February 2025

Shaping the Future of AI Governance: Standards, Risk Management, and Responsible Practices

Key Takeaways from the official side-event to the Paris AI Action Summit 2025



On 11 February 2025, the UN B-Tech Project, the AI Standards Hub, SaferAI and the OECD organized an official side event at the Paris AI Action Summit. This document summarizes key takeaways for two panels in which the UN B-Tech Project was actively involved, which could be fruitful for further discussion by the participating organizations and the field at large. Hence, the discussed themes are not an exhaustive list and are instead a summary of expert views regarding responsible business practices related to the development and use of AI systems as they relate to the advancement and the uptake of the UN Guiding Principles on Business and Human Rights (UNGPs) and the OECD Guidelines for Multinational Enterprises on Responsible Business Conduct (MNE Guidelines).

Panel: "Advancing responsible business conduct in AI: Strengthening policy, investor engagement, and corporate practices"

This panel explored how investors and policymakers can safeguard human rights within AI companies when designing, developing, and deploying AI technologies. Specifically, the panel examined the roles of oversight bodies and investors in promoting responsible AI. The panel pinpointed the crucial role different types of investors play in incentivizing responsible AI development.

1. Human rights should inform technical standards on AI

Stakeholders raised the need for technical standards on AI informed by human rights. As an example, the interplay between legislation and standard-setting in the EU was explained. While laws, such as the EU AI Act, mandate transparency, human oversight, they often do not exactly define what these principles entail in practice and how they can be operationalized. By design, it is European standardisation organisations like CEN and CENELEC that then define the standards that accompany the implementation of the regulatory approaches. Increasingly, these organizations are also going beyond purely technical standards and considering issues such as fundamental rights.

While some have been concerned with introducing this dimension into technical standards, stakeholders emphasized that it is possible to integrate human rights considerations into standard-setting processes. Standardisation processes consist of two steps: Creating a consensus among all relevant stakeholders and then documenting that consensus in a clear and concrete way so that technical people, engineers and programmers can implement it.

In this context, stakeholders mentioned that technical standard-setting processes should ideally orientate themselves towards global standards with regard to codifying responsible business

conduct such as the OECD Guidelines and UNGPs, rather than developing too many parallel instruments and guidelines.

2. Advancing responsible business conduct by reinforcing the different roles and responsibilities of the State and private sector to safeguard human rights in AI

Stakeholders emphasized that AI has risen into public attention extremely fast and could amount to an existential threat to humanity if ungoverned. Driven by private companies and backed by very large investments, the technology is developing exceptionally fast. A participant highlighted how AI affects all sectors and almost everyone and puts a shared responsibility on companies, governments and society.

The private sector has driven AI innovation, supplying products to businesses and governments and wielding substantial influence in shaping global AI governance. The UNGPs constitute a key framework for defining the roles of States and businesses in upholding human rights in the context of AI and emerging technologies. By establishing a level playing field for States and businesses, the UNGPs can mitigate the risks and promote the benefits of AI globally. Accordingly, companies involved in AI development and deployment need to implement measures to safeguard human rights throughout their operations. The UNGPs reaffirm the State's duty to protect human rights, emphasizing the obligation to safeguard against adverse impacts by third parties, including impacts stemming from or linked to corporate activities. That duty extends to the development, deployment and use of AI technologies by companies. The Guiding Principles introduce the concept of a "smart mix" of mandatory and incentive-based measures to require companies to respect human rights. That approach is aimed at achieving policy coherence across national, regional and international frameworks, aligning expectations for corporate conduct.

In this regard, governments should understand more what is happening on the company side but also support projects like B-Tech in applying a human rights lens and discussing these frameworks directly with the industry. The importance of these issues becomes very clear when looking at AI applications in key areas such as health and finance.

3. Investors are increasingly engaging in human rights in AI and welcome global standard-setting

Stakeholders highlighted the need for global standards on the issue, as country by country approaches align less with their global investment operations. These standards can also help companies and investors align on expectations across companies, industries and regions.

A participant mentioned that in general, investors approach AI similarly to other industries when it comes to human rights. They have been engaging with issues such as climate change, human rights, children's rights and biodiversity for many years and can now often apply the same tried and tested frameworks.

