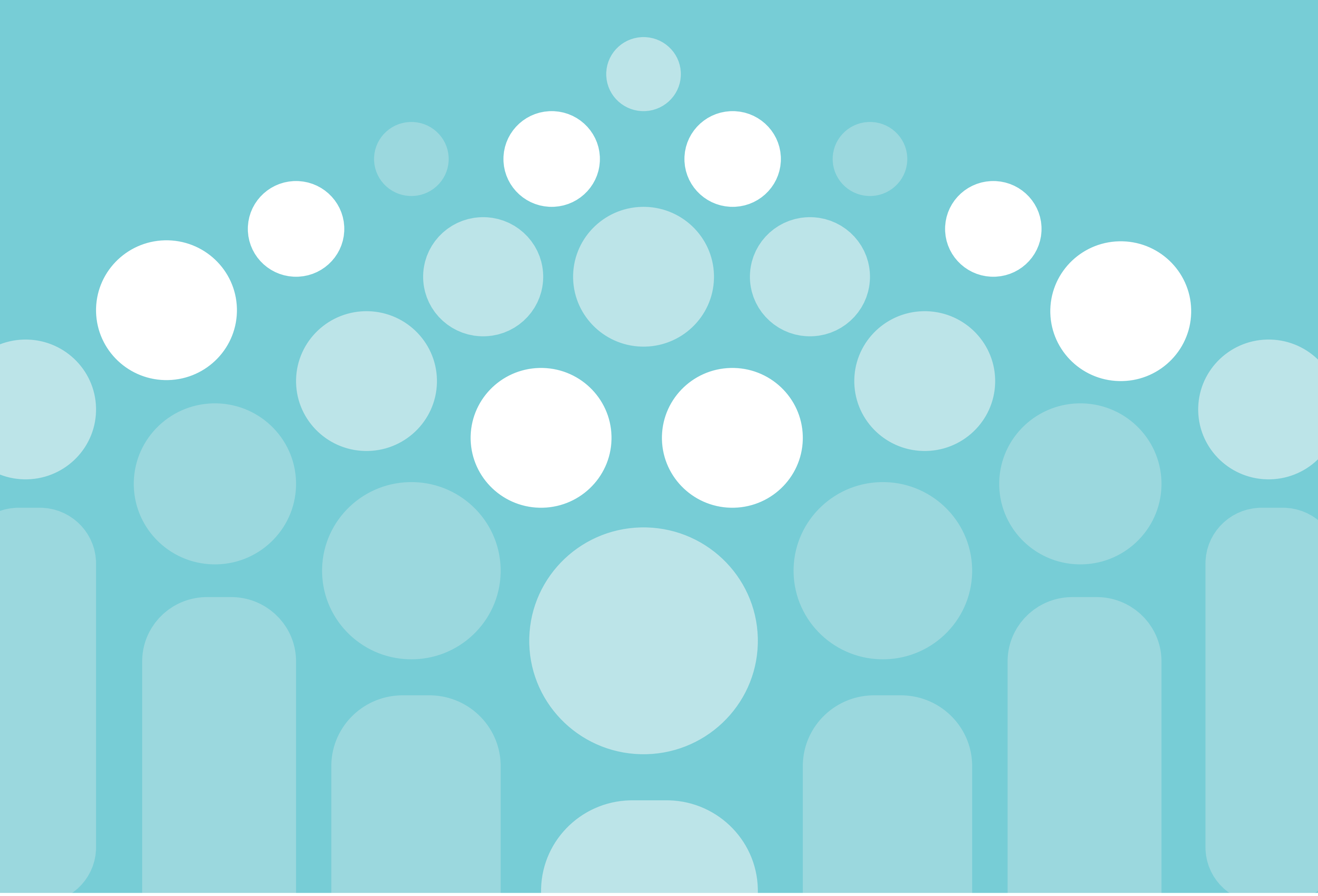

INVESTOR ALLIANCE FOR HUMAN RIGHTS
AN INITIATIVE OF ICCR

SECTOR-WIDE RISK ASSESSMENT: Information and Communications Technology (ICT)

SALIENT ISSUE BRIEFING: Artificial Intelligence-based Technologies

CONTENTS

INVESTOR ALLIANCE FOR HUMAN RIGHTS
AN INITIATIVE OF ICCR

SECTOR-WIDE RISK ASSESSMENT:
Information, Communications and Technology (ICT)

SALIENT ISSUE BRIEFING:
Artificial Intelligence-based Technologies

According to the European Union Commission, an Artificial Intelligence (AI) system is a "machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments." The rapid and widespread growth of innovations in AI is generating significant economic and social benefits, such as improved access to information or the provision of goods and services. However, the use of AI can also pose risks and have a detrimental impact on human rights if designed, managed, or deployed without consideration of real and potential adverse impacts on users and society. This dual nature of AI presents significant opportunities and challenges that need to be adequately balanced for the inclusive and equitable progress of the technology and the well-being of society.

