HRBA@Tech

December 2024

# Towards a Human Rights-Based Approach to New and Emerging Technologies: From Concept to Implementation

*A Series of Papers on*

*A Human Rights-Based Approach to New and Emerging Technologies (HRBA@Tech)*

# Towards a Human Rights-Based Approach to New and Emerging Technologies: From Concept to Implementation

2024



* These papers are the third installation of an annual paper series focusing on a human rights-based approach to new and emerging digital technologies. The first set of papers was published in 2022: *Towards a Human Rights-Based Approach to New and Emerging Technologies: A Framework,* followed by a second batch in 2023: *A Human Rights-Based Approach to AI for Tech Startups and Global Normative Governance*.

# Acknowledgments

This report was jointly produced by the Seoul National University AI Policy Initiative (SAPI) and the Universal Rights Group (URG). Paper 3-1.a was authored by Marc Limon of URG, and paper 3.1.b was authored by Prof. Stephan Sonnenberg with the input of numerous colleagues from URG as well as SAPI. Paper 3-2 was authored by Prof. Stephan Sonnenberg of SAPI, drawing heavily the earlier reports published in this paper series in 2022 and 2023. Prof. Yong Lim served as the Project Manager and supervising editor for all three papers and all accompanying policy engagements and stakeholder consultations.

As in previous years, we would like to thank Ambassador Seong Deok YUN, who serves as the Permanent Representative of the Permanent Mission of the Republic of Korea in Geneva and Secretaries Youngmin KWON and Woohyun KANG, without whose steadfast support and patience this paper series would have never come to pass.

We owe a great debt of gratitude to the Permanent Mission of the Republic of Korea in Geneva for hosting two roundtable discussions in June and October 2024, during which we were able to benefit from the tremendous insights and feedback of the assembled guests representing various United Nations agencies, Geneva-based diplomatic missions, and human rights think tanks. We also would like to thank the Geneva-based staff of the Office of the High Commissioner for Human Rights for generously including us in some of their consultations and for their many valuable inputs into this ongoing paper series.

