Inside Israel's calculated bombing of Gaza". +972 Magazine. 30

- November 2023. <https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/> (accessed 1 May 2026).
- 128 Yachot, N. "Data is control": what we learned from a year investigating the Israeli military's ties to big tech". The Guardian. 30 December 2025. <https://www.theguardian.com/world/2025/dec/30/israeli-military-big-tech> (accessed 1 May 2026).
- 129 Panoptic Tracker. "Facial Recognition Systems in India". <https://panoptic.in/> (accessed 1 May 2026).
- 130 Verma, D. "What we do in the shadows: IFF seeks transparency in how Indian 'smart governments' are using AI". Internet Freedom Foundation. 26 March 2024. <https://internetfreedom.in/transparency-in-government-ai/> (1 May 2026).
- 131 CIVICUS. "2026 State of Civil Society Report". March 2026. <https://publications.civicus.org/publications/2026-state-of-civil-society-report/> (accessed 1 May 2026).
- 132 Danes je nov dan. "The AI Act isn't enough: closing the dangerous loopholes that enable rights violations". EDRi. 13 November 2025. <https://edri.org/our-work/the-ai-act-isnt-enough-closing-the-dangerous-loopholes-that-enable-rights-violations/> (accessed 1 May 2026).
- 133 Yakura, S., Albagli, D., Dokei, T., Mitchell, A.M., Fujino, M., Kamiya, M. and Jackson, J. "Japan's first AI legislation becomes law – Focus is on promoting research and development; no monetary penalties". 14 April 2026. White & Case. <https://www.whitecase.com/insight-alert/japans-first-ai-legislation-becomes-law-focus-promoting-research-and-development-no> (accessed 1 May 2026).
- 134 Mulrenan, S. "Technology: South Korea hopes to become AI leader with Asia's first comprehensive legal regime". International Bar Association. 23 March 2026. <https://www.ibanet.org/Technology-South-Korea-hopes-to-become-AI-leader-with-Asias-first-comprehensive-legal-regime> (accessed 1 May 2026).
- 135 Singh, J. "India orders social media platforms to take down deepfakes faster". TechCrunch. 10 February 2026. <https://techcrunch.com/2026/02/10/india-orders-social-media-platforms-to-take-down-deepfakes-faster/> (accessed 1 May 2026).
- 136 Ibid.
- 137 Ibid.
- 138 PIB. India AI Governance Guidelines. 15 February 2026. <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2228315&reg=3&lang=2> (accessed 3 May 2026).
- 139 SDAIA. AI Ethics Principles. <https://sdaia.gov.sa/en/SDAIA/about/Documents/ai-principles.pdf> (accessed 3 May 2026).
- 140 White & Casey. "AI Watch: Global regulatory tracker – Singapore". 4 March 2026. <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-singapore> (3 May 2026).
- 141 OECD. Recommendation of the Council on Artificial Intelligence. 22 May 2019. <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449> (accessed 3 May 2026).
- 142 African Union. Continental Artificial Intelligence Strategy. July 2024. <https://au.int/sites/default/files/d>
- 143 Kapiyo, V. Personal interview with the author. 5 November 2025.
- 144 García, J.M. Personal interview with the author. 25 November 2025.
- 145 CIVICUS. "Global Summary: civic space

- dynamics". 2025. https://monitor.civicus.org/globalfindings_2025/innumbers/ (accessed 1 May 2026).
- 146 Cruzi, Y. Personal Interview with the author. 18 November 2025.
- 147 Belli, Luca and Curzi, Yasmin and Britto Gaspar, Walter, AI Regulation in Brazil: Advancements, Flows, and need to Learn from the Data Protection Experience (April 16, 2023). Available at SSRN: <https://ssrn.com/abstract=6080726>
- 148 <https://www.techpolicy.press/brazils-ai-law-faces-uncertain-future-as-big-tech-warms-to-trump/>
- 149 Hickok, E., Mohan, S., Pielemeier, J. and Kakkar, J.M. "Multistakeholder Promises and Power Gaps in Global AI Summits". 17 March 2026. Tech Policy Press. <https://www.techpolicy.press/multistakeholder-promises-and-power-gaps-in-global-ai-summits/> (accessed 3 May 2026).
- 150 Civil Society Alliances for Digital Empowerment (CADE). "Civil society groups call for stronger participation in UN AI governance dialogue". 23 April 2026. <https://cadeproject.org/updates/civil-society-groups-call-for-stronger-participation-in-un-ai-governance-dialogue/> (3 May 2026).
- 151 CIVICUS. "'AI-powered weapons depersonalise the violence, making it easier for the military to approve more destruction'". 23 November 2024. <https://lens.civicus.org/interview/ai-powered-weapons-depersonalise-the-violence-making-it-easier-for-the-military-to-approve-more-destruction/> (accessed 1 May 2026).
- 152 No Azure for Apartheid. "The Workers' Toolkit". Undated. <https://noazureforapartheid.com/toolkit/> (accessed 1 May 2026).



Digital Democracy Initiative

The Digital Democracy Initiative (DDI) is a programme to safeguard inclusive democracy and human rights in the digital age. It focuses on supporting local civil society in the Global South, particularly in countries undergoing democratic regression and where civic space is under pressure.

For more information visit:
digitaldemocracyinitiative.net



CIVICUS

CIVICUS is a global alliance of civil society organisations and activists working to strengthen citizen action and civil society throughout the world.