As an approach, participating investors mentioned that they first engage with the companies in order to understand how they operate and to build a relationship. In the next step, they work on making clear their views and values in order to impact the companies, particularly if they can identify risks not yet covered by the companies.

4. There is a race between AI development and regulation

Stakeholders raised the point that while AI is being developed rapidly, regulation is lagging behind. The AI space is continuously innovating, bringing a range of changes to be considered for risk management.

In this context it was also raised that regulation is not only falling behind but also slowing recently.

5. Companies are increasingly engaging on human rights in tech, yet efforts remain uneven

Stakeholders presented that out of 200 companies monitored by the World Benchmarking Alliance, only 15% had some form of published guidance on human rights in 2021, now the share has risen to 37% of companies but this low share was seen as very worrying. Participants highlighted the slow improvement as concerning. This sort of guidance should include impact assessments, due diligence as well as risks emerging from downstream activities of customers.

Best practices play an important role, where educating companies about best practices of other firms can drive the industry to performing better in this regard. Promoting responsible AI in this very engaging way and in close dialogue with companies was described as a promising approach.

6. Venture capital can be a powerful lever in encouraging startups to develop AI responsibly

Participants explained that Venture capital (VC) funds have significant leverage over companies and can thereby influence the companies but also their customers, consumers and the broader stakeholder environment. By influencing how these VCs make investment decisions, and how they support their portfolio companies, it is possible to help these companies lay strong foundations, which means that they are then able to better consider human rights in their decision-making. It was also mentioned that while startup founders often do not participate in this kind of event, VCs can be an effective way of reaching these companies. VCs can also be a good way to access companies that are not yet publicly traded and therefore harder to engage with for traditional investors.

In this context it was raised, that currently many good frameworks exist but many are not very actionable and operate at a high level. This can be very challenging for startups as policies are important but risk being buried and having little practical effect. It is important to work with VCs and to develop guidance and frameworks to show them how these issues can be material, how they can be embedded in their investment decision-making processes and how they can promote awareness among their portfolio companies. Case studies of positive action are important, but those highlighting how things went wrong can be very impressive.

7. Coherence in approaching human rights in AI is important

Finally, stakeholders raised the importance of coherence across approaching the issue of respect for human rights in AI development and deployment. This coherence was said to involve two key dimensions: Alignment and coordination.

In order for all involved stakeholders across governments, international organizations, private actors and civil society to be supportive of each other, it is important to be aligned on general directionality and principles. In this regard participants mentioned that human rights law should be the guiding framework. While approaches based on ethics can be helpful, the concept remains debatable and dependent on cultural understandings.

In addition, coordination across all stakeholders is important. Here institutions like the OECD and UN can play a fundamental role in bringing together all parties. Stakeholders emphasized the importance of events like this one in order to learn from each other but also to make sure that work is not duplicated.

Panel: “Inclusive governance and human rights in global AI standardisation”

The panel on the intersection of human rights, technological innovation, and AI governance highlighted the increasing importance of AI standards in achieving policy objectives. These

standards not only enable innovation but also connect with regulation, assurance mechanisms, and certification, often having legal significance. They are particularly crucial in implementing the EU AI Act, as they provide guidance to protect fundamental rights and public interest. The discussion emphasized the need to critically assess the processes of standard-setting, questioning how standards are produced, who contributes, and whether they adequately protect fundamental rights.

1. Human rights must be embedded in global standards

Stakeholders emphasized that AI systems are developed with priorities that may overlook context-specific considerations of AI deployment and use with regard to cultural norms, legal frameworks, and socioeconomic conditions. When not designed inclusively, these systems can cause harm when interacting with diverse social, economic, and political contexts. Including human rights perspectives in AI design is therefore crucial to avoid biased decisions that negatively impact individuals and communities. Participants noted that the creation of AI standards is a political process, shaped by the priorities of those involved, which ultimately influences how technology is used and under what conditions.