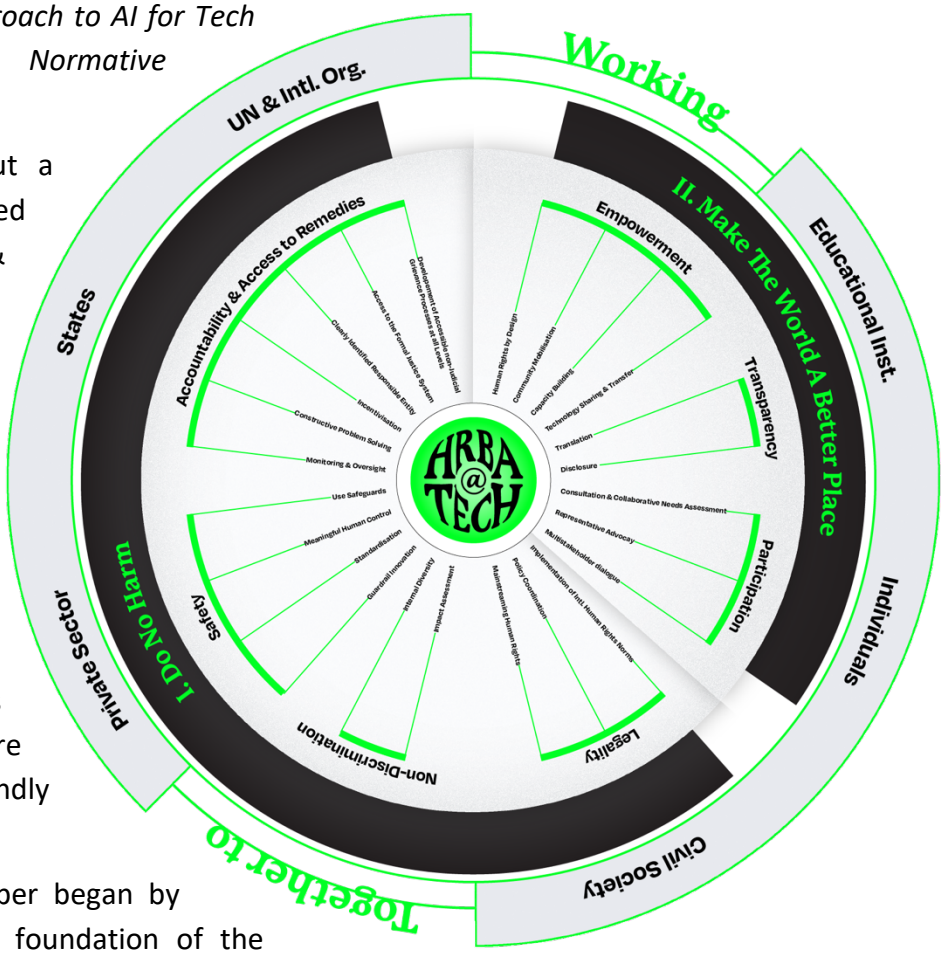
The SAPI team is particularly proud to have been able to involve a group of talented students associated with the University Network for Human Rights in this effort. Lily Ahluwalia, Akram Elkouraichi, Elizabeth Littell, Fiona O'Reilly, Ashleigh Parlman, Eden Mae Richman, Ziyad Rahman, Sage Saling, Jamie Sloves, and Helen Xie from Wesleyan University (Connecticut, USA), as well as Mr. Aly Moose from Yale University (Connecticut, USA) contributed substantially to Paper 3.1.b. Aly Moosa also contributed significantly to the development of the consulting manual described in Paper 3.2, primarily by conceptualizing and testing the accompanying chatbot described in the manual. We would also like to acknowledge the invaluable editorial and substantive inputs of Prof. Buhmsuk Baek and URG's Louis Mason and to this 2024 paper series.

SAPI and URG would like to thank our gracious interviewees and partners, all of whom spent hours speaking to us about their work in this field, often inviting us into sensitive and confidential environments where we were able to move beyond the usual platitudes and learn about the reality of doing this important work. Their inputs and feedback make this project unique. Particular mention this year goes to Woochul Park and Jiwon Jung from NAVER Agenda Research, Jinhwa Ha and Albert Noon from Kakao, Myoungshin Kim from LG AI Research, the participants of the Microsoft Research Asia (MSR) TAB Workshop in Beijing on November 14, 2024, and the many gracious diplomats and specialists in Geneva and Seoul who shared with

us their insights, expertise, and institutional know-how. We owe a great deal of gratitude to each of our interviewees and their senior managers for allowing us to learn from you.

# Introduction

This installation of the HRBA@Tech paper series builds on the foundational vision document released in December 2022 (*Towards a Human Rights-Based Approach to New and Emerging Technologies: A Framework*) as well as the follow-on series of papers released in 2023 (*A Human Rights-Based Approach to AI for Tech Startups and Global Normative Governance*).




The 2022 paper laid out a 'Human Rights Based Approach to New & Emerging Technologies' (HRBA@Tech) built around the recognition that technology is not inherently 'neutral,' and that various stakeholders can and should work together to 'nudge' new and emerging technologies (NETs) in ways that ensure more human-rights friendly social outcomes.

The 2022 Framework Paper began by proposing the normative foundation of the HRBA@Tech model consisting of seven fundamental principles that collectively govern how various stakeholders should engage with NETs. Four of those principles fall under the broad rubric of "do no harm," and broadly describe principles anchored in efforts to prevent violations of human rights. These are (1) Legality, (2) Non-Discrimination & Equality, (3) Safety, and (4) Accountability & Access to a Remedy. The three remaining principles pertain to efforts to "make the world a better place" and focus on (5) Empowerment, (6) Transparency, and (7) Participation. That latter category of principles focuses as much on guaranteeing respect for human *rights* as it also invokes our collective *responsibility* to act as ethical and moral actors when we usher new technologies into existence.