For more information visit:
civicus.org



Digital Democracy
Initiative

CASE STUDY:

AFRICAN LANGUAGES AND NATURAL LANGUAGE PROCESSING (NLP):

CHALLENGES AND
OPPORTUNITIES FOR
CIVIL SOCIETY



CIVICUS



Acknowledgments

Author

Afef Abrougui (Fair Tech)

Design and layout

Monica Molano

We would like to acknowledge and thank all the civil society actors, researchers, and practitioners who generously shared their time and insights through interviews and expert consultations for this research



Digital Democracy Initiative

The Digital Democracy Initiative (DDI) is a programme to safeguard inclusive democracy and human rights in the digital age. It focuses on supporting local civil society in the Global South, particularly in countries undergoing democratic regression and where civic space is under pressure.

For more information visit:

digitaldemocracyinitiative.net



CIVICUS

CIVICUS is a global alliance of civil society organisations and activists working to strengthen citizen action and civil society throughout the world.

For more information visit:

civikus.org



Of the
7000
languages spoken
worldwide, Africa is home to

1/3

of them, making it the most
linguistically diverse continent.

Of the 7,000 languages spoken worldwide, Africa is home to a third of them, making it the most linguistically diverse continent.¹ Despite this diversity, African languages are under-represented in Natural Language processing (NLP), a subfield of artificial intelligence and computational linguistics concerned with enabling computers to understand, generate, and manipulate human language.²

NLP techniques are fundamental to the digital tools and platforms we use daily. They are deployed to detect spam, predict what we are typing, moderate content we post online, recognise our speech and convert it into text or to perform an action, for example, with voice assistants. Large Language Models (LLMs) deploy NLP techniques, and they represent a significant advancement in the field. They are trained in large amounts of datasets and perform a wide range of tasks.³ They power Generative AI models and are a key component of the content moderation operations of social media platforms.

High quality datasets that are free from biases, inaccuracies, and errors are essential in the field of NLP, and large language datasets are necessary for the development of LLMs. These datasets are needed for the training of machine learning models, and when high quality data sets are scarce, there are wide implications for people and their human rights. For instance, their content can be censored because a model misunderstood it to be a violation of platform policies or harmful content. Advocates for African languages, digital rights, and accessibility understand the implications this has in content moderation.

NLP

Natural Language Processing

Natural Language Processing: (NLP): A subfield of computer science and linguistics concerned with understanding and analysing human language. Widely deployed in many of the technologies and software people use on a daily basis, NLP combines statistical models, machine learning, and linguistic rules to perform a wide range of tasks including sentiment analysis, machine translation, and text classification.



"We have a lot of people on these social media platforms who are communicating in local languages. And so, you find that some of the content that is in local languages cannot be detected by the platforms... Moderation systems will also not address content that is in local languages. Sometimes you find harmful content being disseminated widely or being amplified by algorithms. The lack of investment in training the models in local languages means that users will be affected."

— Victor Kapiyo, senior researcher at CIPESA

A number of factors are behind the under-representation of African languages in NLP, including a colonial past that marginalised African languages,⁴ a painful legacy that post-colonial rulers continued by retaining European languages as official or semi-official languages, and not taking adequate measures to preserve local languages.⁵ NLP systems were also not built to meet the needs of many African languages which are structurally different from English, a dominant high-resource language in NLP,⁶ or have various substantial dialectal variations. Another challenge is the lack of access to modern computing power and cloud services necessary for training models.⁷

All these obstacles exclude African languages from the NLP field. As researchers on the challenges and opportunities of NLP in low-resourced Senegalese languages noted: "Despite their societal importance and widespread use in daily communication and media, these languages remain largely excluded from the digital and scientific landscape of NLP. This gap poses a dual challenge: the risk of technological

marginalisation of major linguistic communities, and the missed opportunity to harness NLP for advancing locally grounded research, especially in the social sciences".⁸

The African Union has recognised the language challenge in its Continental African Intelligence Strategy, which includes high level recommendations to promote investment in NLP in Indigenous and local African languages.⁹ The Strategy's fifth guiding principle 'Inclusion and Diversity' states that "the production, development, use and assessment of AI in Africa will be inclusive, non-discriminatory, leaving no one and no place behind and benefiting everyone, and respectful of the diversity of African people, cultures, languages, gender dimensions and nations".

To address this gap, civil society and academic initiatives have been launched across the continent. Masakhane, "a grassroots organisation whose mission is to strengthen and spur NLP research in African languages, for Africans, by Africans"¹⁰ has been building open and high-quality data sets that support African language. This includes text and voice datasets in East African languages to support "building natural language processing models that range from spell and grammar checking, sentiment analysis, topic modeling, text summarization, misinformation and fake news detection, machine translation and automatic speech recognition models".¹¹ Another project by the Masakhane community resulted in "the first large, publicly available, high-quality dataset for named entity recognition (NER) in ten African languages".¹² An important task in NLP, NER is not only essential for automated products and functions like spell-checking and voice assistant but "it also enables identifying African names, places, and organizations for information retrieval".¹³ In another initiative, Code for Africa runs an AI Sandbox with its self-hosted LLMs allowing

partners to prototype and experiment with AI models in a secure environment and with the support of a team of engineers.¹⁴ AI sandboxes are isolated testing environments where developers and researchers can experiment with AI and trial new models.