Weak standards can normalize unethical practices, such as lacking privacy safeguards in biometric systems, leading to mass surveillance. Embedding human rights in AI standardisation is essential to guide technology design and deployment responsibly. As an example, participants mentioned a clear need for protections in the data value chain, where many workers involved in building AI models lack rights due to the informal nature of their work. It requires understanding this ecosystem to introduce adequate protections. Here lessons can be learned from responsible business practices in other value chains, and existing frameworks can be adapted to address these new challenges. Bridging and integrating responsible AI initiatives with human rights efforts within organizations is the first element to build practical guidance. Collaborative efforts with entities like the OECD and B-Tech can test and refine practices in applied settings, embedding them into broader frameworks for responsible AI. It's important to avoid creating too many competing guidelines, and instead focus on developing practical, embedded guidance within broader frameworks and bridging between different communities to avoid the diffusion of responsibility.

The role of standardisation bodies like the International Telecommunication Union (ITU) was highlighted as increasingly important in this context. The ITU is involved in designing global technical standards, driven by member States, industry, academia and civil society, through expert-led study groups seeking consensus for technical standards. Increasingly it is incorporating human rights into technical standards, responding to the UN Human Rights Council's call for cooperation to ensure human rights protection in digital spaces as a third lens alongside technical accuracy and commercial viability. Guided by the Global Digital Compact, the ITU is developing over 100 AI-related standards focused on safety, reliability, sustainability and human rights, promoting global AI standards collaboration and contributing to projects like AI watermarking and deepfake detection to combat misinformation. The World Telecommunication Standardisation Assembly, held every four years, sets ITU's strategy; recent resolutions highlight ITU's role in AI standardisation and developing metaverse standards with a focus on human rights, marking a pivotal move towards explicit human rights considerations.

2. Establishing legal and enforceable rules beyond “ethics” and based on human rights is crucial

Stakeholders mentioned human rights as crucial in establishing enforceable legal rules, which can aid judicial systems, particularly in regions like Latin America where regulations are evolving, by offering interpretation of human rights in relation to AI and technology. Incorporating human rights into standards provides clarity on the risks posed by systems and aids in legally defending human rights in increasingly complex societies.

While ethics are valuable in guiding AI standards, they may not adequately address the diversity and differing worldviews of traditional communities; integrating human rights offers a clearer and more inclusive approach, as well as a universally agreed global framework. Human rights offer global consensus, providing clarity for assessing harms and determining redress measures across diverse communities.

3. Technical and human rights communities must work together/ Human rights principles should be translated to actionable guidelines

Participants discussed the need for human rights principles to be translated into actionable guidelines to inform technical standard-setting processes and for human rights communities working together with technical experts. This was corroborated by a case study tracking the interactions of researchers, engineers, product managers regarding human rights within a company. Integrating AI guidelines within existing human rights due diligence frameworks requires ongoing collaboration with diverse organizations to effectively implement practical solutions while considering the impact on affected communities. In this context, stakeholders highlighted a need for greater openness and integration across various communities, as current discussions on topics like auditing, safety, and human rights are occurring in separate spaces. More consistent collaboration could enhance understanding and implementation of responsible AI practices.

4. Barriers remain to effective multistakeholder collaboration in technical standard-setting processes

Without diverse and inclusive input, AI standards risk reinforcing inequalities, legitimizing harmful practices, and benefiting dominant stakeholders. Technology and design processes can reflect societal values, priorities, and power structures of the societies that create them. Design decisions in AI, like data selection and model architecture, involve trade-offs with ethical, social, economic, and political implications and must be taken with the realities of all affected stakeholders in mind.

In this regard, civil society organizations face a number of barriers:

1. Financial constraints, such as costly memberships and frequent travel requirements, create significant barriers for civil society organizations from marginalized communities or the global South.

2. A lack of technical expertise in specialized areas like cryptographic security and machine learning makes it challenging for civil society organizations to participate meaningfully in discussions.

3. Opaque and exclusionary processes within standards bodies restrict access to documents and transparency, making it difficult for civil society to engage and shape decisions.

4. Industry and government dominance in standardisation spaces often sidelines human rights considerations, but building coalitions and bridging expertise can help civil society advocate for more inclusive and transparent processes.

In this context, participants mentioned that bridging technical and human rights expertise can help with some issues, empowering civil society to navigate AI standardisation complexities and advocate for structural reforms to make standard-setting bodies more inclusive and transparent. Initiatives and fellowships can help overcome barriers to entry. It was also mentioned that the ITU encourages civil society participation in developing technical standards and is working on

tools like a human rights taxonomy and impact assessment framework to bridge the gap between human rights experts and the technical community.