The OECD defines Generative AI (genAI) as a category of AI that can create new content such as text, images, videos, and music. According to the OECD, genAI has the potential to revolutionize entire industries and society but also exacerbates challenges that policymakers must confront. As GenAI gained widespread attention globally in late 2022 with the creation of text-to-image generators and Large Language Models (LLMs) like ChatGPT, all references made to AI in this briefing apply to genAI as well, unless otherwise indicated.

HOW DOES AI IMPACT HUMAN RIGHTS

The UN International Bill of Rights, which consists of the Universal Declaration of Human Rights (UDHR), the International Covenant on Civil and Political Rights (ICCPR), and the International Covenant on Economic, Social, and Cultural Rights (ICESCR), is the binding international system applicable in understanding the impact of AI on human rights. Applied in the context of the UN Guiding Principles on Business and Human Rights, it provides the framework to understand and assess, the duty of companies to respect human rights in their design, deployment and operation of AI systems. See Annex A for key international and regional AI governance laws and standards.

Privacy and Data Protection

> *The right to privacy is recognized in Article 12 of the Universal Declaration of Human Rights (UDHR) and Article 17 of the International Covenant on Civil and Political Rights (ICCPR).*

In 2021, the UN High Commissioner of Human Rights published a report about the widespread use of AI by states and businesses, including profiling, automated decision-making, and machine learning technologies that interfere with the right to privacy, primarily through the increased collection and use of personal data, often without informed and specific consent. AI often relies on the large-scale collection, storage, and processing of personal data without users' knowledge or effective and informed consent.

AI-driven systems, such as automated assistants and facial recognition technologies (FRTs), often collect vast amounts of data without users' informed consent, raising serious privacy concerns. For genAI, companies have intensively harvested data for training, including protected and copyrighted content from the arts and entertainment industry. FRTs are increasingly pervasive, particularly in neighborhoods of color, where their presence is most concentrated. These AI systems collect data on individuals at home, work, and in public spaces, enabling profiling, tracking, and identification that fundamentally alters expectations of privacy. Often operating without transparency or accountability, these systems sort, score, and rank individuals, leaving little room for recourse. Other companies increasingly use AI-powered tools to enhance their operations or deploy their products and services.

For more details, please refer to our Salient Risk Briefing on Privacy and Data Protection.

Freedom of Opinion and Expression

> *Article 19 of the UDHR states that the right to freedom of opinion and expression "includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers."*

A report of the UN's Special Rapporteur highlights the chilling effect on freedom of expression by disinformation and other forms of manipulation of online content. Social media platforms, for example, depend on algorithmic decision-making to rank posts or moderate and remove content on the feeds of users, effectively acting as gatekeepers to freedom of expression. Algorithmic filters may remove content on human rights abuses or block activist groups, thereby limiting public awareness. Additionally, bot accounts and doxxing tactics amplify harassment, often targeting vulnerable groups, such as women, journalists, and activists, pressuring them into self-censorship. Governments, especially authoritarian regimes, may exploit AI for selective, retroactive censorship and predictive control of dissent, using online censorship to silence opposition. AI-driven surveillance, such as facial recognition, further chills freedom of expression, as people may self-censor to avoid becoming targets.

For more details, please refer to our Salient Risk Briefing on Freedom of Opinion and Expression.

Conflict and Security

> *The UDHR states that "everyone has the right to life, liberty and security," which is reaffirmed through Articles 6(1) and 9(1) of the ICCPR. Article 20(2) of the ICCPR prohibits "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence."*

Governments, particularly authoritarian regimes, can weaponize AI-driven technology for pre-emptive and retroactive censorship, silencing dissent before it even emerges. By analyzing past behavior, social connections, and speech patterns of specific individuals, AI can identify potential critics in advance, enabling authorities to erase, manipulate, or suppress online content at will. This level of predictive control stifles opposition but also ensures that only state-approved narratives dominate public discourse. Moreover, AI-enhanced surveillance through facial recognition can arbitrarily restrict the freedom of movement of people and lead to unlawful detention, torture, forced disappearances, and killings.

The design and deployment of AI-based technologies by companies in conflict zones may significantly impact security and human rights. Algorithmic content moderation often amplifies "sensational" content, which may include extremist views or incitements to violence by governments and non-state groups. For instance, the spread of disinformation or "deep fakes" may heighten risks for vulnerable communities in conflict-affected regions, as seen with the rapid dissemination of false narratives during multiple crises. Many countries have indicated an interest in using facial recognition to track specific groups, like journalists, human rights defenders, or religious minorities, posing serious threats to individual liberty.