Despite the lack of high-quality data and challenges in access to infrastructure and computing power, civil society in Africa has been experimenting with building their own small models, trained on small datasets and for specific purposes rather than the general-purpose LLMs. Civil society is “leveraging AI for advocacy, factchecking, monitoring (of elections, hate speech and disinformation), digital literacy, public accountability, and community engagement”, according to CIPESA’s 2025 State of Internet Freedom Report, which assessed implications of AI on digital democracy in 14 African countries.¹⁵



“In the women’s sexual health and reproductive sector, a lot of the organisations have been able to draw patterns based on intake of complaints or issues to do with women’s SRHR [Sexual and Reproductive Health and Rights] issues. And they have been able to conduct research both, qualitative and quantitative, and some have built small models. You hear them saying we have built this app, we have built this tool to just see if it can teach women to do this, and when you explore some of these tools, one or another, you find that actually you see it is based out of data they have collected”.

— Irene Mwendwa, Nairobi-based consultant on technology, AI and law

She mentioned as an example, Zuzi, an “African, and trauma-informed” AI chatbot developed by South African organisation GRIT

to provide information on the rights and legal services available to survivors of gender-based violence.¹⁶

In other examples in Tunisia, independent media outlet Nawaat launched Nawaat Chronicles,¹⁷ an AI-enhanced archive platform allowing readers, researchers, students and journalists to easily access and search an archive of over 22 years of independent journalism on Tunisia in Arabic, French, and English. In Kenya, Mzalendo Watch has been developing AI tools to facilitate access to parliamentary data and proceedings.¹⁸ The latter is one of several other examples of CSOs across the continent, including in Nigeria, South Africa, and Zimbabwe, developing AI tools to facilitate access to and analysis of official documents and improve transparency.¹⁹ In Senegal, Jangat, an AI-powered platform, allowed voters to compare the electoral programmes of candidates in the 2024 election.²⁰

“I have been able to see a lot of opt-in by citizens, opt-in from a point of civic duty or civic participation, where individuals are just using their knowledge on building technologies as computer scientists, as data scientists, some self-taught, some [are] young people who do not even have jobs, to really showcase how technology can be used to create social cohesion and civic engagement or public participation here in Kenya and other countries”, said Mwendwa. She added that it is unfortunate that participation and creativity are “inhibited” by government restrictions such as app restrictions and network shutdowns, as was the case during the Tanzanian election in 2025. “Africa can also harness the power of these same AI technologies to grow our democracies, our economic opportunities. At the same time, that same prowess is being used against them [African youth], communicating two very different things for young people who are just trying to seek opportunities”, she added.

REFERENCES

- 1 UNESCO. "UNESCO and the promotion of languages in Africa: cultural diversity and multilingualism". 27 February 2025. <https://www.unesco.org/en/articles/unesco-and-promotion-languages-africa-cultural-diversity-and-multilingualism> (accessed 1 May 2026).
- 2 Comic, M. "NLP vs. LLMs: What's the difference"? Lokalise. 15 May 2025. <https://lokalise.com/blog/nlp-vs-llm/> (accessed 1 May 2026).
- 3 Ibid.
- 4 Masakhane. "Our Mission". <https://www.masakhane.io/home> (accessed 1 May 2026)
- 5 Amorim, R, Baltazar, R. & Soares, I. "Linguistic legacies of British and Portuguese (de)colonization in Africa: (un)successful common bonds"? Centro de Administração e Políticas Públicas, Instituto Superior de Ciências e Políticas Públicas, Universidade de Lisboa, Portugal. 2020 (accessed 1 May 2026).
- 6 Korir, K. "The Untapped Potential: African Languages in Natural Language Processing". Medium. 27 February 2025. <https://medium.com/@kiplangatkorir/the-untapped-potential-african-languages-in-natural-language-processing-7478b78ef0bd> (1 May 2026).
- 7 Mbaye, D., Mbengue, T. D. P., Seye, M. R., Diallo, M., Ndiaye, M. L., Adjanohoun, D. S., Sow, D., Wade, C. S., Munyaka, J. C. B., & Chenal, J. (2026). Opportunities and Challenges of Natural Language Processing for Low-Resource Senegalese Languages in Social Science Research. Preprints. <https://doi.org/10.20944/preprints202601.1124.v1>
- 8 Ibid.
- 9 African Union. Continental Artificial Intelligence Strategy. July 2024. https://au.int/sites/default/files/documents/44004-doc-EN-Continental_AI_Strategy_July_2024.pdf (accessed 1 May 2026).
- 10 Masakhane. "Our Mission". <https://www.masakhane.io/home> (accessed 1 May 2026)
- 11 Masakhane. "MakerereNLP: Text & Speech for East Africa". Undated. <https://www.masakhane.io/ongoing-projects/makererenlp-text-speech-for-east-africa> (accessed 1 May 2026).
- 12 David Ifeoluwa Adelani, Jade Abbott, Graham Neubig, Daniel D'souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, Stephen Mayhew, Israel Abebe Azime, Shamsuddeen H. Muhammad, Chris Chinenye Emezue, Joyce Nakatumba-Nabende, Perez Ogayo, Aremu Anuoluwapo, Catherine Gitau, Derguene Mbaye, Jesujoba Alabi, Seid Muhie Yimam, Tajuddeen Rabi Gwadabe, Ignatius Ezeani, Rubungo Andre Niyongabo, Jonathan Mukiibi, Verrah Otiende, Irero Orife, Davis David, Samba Ngom, Tosin Adewumi, Paul Rayson, Mofetoluwa Adeyemi, Gerald Muriuki, Emmanuel Anebi, Chiamaka Chukwuneke, Nkiruka Odu, Eric Peter Wairagala, Samuel Oyerinde, Clemencia Siro, Tobius Saul Bateesa, Temilola Oloyede, Yvonne Wambui, Victor Akinode, Deborah Nabagereka, Maurice Katusiime, Ayodele Awokoya, Mouhamadane MBOUP, Dibora Gebreyohannes, Henok Tilaye, Kelechi Nwaike, Degaga Wolde, Abdoulaye Faye, Blessing Sibanda, Orevaoghene Ahia, Bonaventure F. P. Dossou, Kelechi Ogueji, Thierno Ibrahima DIOP, Abdoulaye Diallo, Adewale Akinfaderin, Tendai

Marengereke, and Salomey Osei. 2021. MasakhaNER: Named Entity Recognition for African Languages. Transactions of the Association for Computational Linguistics, 9:1116–1131.

13 Ibid.

14 Code for Africa. “Actionable Information”. Undated. <https://codeforafrica.org> (accessed 1 May 2026).

15 CIPESA. State of Internet Freedom in Africa Report. 2025. <https://cipesa.org/2025/09/state-of-internet-freedom-in-africa-report/> (accessed 1 May 2026).

16 “Our Tech”. GRIT. Undated. <https://www.grit-gbv.org/technology> (accessed 2 May 2026).

17 <https://chronicles.nawaat.org/>

18 Mzalendo. “AI Tools”. <https://mzalendo.com/ai-tools/> (accessed 1 May 2026).

19 CIPESA. State of Internet Freedom in Africa Report. 2025. <https://cipesa.org/2025/09/state-of-internet-freedom-in-africa-report/> (accessed 1 May 2026).

20 Ibid.



Digital Democracy Initiative

The Digital Democracy Initiative (DDI) is a programme to safeguard inclusive democracy and human rights in the digital age. It focuses on supporting local civil society in the Global South, particularly in countries undergoing democratic regression and where civic space is under pressure.

For more information visit:

digitaldemocracyinitiative.net



CIVICUS

CIVICUS is a global alliance of civil society organisations and activists working to strengthen citizen action and civil society throughout the world.

For more information visit:

civicus.org



Digital Democracy
Initiative

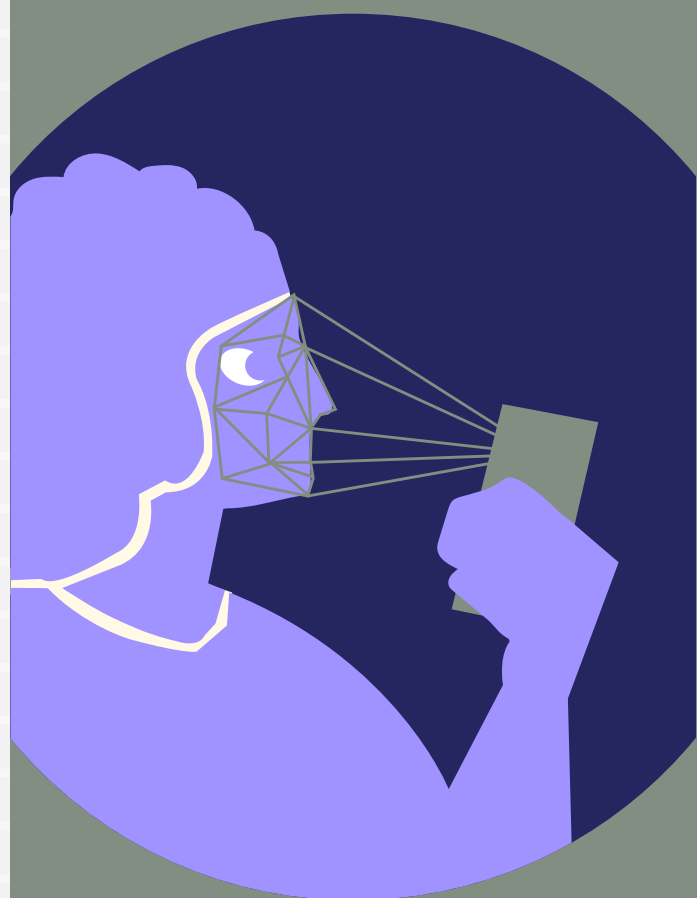
CASE STUDY:

**AI ACCOUNTABILITY
IN THE CONTEXT OF
GENOCIDE AND
OCCUPATION:**

“LARGER THAN
JUST AI ITSELF”



CIVICUS



Acknowledgments

Author

Afef Abrougui (Fair Tech)

Design and layout

Monica Molano

We would like to acknowledge and thank all the civil society actors, researchers, and practitioners who generously shared their time and insights through interviews and expert consultations for this research



Digital Democracy Initiative

The Digital Democracy Initiative (DDI) is a programme to safeguard inclusive democracy and human rights in the digital age. It focuses on supporting local civil society in the Global South, particularly in countries undergoing democratic regression and where civic space is under pressure.

For more information visit:

digitaldemocracyinitiative.net



CIVICUS

CIVICUS is a global alliance of civil society organisations and activists working to strengthen citizen action and civil society throughout the world.

For more information visit:

civikus.org



Palestinians in the Occupied Palestinian Territories (OPT) feel the impact of AI technologies deployed by Israel to control and subjugate them. Israel deploys biased and flawed algorithms to monitor what Palestinians are saying online under the pretext of preventing terror attacks,¹ and thereby criminalising Palestinians' calls for resistance to the occupation and an end to the genocide.² In the West Bank, networks of facial recognition track and monitor Palestinians, which hinders their freedom of movement, association, and right to peaceful assembly. The surveillance can also restrict their rights to access health services, engage in entrepreneurial activities, and worship freely.³ In the genocide in Gaza, the Israeli military has used machine learning tools that rely on mass (biometric) surveillance to track, generate, and bomb targets,⁴ in gross violation of International Humanitarian Law (IHL), in particular military necessity, proportionality, distinction between combatants and civilians, and precaution.⁵

The private sector, including tech companies, has been benefiting from the Israeli occupation and the genocide in Gaza.⁶ For example, the Israeli military and the apartheid regime rely on cloud storage provided by tech giants, including Google, Amazon, and Microsoft, to collect massive amounts of data on Palestinians through surveillance and privacy-infringing practices.