Lethal autonomous weapons further increase the risks of harm in conflict settings, as they operate without direct human intervention. AI-driven predictive tools, such as those for pretrial risk assessment, policing "hotspots," (e.g., Shotspotter's use in majority-minority neighborhoods with racially discriminatory impact) or integrating social media with public databses (e.g., the use of ClearviewAI's facial recognition database) that may lead to arbitrary detentions, undermining rights to liberty and security.

For more details, please refer to our Salient Risk Briefing on Conflict and Security.

Non-Discrimination

> *Articles 2 and 7 of the UDHR states, "Everyone is entitled to all the rights and freedoms without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status" and "all are equal before the law and are entitled to equal protection against any discrimination in violation of this Declaration and against any incitement to such discrimination."*

A 2024 report titled "Racism and AI: Bias from the past leads to bias in the future", by the Office of the High Commissioner for Human Rights, highlights widespread concerns about the harmful assumption that technology is neutral and objective, leading to the perpetuation of racial discrimination by AI-based technologies.

The lack of diversity in AI research and development exacerbates existing societal harms by embedding historical biases into technological systems. Facial recognition algorithms have been shown to misidentify women and people of color at disproportionately high rates, leading to wrongful arrests and reinforcing systemic discrimination in law enforcement. Deepfake technology and nonconsensual nude imagery overwhelmingly target women, particularly those from marginalized communities, perpetuating gendered power imbalances and digital violence (e.g., the Lensa app).

Often referred to as the "WEIRD" (Western, Educated, Industrialized, Rich, Democratic) bias, most AI developers and datasets are rooted in Western cultural and geographic regions, resulting in AI systems reflecting predominantly Western values, assumptions, and perspectives. Analysis of widely used genAI training datasets has revealed troubling patterns of emotion detection, misogyny, explicit content, and harmful stereotypes. Algorithmic decision-making tools replicate the biases of their creators or the skewed training data sets used, leading to discriminatory outcomes in areas like pretrial detention (e.g., COMPAS, the risk assessment tool's bias against people of color), housing (e.g., enabling discrimination in advertising), employment (e.g., Amazon's biased hiring tool), and credit (e.g., algorithms lowering minorities' credit scores due to biased or missing socio-economic data). Hence, proactive efforts need to be made to address such biases in data sets that serve as the backbone for many AI applications and spread across platforms.

Predictive policing tools, which rely on biased historical data to "predict" where crime may occur, have led to disproportionate surveillance, especially among communities of color. This has created a vicious cycle where predictions drive intensified policing in certain communities, further entrenching discrimination as more biased data is collected and reinforced. Unregulated data collection through biometric identification and mass surveillance technologies, like those used by Palantir to assist ICE in tracking migrants and human rights defenders, exacerbate the targeting of vulnerable communities, violating privacy and civil liberties.

For more details, please refer to our Salient Risk Briefing on Discrimination.

Political Participation

> *Civil liberties, essential for democracy as detailed in Article 21 of the UDHR and Article 25 of the ICCPR, establish our right "to take part in the conduct of public affairs, directly or through freely chosen representatives," that ensure that the "will of the people shall be the basis of the authority of government," and that "this shall be expressed in periodic and genuine elections which shall be by universal and equal suffrage."*

AI assists technology companies in moderating content on public platforms based on terms of service and government requests, but its limitations pose significant risks, including election interference. Algorithms often struggle with contextual understanding, leading to the removal of legitimate political discourse while allowing misinformation to spread unchecked. Additionally, as discussed above, AI systems created by technology companies and trained on biased datasets may perpetuate systemic discrimination, disproportionately silencing marginalized groups or political opponents. Government requests for content removal can further complicate this issue, as they may be used to suppress dissenting voices, shape public perception, or control dominant narratives. This creates a risk of election interference, ultimately undermining democratic processes and public trust.

AI and its underlying algorithms and datasets significantly impact human rights, particularly the right to political participation, which relies on access to accurate information and the freedom to engage in political activities without fear of discrimination, surveillance, or reprisals. AI can enhance political engagement by enabling personalized communication (e.g., chatbots for voter Q&A), improving access to information (e.g., videos in different languages, AI-powered fact-checking, etc.), and boosting participation (e.g., automated voting reminders). It also combats misinformation and promotes inclusivity with tools like real-time translation in different languages and accessibility features, albeit with some limitations. However, biased datasets can undermine these benefits by distorting political messaging, amplifying existing inequalities, and limiting representation. If AI systems rely on skewed data, they may reinforce discriminatory narratives, suppress marginalized voices, or disproportionately flag content from certain groups as misinformation. The lack of transparency in automated decision-making further restricts scrutiny and access to remedies. Gender bias plays a role as well, as women politicians are disproportionately targeted by AI-driven tools, including deepfakes, which can be weaponized to discredit and silence them.