Local, regional, and international civil society has engaged in different tactics and strategies to investigate and document the impacts of these technologies on Palestinians and hold the developers, deployers, and enablers to account.



"We are looking at a situation where Israel is using AI tools in a horrendous, horrifying way and we are calling for accountability for any company or anyone who is providing them with this assistance. It is how we attempt to hold Israel accountable for war crimes and especially if we are able to impact the private companies who are involved in providing assistance to power Israel's AI systems and tools"

— Jalal Abukhater, Policy Manager at Tamleh, regarding his organisation's approach to AI accountability in the OPT.

He added that, "targeting Israel is not an easy task. We do not do that. We do not engage Israel. Israel is committing war crimes. We are engaging the companies who are seen to be complicit in those war crimes by providing assistance when it comes to AI". The organisation has also advocated at the EU level to call for stricter controls on exports of EU-manufactured technologies to Israel, which are then deployed against Palestinians.

Tech workers also have been organising through campaigns such as No Azure for Apartheid⁷ and No Tech for Apartheid to demand that Microsoft, Google, and Amazon end their cloud storage contracts with the Israeli government. Campaigns include internal petitions, sit-ins, walk-outs and public pressure. In some cases, employees have faced retaliation⁸ for their activism, including termination, doxxing, and surveillance⁹. Activists have also protested at events such as the United Nations AI for Good Summit in Geneva in July 2025, calling out those actors that have enabled Israel's AI-assisted genocide against Palestinians in Gaza¹⁰ as well as at a Google event¹¹ at the AI+ Expo in Washington D.C.¹²

Sustained pressure has paid off, in some cases, with Microsoft announcing in September 2025 that it had terminated an infamous Israeli military unit's access to "specific cloud storage and AI services and technologies", although the company still provides other services to the Israeli military.¹³ The unit in question, Unit 8200, runs a massive surveillance program that has "collected, stored, replayed and analyzed millions of civilian phone calls from Gaza and the West Bank", as documented by a joint investigation between the Guardian, Israeli-Palestinian publication +972 Magazine and the Hebrew-language media outlet Local Call.¹⁴

In addition, civil society groups and researchers successfully pushed for a ban on fully autonomous weapons at the UN Convention on Certain Conventional Weapons,¹⁵ with the UN Secretary-General calling for Member States to set clear regulations and prohibitions on such systems by 2026.¹⁶ Strategic litigation has been used as a path of resistance against European governments for their exports of weapons to Israel, for example, in the UK¹⁷ and the Netherlands.¹⁸

Despite these efforts, and evidence of the impacts of these technologies, accountability remains elusive. Globally, amid an unchecked AI race, there is a lack of willingness to institute protections from AI's most salient harms. For Palestinians in the West Bank and Gaza, the occupation deprives them of the most basic protection, including due process and self-determination. AI accountability starts with holding Israel to account for its occupation and war crimes, as Abukhater reiterated:

Sustained pressure, tangible results



July 2025

Activists protest at the UN AI for Good Summit in Geneva, and at a Google event at the AI+ Expo in Washington D.C.



September 2025

Microsoft announces it terminated Unit 8200's access to "specific cloud storage and AI services and technologies"



2026

Deadline set by the UN Secretary-General for Member States to establish clear regulations and prohibitions on fully autonomous weapons systems

"There is no infrastructure to protect people, to regulate the use of AI tools. A big obstacle is that we are under occupation, there is no sovereignty, there is no self-determination. People are not able to just go to a court and say that my rights are being violated by the use of this AI system, and this court would listen to us, there is no such thing. The AI in Palestine is being used to commit crimes against humanity in a systematic and regular way. So, for us, accountability starts with accountability for Israel not just on AI, but to hold Israel accountable for the illegal occupation of occupied territory. We call on countries to respect their obligations and duties towards not supporting Israel's occupation, and there is an ICJ court ruling that ordered Israel to cease or end the occupation, and third countries have a responsibility to respect the ICJ orders. So for us, the issue of accountability is much larger than just AI itself".

The AI technologies of repression that Israel deploys to subjugate and control Palestinians and commit genocide in Gaza have even wider repercussions for the entire world. Israel's military industrial complex, a leading exporter of military and surveillance technology around the world, has long made Palestine a testing ground for its latest weapons and technologies of repression, including border and surveillance systems, and spyware and attack drones. Palestine is "Israel's workshop, where an occupied nation on its doorstep provides millions of subjugated people as a laboratory for the most precise and successful methods of domination", wrote Antony Loewenstein in his book, *The Palestine Laboratory*.¹⁹

As Dima Samaro, a Palestinian lawyer and researcher, and director of Skyline International for Human Rights, told CIVICUS in an interview in July 2025, "Gaza serves as a warning".²⁰ "As militarised AI becomes normalised, the language of 'precision' and 'efficiency' masks atrocity. The lesson from Gaza is clear: when AI-powered machines control who lives, human rights die. This transcends Palestine's tragedy – it foreshadows everyone's future", she added.

Despite the obstacles and the lack of resources, civil society groups and activists in Palestine and beyond continue to demand accountability and justice through public campaigns, strategic litigation, disruptive actions, worker campaigns and boycotts. The different campaigns and coalitions that have emerged from this fight need sustained support in the face of powerful governments that refuse to regulate dangerous AI systems and an industry with the resources and influence to undermine any form of regulation that could reign in their technologies and therefore affect their profits.



"Gaza serves as a warning. As militarised AI becomes normalised, the language of 'precision' and 'efficiency' masks atrocity. The lesson from Gaza is clear: when AI-powered machines control who lives, human rights die. This transcends Palestine's tragedy – it foreshadows everyone's future"

— Dima Samaro, Palestinian lawyer and researcher, director of Skyline International for Human Rights.



"The obstacles we are facing is that we are going up against very big, large companies, huge companies. And our concerns are not really backed up by institutions or a regulatory body that is actually influential or powerful. We are talking as a civil society with very limited resources on those big issues, but we are facing a larger problem when it comes to lack of regulation, lack of infrastructure and lack of capacities to influence the policies on the AI tools that are used here"

— Jalal Abukhater, Policy Manager at Tamleh

REFERENCES

- 1 AP. "Israel claims 200 attacks predicted, prevented with data tech". 12 June 2018. <https://www.cbsnews.com/news/israel-data-algorithms-predict-terrorism-palestinians-privacy-civil-liberties/> (accessed 4 May 2026).
- 2 Aljazeera. "Dareen Tatour sentenced to five months in prison over poem". 31 July 2018. <https://www.aljazeera.com/news/2018/7/31/dareen-tatour-sentenced-to-five-months-in-prison-over-poem> (accessed 30 April 2026).
- 3 Amnesty International. "Israel and Occupied Palestinian Territories: Automated Apartheid: How facial recognition fragments, segregates and controls Palestinians in the OPT". <https://www.amnesty.org/en/documents/mde15/6701/2023/en/> 2 May 2023. (accessed 4 May 2026).
- 4 Human Rights Watch. "Questions and Answers: Israeli Military's Use of Digital Tools in Gaza". 10 September 2014. https://www.hrw.org/news/2024/09/10/questions-and-answers-israeli-militarys-use-of-digital-tools-in-gaza#_What_are_some (accessed 1 May 2026).
- 5 Düz, S. & Koçakoğlu, M.S. 'Deadly Algorithms: Destructive Role of Artificial Intelligence in Gaza War'. Foundation For Political, Economic And Social Research. February 2020. <https://media.setav.org/en/file/2025/02/deadly-algorithms-destructive-role-of-artificial-intelligence-in-gaza-war.pdf>
- 6 OHCHR. "A/HRC/59/23: From economy of occupation to economy of genocide – Report of the Special Rapporteur on the situation of human rights in the Palestinian territories occupied since 1967". 2 July 2025. <https://www.ohchr.org/en/documents/country-reports/ahrc5923-economy-occupation-economy-genocide-report-special-rapporteur> (accessed 4 May 2026).
- 7 No Azure for Apartheid. "The Workers' Toolkit". Undated. <https://noazureforapartheid.com/toolkit/> (accessed 1 May 2026).
- 8 No Azure for Apartheid. "Microsoft: End Complicity in Apartheid and Genocide". Action Network. Undated. <https://actionnetwork.org/petitions/microsoft-end-complicity-in-apartheid-and-genocide> (accessed 4 May 2026).
- 9 Day, M. and Ford, B. "Microsoft Asked FBI for Help Tracking Palestinian Protests". Bloomberg. 26 August 2025. <https://www.bloomberg.com/news/articles/2025-08-26/microsoft-asked-fbi-for-help-with-israel-gaza-protests?srnd=phx-technology&embedded-checkout=true> (accessed 4 May 2026).
- 10 Binnur Donmez, B. "Protest in Geneva slams AI for Good Summit over big tech's reported role in Gaza war". Anadolu Agency. 28 July 2025. <https://www.aa.com.tr/en/europe/protest-in-geneva-slams-ai-for-good-summit-over-big-techs-reported-role-in-gaza-war/3625487> (accessed 4 May 2026).
- 11 Ibid.
- 12 CODEPINK. ""You are killing people in Palestine!" CODEPINK and No Tech for Apartheid Disrupt Google at AI+ Expo". 3 June 2025. <https://www.codepink.org/googledisruptjune3> (accessed 4 May 2026).
- 13 Amnesty International. "Israel/IOPT: Microsoft's move to block Israeli military unit's access to its mass surveillance technology is a moment for corporate reckoning". 26 September 2025. <https://www.amnesty.org/en/latest/news/2025/09/microsoft-block-israel-military-unit-from-using-its-rapporteur> (accessed 4 May 2026).

- [technology/](#) (accessed 4 May 2026).
- 14 Ibid.
- 15 CIVICUS. "AI-powered weapons depersonalise the violence, making it easier for the military to approve more destruction". 23 November 2024. <https://lens.civicus.org/interview/ai-powered-weapons-depersonalise-the-violence-making-it-easier-for-the-military-to-approve-more-destruction/> (accessed 1 May 2026).
- 16 "UN News. 'Politically unacceptable, morally repugnant': UN chief calls for global ban on 'killer robots'". 14 May 2025. <https://news.un.org/en/story/2025/05/1163256> (accessed 4 May 2026).
- 17 Amnesty International. "Case of Al Haq v Secretary of State for Business and Trade: Written submissions of Human Rights Watch and Amnesty International UK". 13 May 2025. <https://www.amnesty.org.uk/knowledge-hub/all-resources/submission-aiuk-ban-arms-sales-israel/> (accessed 4 May 2026).
- 18 ELSC. "Court appeal: confronting the role of the Dutch state in genocide". <https://elsc.support/court-appeal-confronting-the-role-of-the-dutch-state-in-genocide/> (accessed 4 May 2026).
- 19 Loewenstein, A. The Palestine Laboratory: How Israel exports the technology of occupation around the world. Verso Books, 2023
- 20 CIVICUS. "The lesson from Gaza is clear: when AI-powered machines control who lives, human rights die". 16 July 2025. <https://lens.civicus.org/interview/the-lesson-from-gaza-is-clear-when-ai-powered-machines-control-who-lives-human-rights-die/> (accessed 3 May 2026).