There are three applications of AI in the information environment that raise concerns, including for the right to political participation:

1. **Content display and personalization**, where algorithms may create echo chambers by reinforcing biases and limiting exposure to diverse political views;

2. **Content moderation and removal**, where AI systems may inadvertently remove legitimate content due to context misinterpretation or discriminatory assumptions; and

3. **Profiling, advertising, and targeting**, where micro-targeting can manipulate voters, spread disinformation, and suppress voter participation, threatening diverse political discourse.

For more details, please refer to our Salient Risk Briefing on Political Participation.

THE 'BUSINESS CASE' FOR RIGHTS-RESPECTING AI

Companies that design, use, or deploy AI-enabled technologies must proactively identify, assess, and address the salient human rights risks discussed above. Failure to do so can and has resulted in legal, reputational, and financial consequences. Human rights organizations have criticized technology companies for their business practices and algorithms that contribute to human rights abuses, including egregious ones in conflict settings. The commercial aspects of AI are increasingly being scrutinized for their negative impacts on human rights, demanding action from technology leaders to prioritize responsible AI development and build public trust to enhance the reputation and long-term sustainability of the business.

Legal and Reputational Risks:

In Myanmar and Ethiopia, Facebook's algorithm was implicated in spreading hate speech that fuelled violence against the Rohingya minority and Tigrayans, respectively, contributing to allegations of genocide. In Ethiopia, a $1.6 billion lawsuit was filed in a Kenyan court based on claims that Meta failed to adequately moderate hate speech, contributing to the death of an Ethiopian professor after Facebook posts reportedly incited violence against him based on his ethnicity. This lawsuit is especially significant because Meta is not formally registered in Kenya, but because the alleged violations occurred in that jurisdiction, the court deemed Meta liable.

Financial Risks:

- There is a growing trend in the technology industry where intense competition and pressure to be first-to-market often leads to the premature release of AI and other emerging technologies, potentially compromising user safety. After the launch of Google's Bard AI, the company's share prices dropped by $100 billion in market value when the chatbot shared inaccurate information in a promotional video. This incident underscored the financial risks of inadequate product testing in the rush to compete, specifically in response to OpenAI's ChatGPT.
- Surveillance technologies deployed by companies such as Clearview AI have raised significant privacy concerns, as their facial recognition software has been used without individuals' consent and in ways that could lead to unlawful surveillance and tracking. Clearview AI has faced multiple fines (e.g., $22 million in France and $33.1 million in the Netherlands) in different countries due to privacy concerns related to its facial recognition practices.

Reputational Risks:

- In July 2024, ASN Impact Investors, a major Dutch investment firm, urged TKH Group, a technology company that focuses on customer-centric solutions for sustainable innovations, to implement human rights due diligence policies within a year or face divestment. This was in response to Amnesty International's revelation that TKH Group was using surveillance cameras in occupied East Jerusalem, allegedly facilitating human rights violations against Palestinians.
- Amazon has faced significant penalties due to its AI-driven employee surveillance practices, particularly for excessively monitoring productivity in warehouses. Recently, the French data protection authority fined Amazon $35 million for overstepping privacy boundaries in its French warehouse operations.

These examples highlight the urgent need for technology leaders to conduct comprehensive impact assessments on an ongoing basis and in particular, prior to the market launch of AI applications and services. This involves engaging with affected stakeholders and ensuring transparency in their operations and decision-making processes. By proactively addressing these human rights risks, technology companies would not only avoid adverse consequences that may open them up to financial and legal risks but also contribute positively to societal well-being and uphold fundamental human rights.

The opacity of AI decision-making, referring to the lack of transparency in how algorithms process data and make decisions, raises significant concerns about the accuracy and fairness of using personal data, as well as individuals' rights over their information. The adaptability of machine-learning systems, while a strength, can also weaken human oversight, making it harder to anticipate and mitigate potential threats to human rights. As emphasized in Article 14 of the EU AI Act, human agency must remain central to AI governance, ensuring that technology serves rather than undermines the safety and fundamental rights of people. This places a clear responsibility on technology companies to implement robust oversight mechanisms that prevent AI systems from inflicting harm or perpetuating systemic injustices.

Moreover, regulatory landscapes are constantly evolving, with increasing demands for transparency and accountability. This includes the imposition of strict requirements on data privacy and AI governance. Companies that anticipate and comply with these regulations would avoid hefty penalties and gain a competitive edge by demonstrating their commitment to responsible practices. A major challenge is that technology has outpaced regulations, making compliance a complex issue. AI systems already in use may lack the transparency and traceability required by emerging global regulations. Investors may need to consider the potential requirement for companies to backtrack and adapt, especially to meet standards like the EU AI Act.

In conclusion, the commercial aspects of the development and deployment of AI are under increasing scrutiny. Technology companies can mitigate legal, reputational, and financial risks while fostering trust, resilience, and sustainable business practices. This strategic approach not only ensures compliance with evolving regulations but also enhances long-term value, business viability, and beneficial societal impact.