Digital Democracy Initiative

The Digital Democracy Initiative (DDI) is a programme to safeguard inclusive democracy and human rights in the digital age. It focuses on supporting local civil society in the Global South, particularly in countries undergoing democratic regression and where civic space is under pressure.

For more information visit:

digitaldemocracyinitiative.net



CIVICUS

CIVICUS is a global alliance of civil society organisations and activists working to strengthen citizen action and civil society throughout the world.

For more information visit:

civikus.org



Digital Democracy
Initiative

CASE STUDY:

**BRAZILIAN CIVIL
SOCIETY'S LONG
FIGHT FOR HUMAN
RIGHTS IN AI
GOVERNANCE**



CIVICUS



Acknowledgments

Author

Afef Abrougui (Fair Tech)

Design and layout

Monica Molano

We would like to acknowledge and thank all the civil society actors, researchers, and practitioners who generously shared their time and insights through interviews and expert consultations for this research



Digital Democracy Initiative

The Digital Democracy Initiative (DDI) is a programme to safeguard inclusive democracy and human rights in the digital age. It focuses on supporting local civil society in the Global South, particularly in countries undergoing democratic regression and where civic space is under pressure.

For more information visit:

digitaldemocracyinitiative.net



CIVICUS

CIVICUS is a global alliance of civil society organisations and activists working to strengthen citizen action and civil society throughout the world.

For more information visit:

civikus.org

In December 2024, the Federal Senate in Brazil approved AI-related legislation, Bill No. 2338/2023. This comprehensive law sets out principles, rules, and mechanisms for regulating AI in a way that is aligned with fundamental rights and democratic values. It adopts a risk-based approach to AI governance, similar to the EU AI Act, establishing three tiers of AI systems.¹ Systems that pose excessive risk shall be prohibited; high risk shall be heavily regulated; and other systems must meet basic requirements of regulation. The legislation also mandates assessments of algorithmic impact for high-risk systems and human oversight and strict requirements for AI use in the public sector. It further establishes several rights for individuals interacting with AI systems, including the right to explanation and the right to contest decisions that significantly impact them.

At a glance: Bill 2338/2023

- **Prohibited:** systems that pose excessive risk
- **Heavily regulated:** high-risk systems
- **Basic requirements:** all other systems

This law was first proposed in 2023 but had to be reviewed and voted on by the House of Representatives and adopted by the President. It is still unclear when it will be adopted and how the bill will look before it is adopted. However, the law is considered a win for civil society in Brazil which has been advocating for human rights-based regulation of AI.²

In fact, another bill that was first proposed to

regulate AI in the country, draft Bill no.21/2020 adopted “a very light touch and minimalistic approach” with “several substantial gaps, being either laconic or vague as regards some key elements such as risk assessment and oversight mechanisms”.³ This bill “was really favoured by the industry,” said Yamin Curzi, a professor at the Centre for Technology and Society at Fundação Getulio Vargas (FGV) Law School in Rio de Janeiro. “But of course, civil society pushed a lot against this, because it would not give the process of accountability and the due process that we needed to have, actually, governance of the AI systems. We would have only the industry regulating itself. So then, civil society pushed against this. We had a lot of push by civil society organizations back then and a lot of debates, also in academia”. As a result of the civil society pressure, including Coalizão Direitos na Rede (CDR), a coalition of 51 academic and civil society organisations in defense of digital rights,⁴ Brazil’s Senate launched multistakeholder public hearings and consultations for a substitute bill.



academic and civil
society organisations
form the **Coalizão
Direitos na Rede (CDR)**

As part of its efforts, CDR, which has conducted campaigning and advocacy on AI regulation in the country, published petitions,⁵ sent letters to lawmakers,⁶ submitted input for an alternative bill,⁷ and engaged directly with members of parliament.⁸ CDR has also

conducted legal analysis and campaigned for a ban on the use of facial recognition technology in public spaces.⁹

While civil society has been successful in pushing back against the initial AI draft bill and shaping a more robust alternative, a long road still lies ahead, as the tech industry and lawmakers can still undermine human rights-based regulation of AI. Already the version of Bill No. 2338/2023 that was approved by the Senate in December 2024 had been re-shaped. For example, under pressure from technology companies, algorithmic systems to recommend content were not listed under the high-risk category and most references to ‘information integrity’ or ‘freedom of expression’ were removed after Big Tech companies and members of the opposition equated such references with ‘censorship’.¹⁰ Pressure came not only from Big Tech companies, as Laís Martins, a Brazilian investigative journalist based in São Paulo, wrote for Policy Tech Press: “Advocates for a weaker regulatory approach (or no regulation at all) were joined by Brazilian industry groups and data center executives, all echoing a big tech narrative that regulation would hinder innovation. To some extent, their efforts were successful in weakening what started out as a very ambitious bill”.

“Both bills are still alive because people are trying to push for AI regulation also with this principled logical approach still nowadays...Bill 2338/2023 is being discussed. It was approved in the National Chamber, but because when the Senate approves a bill, it goes back to the National Deputy Chamber for revisions, and then again goes to the Senate, to be sent out to the federal government for final ratification, before it is signed to be enforced. They try to retrieve some of the due process rules because of the lobby of big tech, especially by Microsoft, but also, AI enterprises and associations of industries, especially linked

to the agribusiness here in Brazil because we have tons of AI deployments in agribusiness”, asserted Professor Curzi.

While civil society continues the fight,¹¹ they face obstacles in their advocacy efforts: “most of these [government] agencies are based in Brasilia, the capital of Brazil. Civil society is not there. Civil society is in Rio, in Sao Paulo, in the Northeast but not in Brasilia. It is really expensive to have a lawyer. It is really expensive for them to do advocacy all the time. And for the enterprises, it is not. They go all the time there, they have this capacity of being in private spaces with the regulators..and they established links to and relationships with these people [regulators and lawmakers]”, Curzi declared.

In the meantime, Brazilians belonging to excluded groups, including Black and Indigenous people, continue to face discrimination in AI systems. Facial recognition cameras are widely deployed in public and private spaces with little protection, transparency, and oversight, thus posing risks to privacy, freedom of expression, and freedom of association. Brazil’s civic space was rated as obstructed by the CIVICUS Monitor in 2025, a rating it has retained since 2018.

“

“It is really expensive to have a lawyer. It is really expensive for them to do advocacy all the time. And for the enterprises, it is not. They go all the time there, they have this capacity of being in private spaces with the regulators.”

— Yasmin Curzi, professor, Centre for Technology and Society, FGV Law School.

REFERENCES

- 1 Atanasovska, D. and Robeli, L. "Brazil's AI Act: A New Era of AI Regulation". GDPR Local. 26 February 2026. <https://gdprlocal.com/brazils-ai-act-a-new-era-of-ai-regulation/> (accessed 1 May 2026).
- 2 Coalizão Direitos na Rede. "Regulação iA O Que Tenho a Ver Com Isso". August 2023. <https://direitosnarede.org.br/campanha/inteligencia-artificial-o-que-eu-tenho-a-ver-com-isso/> (accessed 1 May 2026).
- 3 Belli, Luca and Curzi, Yasmin and Britto Gaspar, Walter, AI Regulation in Brazil: Advancements, Flows, and need to Learn from the Data Protection Experience (April 16, 2023). Available at SSRN: <https://ssrn.com/abstract=6080726> or <http://dx.doi.org/10.2139/ssrn.6080726>
- 4 Coalizão Direitos na Rede. "CARTA DE APOIO AO PL 2338/2023". 14 June 2023. <https://direitosnarede.org.br/2023/06/14/carta-de-apoio-ao-pl-2338-2023/> (accessed 1 May 2026).
- 5 Coalizão Direitos na Rede. "¡Firma la petición! POR UNA REGULACIÓN DE IA PARTICIPATIVA Y RESPONSABLE". 14 April 2023. <https://direitosnarede.org.br/2023/04/14/firma-la-peticion-por-una-regulacion-de-ia-participativa-y-responsable/> (accessed 1 May 2026).
- 6 Coalizão Direitos na Rede. "Regulação da IA protetiva de direitos, já! Urgência de transparência e maior participação da sociedade civil nos debates do PL 2338". 10 December 2025. <https://direitosnarede.org.br/2025/12/10/regulacao-da-ia-protetiva-de-direitos-ja-e-maior-participacao-sociedade-civil-pl-2338/> (accessed 1 May 2026).
- 7 Coalizão Direitos na Rede. "Regulação de Inteligência Artificial: Contribuição da CDR para a Comissão de Juristas do Senado". 6 July 2022. <https://direitosnarede.org.br/2022/07/06/regulacao-de-inteligencia-artificial-contribuicao-da-cdr-para-a-comissao-de-juristas-do-senado/> (accessed 1 May 2026).
- 8 Coalizão Direitos na Rede. "Coalizão Direitos na Rede intensifica diálogo sobre IA no Congresso e defende regulação com foco na proteção de direitos". 5 November 2025. <https://direitosnarede.org.br/2025/11/05/coalizao-direitos-na-rede-intensifica-dialogo-sobre-ia-no-congresso-e-defende-regulacao-com-foco-na-protecao-de-direitos/> (accessed 1 May 2026).
- 9 "A Call for General Ban on the Use of Facial Recognition Technologies on Public Security". <https://tiremeurostodasuamira.org.br/en/home-eng/> (accessed 1 May 2026).
- 10 Martins, L. "Brazil's AI Law Faces Uncertain Future as Big Tech Warms to Trump". Tech Policy Press. 4 February 2025. <https://www.techpolicy.press/brazils-ai-law-faces-uncertain-future-as-big-tech-warms-to-trump/> (accessed 1 May 2026).
- 11 Coalizão Direitos na Rede. "Regulação da IA protetiva de direitos, já! Urgência de transparência e maior participação da sociedade civil nos debates do PL 2338". 10 December 2025. <https://direitosnarede.org.br/2025/12/10/regulacao-da-ia-protetiva-de-direitos-ja-e-maior-participacao-sociedade-civil-pl-2338/> (accessed 1 May 2026).



Digital Democracy Initiative

The Digital Democracy Initiative (DDI) is a programme to safeguard inclusive democracy and human rights in the digital age. It focuses on supporting local civil society in the Global South, particularly in countries undergoing democratic regression and where civic space is under pressure.

For more information visit:
digitaldemocracyinitiative.net



CIVICUS

CIVICUS is a global alliance of civil society organisations and activists working to strengthen citizen action and civil society throughout the world.

For more information visit:
civikus.org