HUMAN RIGHTS GUIDANCE FOR RIGHTS-RESPECTING AI DEVELOPMENT AND DEPLOYMENT

Grounded in the UN Guiding Principles on Business and Human Rights (UNGPs), this guidance helps investors engage with both technology companies developing AI and other companies using or deploying it, ensuring they prevent, mitigate, and address key human rights impacts.

Application of UNGPs	Implementation	Good practice recommendation
Develop or update human rights policy commitment to embed responsible AI principles aligned with international laws, standards, and frameworks	Companies should consult with civil society and other stakeholders in developing or updating policy commitments on AI that recognize international human rights frameworks. The policy should be discussed, reviewed, and approved by senior leadership, distributed internally, and shared publicly with all users, customers, business partners, and suppliers through terms of service, codes of conduct, or contracts. Furthermore, the policy's implementation should be adequately resourced, including with appropriate expertise (internal and/or external) and regularly reported to senior leadership, including the Board of Directors.	Companies should establish corporate policies, training, and oversight mechanisms, as well as technical guidance for all personnel (senior leadership, engineers, product designers, data enrichment workers, etc.) involved in the life cycle of AI-based products and services (design, deployment, and implementation) that protect human rights and guide all business operations and relationships.

Assess actual and potential impacts of development and deployment of AI-based technologies to identify and prioritize salient human rights risks	Through engagement with affected stakeholders and relevant experts, companies should identify and assess actual and potential adverse impacts that AI-based technological products and services in their value chain, with a specific focus on high-risk use cases of AI, can have on their consumers/users and society at large. This should also include assessing how AI is being used to drive business choices and decision-making within their business models. The focus should be on addressing salient risks to human rights, particularly in vulnerable and marginalized communities.	Impact assessments should be conducted on an ongoing process, before entering new markets, new business relationships, and new or updated technological applications, especially in new or changed operating environments. In-depth, stand-alone assessments may be needed for high-risk use cases of AI and deployment, such as in conflict-affected and high-risk countries (e.g., data centers in Saudi Arabia) as well as severe actual and potential impacts – AI systems that reinforce discrimination, child safety online, surveillance, and political repression related to facial recognition technologies or risks to freedom of expression and information, privacy.
Integrate responsible AI principles into company policies and activities based on findings of assessments	Drawing on regular impact assessments, companies should integrate the assessment findings across relevant internal functions and processes (with clear assignment of roles, responsibilities, and resources and senior-level oversight). Where a company is causing or contributing to real or potential adverse impacts due to its AI-based products and services, it should take steps to promptly cease the activity or use its leverage to mitigate the impact. Where companies are directly linked to adverse impacts through business relationships, they should seek to increase and utilize their leverage to address them.	Incorporate public Voluntary Commitments from Leading Artificial Intelligence Companies (adopted on July 21, 2023). Increase leverage by acting collectively. For example, companies may join multistakeholder initiatives (e.g., Tech Coalition or the Global Partnership on Artificial Intelligence) for technology companies working to address risks of AI-based products and services. Take steps to increase accessibility of AI-based products and services while ensuring they respect the needs of impacted marginalized communities.
Monitor performance to ensure the effectiveness of measures to address adverse impacts	Companies should use appropriate qualitative and quantitative indicators and draw on feedback from internal and external stakeholders and experts, including impacted communities and other stakeholders in tracking performance.	Document the amount of content removed historically, on what grounds, and lessons learned (e.g., the shutdown of Meta's CrowdTangle threatens the availability of archival data).

		Post-market monitoring to ensure that AI systems using algorithms, machine learning, and automated decision-making are not having unintended negative impacts, such as discriminatory outcomes. The data sets used to train the AI system, the algorithm/model selected, and the data inputs (use cases) should be fully auditable, and audit results should be disclosed.
Publicly communicate efforts to address the human rights impacts of the use of AI	Adequate information must be included to ensure evaluation of the company's efforts to respond to salient risks emerging from the use of AI-based technologies. Separate formal reporting should occur where risks of severe impacts exist, for example, Meta's human rights impact assessments.	Companies should communicate, among other things: when and how AI technologies are deployed; the logic used by those systems; policies that direct their use; which decisions are made by automated systems and/or human review; and when personal data will become part of a dataset and how it will be used. This information should, at a minimum, be contained in annual human rights and sustainability reports.
Remediate harms	When companies cause or contribute to adverse impacts on human rights, they should provide for or cooperate in remediation through legitimate judicial and non-judicial grievance mechanisms, as appropriate. Operational level mechanisms must be aligned with the UNGPs effectiveness criteria, including being accessible to impacted communities. Outcomes of the grievance mechanism should flow into risk assessment processes.	Create reporting mechanisms (e.g., Microsoft's anonymous reporting on human rights practices) so that impacted communities can raise concerns about violations of human rights. These mechanisms should be open to workers, suppliers, CSOs, communities, whistleblowers, journalists, etc.

GUIDING QUESTIONS ON RIGHTS RESPECTING AI

The following questions are intended as a starting point for investors who are engaging with technology companies to help them evaluate if the companies are making adequate efforts to fulfill their responsibility to respect human rights in their operations and business relationships related to AI-based technologies.

Human rights commitment

→ Has the company adopted a public-facing policy commitment to responsible AI in line with international standards? If yes, does the policy include commitments to human rights? Is the policy endorsed by the Board and CEO of the company?

→ Does the company have safeguarding policies and measures to protect users of their AI-based products and services?

→ Does the commitment apply to suppliers and other business partners throughout the company's value chain, which includes private corporate customers and state actors/customers, and are they communicated as a responsibility, guiding business relationships?

Governance

→ Does the Board exercise direct oversight over AI-related risks to the human rights of users? Does Board membership or appointed board committees include people with expertise and relevant experience (external specialists, policy professionals, human rights advocates, domain experts) on issues related to AI? Does the Board regularly review connections between the company's operations related to its AI-based products and services and its impacts on human rights?

→ Are the company's lobbying and political contributions aligned with commitments to protect human rights?

→ Is respect for human rights incorporated into the company's business strategy and processes for developing and deploying its AI-based technologies?

Assessing impacts, risks, and opportunities

→ Does the company conduct a human rights risk assessment, either as a stand-alone or integrated assessment, of all its operations that specifically focuses on the development and deployment of AI technologies and its impacts?

→ Has the company identified and assessed the most salient risks to human rights caused or exacerbated by their development and/or deployment of AI technologies, including unintended risks or harms?

→ Does the company report externally on salient human rights impacts and how often does the company publish such reports? Has the company shared learnings from any risk and impact assessments on human rights in industry forums?

→ Does the company consult with stakeholders in the assessment of human rights? How frequently does the company have consultations with human rights experts? How are their inputs integrated into business activities, particularly as it relates to the company's development or deployment of AI technologies?

Responding to risks and impacts

→ Has the company developed steps to prevent and mitigate adverse impacts of its AI-based products and services on human rights and how does it evaluate the effectiveness of its efforts?

→ Does the company participate in multi-stakeholder engagements with industry peers, NGOs, civil society organizations, and governments that support the advancing of human rights and addressing risks that AI poses to human rights?

- How does the company use its leverage in business relationships to reduce adverse impacts on human rights?

- Does the company have policies on integrating safety-by-design principles into developing its AI-based products and services? How does the company ensure AI systems are not reinforcing bias and causing unintended harm to marginalized communities due to biased data sets?

Embedding commitments internally and externally

- How will the company disseminate its AI and human rights commitments to internal and external stakeholders and throughout its value chain? How does the company ensure its partners, suppliers, and other key relationships are following the company's policy commitments? Does the company offer training to personnel and business partners on its human rights commitments?

- How does the company address cases of users, including private corporate customer and state actors and customers, violating the terms of services of its products and services?

Ensuring access to remedy

- Does the company provide or participate in timely, accessible, and effective grievance mechanisms to offer impacted communities and/or their representatives access to remedy when their rights have been harmed?

- Does the company periodically assess and disclose information about the effectiveness of grievance mechanisms?

- Does the company proactively notify users when they believe harm has occurred?

ANNEX A:
International and Regional Standards Related to Development and Deployment of AI

The following are the main international and regional policy frameworks that underscore the human rights risks resulting from the design and use of AI-based technologies. These frameworks are aligned and complement each other in their AI principles.

2018

The European Union (EU) passed the General Data Protection Regulation (GDPR), a privacy and security law that imposes obligations on organizations across the world if they target or collect data related to people in the EU. A unique feature of the GDPR is a provision that enables users to withdraw consent that they may have previously granted for data that was already made public.

2021

UNESCO adopted the first global standard on AI titled "Recommendation on the Ethics of Artificial Intelligence", which is applicable to all 194 member states. The Recommendation prioritizes human rights and dignity, emphasizing transparency, fairness, and the need for human oversight in AI systems.

2023

The UN General Assembly adopted a resolution on the promotion and protection of human rights in the context of digital technologies. Its clause 20(b) highlights the need "to prevent harm to individuals caused by artificial intelligence systems and to refrain from or cease the use of artificial intelligence applications that are impossible to operate in compliance with international human rights law or that pose undue risks to the enjoyment of human rights, unless and until the adequate safeguards to protect human rights and fundamental freedoms are in place".

MARCH 2024

The UN General Assembly adopted a landmark resolution on the promotion of "safe, secure and trustworthy" AI systems that will also benefit sustainable development for all. In September 2024, the Global Digital Compact was opened for endorsement as a comprehensive framework for global governance of digital technology and artificial intelligence.

MAY 2024

The OECD AI principles were updated and adopted by 47 countries, including the G20 countries, to focus on international cooperation in fostering AI development that is ethical and promotes economic growth. The principles promote the use of AI that is innovative and trustworthy and that respects human rights and democratic values. These principles provided the "first intergovernmental standard for trustworthy AI, focusing on human-centered values, transparency, fairness, and accountability."

The Council of Europe Framework Convention on artificial intelligence and human rights, democracy, and the rule of law was adopted by the Council of Europe Committee of Ministers on May 17. It has been signed by the European Union, the United States of America, the United Kingdom, Israel, Norway, Georgia, Andorra, Iceland, San Marino, and the Republic of Moldova. The treaty provides a legal framework covering the entire lifecycle of AI systems. It promotes AI progress and innovation, while managing the risks it may pose to human rights, democracy, and the rule of law. To remain effective over the long term, the regulatory framework is designed to be adaptable, applying consistently across different technologies, including future innovations.

JULY 2024

Africa's Continental AI Strategy, endorsed by the African Union Executive Council in July 2024, calls for unified national approaches among African Union Member States to navigate the complexities of AI-driven change, aiming to strengthen regional and global cooperation and position Africa as a leader in inclusive and responsible AI development.

AUGUST 2024

In the European Union, the Artificial Intelligence Act (EU AI Act) was signed into law in August 2024 as a single regulatory framework for AI to be applied mandatorily across all member countries, including the private sector. The AI Act seeks to regulate the development and use of AI-based technologies by providing developers and deployers of AI-based solutions with clear obligations and requirements related to the specific uses of AI, including requiring fundamental rights impact assessments (FRIAs) prior to the deployment of high-risk AI systems (e.g., biometric systems). A detailed overview of the key provisions is available here. The EU AI Act aims to promote the adoption of 'trustworthy AI' while safeguarding the rights of impacted individuals and communities. While already signed into law, the majority of its provisions will come into force in August 2026. It offers a tiered, risk-based approach to different AI systems. The EU AI Act has extra-territorial reach as it covers (1) companies developing and deploying within the EU; (2) companies domiciled outside the EU but deploying in the EU; and (3) information transferring via the EU.

JANUARY
2025

In the United States, the White House's Executive Order dated 23 January 2025 calls for the removal of barriers to American leadership in AI by establishing policy for promoting human progress, economic competitiveness, and national security. The U.S. Government is currently drafting an AI Action Plan based on these goals. The now-repealed Blueprint for an AI Bill of Rights and the 2023 Executive Order on Safe, Secure, and Trustworthy AI constituted a set of standards and guidelines to address AI challenges and opportunities. It provided for the government executive departments to formulate industry standards, guidelines, practices, and regulations for AI development and usage. This included calls for comprehensive action to strengthen AI design, AI safety and security, protect privacy, advance equity, prevent algorithmic discrimination, promote innovation, and more.

MARCH
2025

The ASEAN Responsible AI Roadmap (2025-2030) provides actionable steps for policymakers and stakeholders in the region to create conditions that facilitate the development of responsible AI in the region and for member states to leverage and enable AI in a meaningful, impactful, and sustainable manner by 2030.

Many countries are laying the foundations of AI governance through national and regional legislation, which are influenced by and, in some instances, have paved the way for global standards and regulations on AI-based technologies. The Center for AI and Digital Policy assesses the AI policies and practices of different countries across the world in its AI and Democratic Values Index. Canada, Japan, Netherlands, Korea, United Kingdom lead in the rankings. The 2025 edition can be found here.

ANNEX B: Investor Efforts

Investors are taking steps to prevent and mitigate adverse impacts on human rights by holding technology companies accountable for their design, deployment, and use of AI-based technologies. Here are some examples:

→ **Investor Engagement Initiatives:** There are several collective investor initiatives focused on tech companies' corporate accountability, including the Investor Alliance's digital rights and AI accountability engagement launched in 2021. Other efforts include the World Benchmarking Alliance's Collective Impact Coalition for Ethical AI to advance ethical AI policies and practices of technology companies, the Council on Ethics of Sweden's collaborative investor engagement with seven global technology giants to strengthen their management of human rights risks and impacts, and Candriam's facial recognition tech (FRT) initiative and report advocating for safe FRT development and deployment.

→ **Shareholder Proposals:** In 2024, investors filed 14 proposals with technology companies calling for transparency and accountability in the development and use of AI across their business operations and provision of services, citing adverse impacts on users and society due to the lack of oversight and accountability as a material risk to corporate value. Some examples include: Arjuna Capital and Open Mic's proposals with Microsoft, Meta and Alphabet on the misinformation and disinformation risks of Generative Artificial Intelligence tools, SHARE's and Mercy Investments proposal on the human rights impact of AI-driven advertising practices filed with Alphabet and Meta respectively, Trillium Asset Management Corporation's proposal with Alphabet on the issue of Board Oversight and governance matters in relation to AI principles, and AFL-CIO's proposal with Amazon on human rights issues as well as Apple and other companies demanding greater transparency on AI deployment. A number of these proposals have been refiled in 2025 and together, they build on the shareholder proposals filed previously in 2022 and 2023 to raise human rights concerns related to inadequate content moderation, proliferation of hate speech, lack of transparency and accountability due to opaque algorithms and AI, violations of privacy rights, and advertising business model risks.

→ **Investor Advocacy for Rights-Respecting Regulation:** Investors have understood the importance of regulatory measures for responsible business conduct, thereby enabling rights-respecting decisions throughout the investment lifecycle. The EU Digital Services Act was supported in a statement signed by 65 investors representing over US 8.7 trillion in assets, urging for additional measures to mitigate risks associated with algorithmic driven surveillance in online advertising and content management. The investor statement on the EU AI Act was endorsed by 149 global institutional investors signatories, representing over US$1.66 trillion in assets. It called for the adoption of human rights impact assessment requirements for developing and deploying AI systems, public database requirements to ensure meaningful transparency, and appropriate prohibitions and safeguards for high-risk AI systems.

→ **Adding Investor Voice in Public Consultations:** In July 2023, Open Mic, together with NEI Investments, Arjuna Capital, the Seventh Generation Interfaith Coalition for Responsible Investment, Azzad Asset Management, Zevin Asset Management, Heartland Initiative and the Investor Alliance for Human Rights filed a joint public comment recommending that the U.S. National Telecommunications and Information Administration (NTIA) develop critical guardrails on AI development and use for the potential benefits of AI technology to be realized. The comments address the lack of transparency in the AI sector, deficit standards for nascent AI auditing and assessment field, and the absence of federal privacy and liability frameworks to govern company responsibility for the inputs and outputs of their AI models. The Investor Alliance also submitted an independent submission to NTIA in 2023 and inputs on the development of an AI Action Plan in the U.S in March 2025.

ANNEX C: Resources

- 59th HRC Submission on the Use of AI and the UNGPs, ICCR and the Investor Alliance for Human Rights, January 2025
- Human Rights Across the Generative AI Value Chain, BSR, February 2025
- Intro to AI: a beginner's guide to artificial intelligence, MIT Technology Review, October 2024
- Dehumanization, Discrimination and Deskilling: The Impact of Digital Tech on Low-Wage Workers-The Case for Shareholder Engagement, ICCR, September 2024
- Navigating the nexus of AI and human rights: A toolkit for investors in a world of rapid change, Responsible Investment Association Australasia, June 2024
- Advancing Responsible Development and Deployment of Generative AI, U.N. B-Tech Project, November 2023
- Responsible AI and Human Rights: An Overview of Company Practices, U.N. B-Tech Project, November 2023
- Taxonomy of Human Rights Risks Connected to Generative AI,U.N. B-Tech Project, November 2023
- Taking Action to Address Human Rights Risks Related to End-Use, U.N. B-Tech Project, November 2023
- Intro to Generative A.I. Aspen Digital, June 2023
- A.I. 101, Aspen Digital, June 2023
- Artificial Intelligence: A Rights-Based Blueprint for Business, BSR, August 2018
- Preliminary Standards for Generative AI, Ranking Digital Rights
- AI Risk Repository, MIT
- Global AI Policy tracker International Association of Privacy Professionals
- Heightened Human Rights Due Diligence for Business in Conflict-Affected Contexts: A Guide,The UNDP and the UN Working Group on Business and Human Rights

INVESTOR ALLIANCE
FOR **HUMAN RIGHTS**
AN INITIATIVE OF ICCR

The Investor Alliance for Human Rights is a collective action platform for responsible investment that is grounded in respect for people's fundamental rights. We are a membership-based, non-profit initiative focusing on the investor responsibility to respect human rights, corporate engagements that drive responsible business conduct, and standard-setting activities that push for robust business and human rights policies. Our membership is currently comprised of over 240 institutional investors, including asset management firms, trade union funds, public pension funds, foundations, endowments, faith-based organizations, and family funds. Our members currently represent a total of over US$20 trillion in assets under management and 20 countries. The Investor Alliance for Human Rights is an initiative of ICCR. For more information, please visit our website and follow us on LinkedIn and BlueSky.

This briefing was developed by the Investor Alliance for Human Rights, with Aditi Rukhaiyar as the principal author. We would like to thank Jessica Wan at Redwheel, Larisa Ruoff at Loring, Wolcott & Coolidge, Matt Mahmoudi at Amnesty International, and Access Now for their feedback.