The 2022 report went beyond the mere articulation of guiding normative principles. It also explored three subordinate aspects of what it takes to make those principles real. The first has to with "The How" – identifying 24 concrete processes by which various stakeholders –

sometimes working alone but more frequently acting in concert with others – collectively 'nudge' NETs in the direction of human rights. These 24 processes, the paper contended, are more tangible than the aforementioned seven principles, and thereby offer decision makers more practical guidance as they work to bring human rights considerations into the design and deployment of NETs. The focus on concrete processes distinguishes the HRBA@Tech model from the majority of other ethics or rights-based frameworks.

Second, the HRBA@Tech model distilled the full universe of actors in this arena into six core stakeholder categories ("the Who"). These six stakeholder categories – sometimes by means of collaborative interactions and other-times by more adversarial means – collectively interact to 'nudge' NETs in the direction of human rights. This model expands on the traditional model of human rights which really unfolds only within the relationship between a State and an individual, and even the more recent focus on businesses as relevant actors in the area of human rights. By adding also a description of the role that international organizations, educational institutions, and civil society actors can and often do play in efforts to influence the way NETs interact with society this model more appropriately addresses the important role differentiation within any such governance effort. Finally, the HRBA@Tech model began to project the 24 individual processes onto a generic technology lifecycle (TLC), illustrating how some processes are more impactful at certain points along the TLC than others ("The When"), and thereby allowing decision makers to also think about efficiency.

In 2023, SAPI and URG produced a second installment of the paper series that accomplished two tasks. The first was to 'translate' the 2022 HRBA@Tech model into the world of Artificial Intelligence (AI), and more specifically the world of startups trying to develop and deploy AI or AI-enabled technologies. The 2022 HRBA@Tech model was designed to apply to *any* NET – not just digital NETs and certainly not only AI. Given the global focus on efforts to produce "safe" and "trustworthy" AI, however, the authors of the second paper series compared what the HRBA@Tech model might *suggest* those tech entrepreneurs should be doing with the descriptions of what they were *actually* doing in real life. The first part of that paper series focused on the efforts of small and medium enterprises (AI Startups) to put in place measures designed to prevent AI from violating human rights. These efforts were then compared with the principles and processes listed under the "do-no-harm" heading of the HRBA@Tech model. IN a second paper, researchers examined what tech startups were doing to deploy AI and AI-based solutions in the fight against climate change, comparing those findings with the "make the world a better place" principles and associated processes. Finally, the 2023 paper explored the contours of a proposal to establish a new Special Mechanism under the HRC.

At the outset of 2024, the authors already knew that more work was necessary to elaborate upon what reforms, if any, might be necessary to optimally situate the Human Rights Council and other Geneva-based human rights institutions to engage in efforts to govern NETs. Furthermore, the authors determined to 'translate' the content of the 2022 HRBA@Tech model into a concrete training and management consulting agenda, making it easier to

imagine how one might use the model to drive concrete institutional change. 2024 marked a new high-point in discussions at the UN and elsewhere over how to regulate AI, and a consensus seemed to be emerging that human rights norms and institutions should serve as a crucial "anchor" to any efforts to deploy AI and other NETs to solve humanity's knottiest problems.

### Paper 3-1.a: Establishment of a Human Rights Council Working Group on new and emerging digital technologies and human rights: Background Paper

Principal Author: Universal Rights Group (URG) - Geneva

Focus: Following years of informal consultations between States, civil society actors, technology companies (including multinationals and SMEs), academia, and UN experts, this background paper makes the case for the Human Rights Council to establish a mechanism for new and emerging digital technologies and human rights.

### Paper 3-1-b: The historical and disciplinary backdrop to current efforts to enhance the UN Human Rights Council's capacity to promote a human rights-based approach to AI and other New and Emerging Digital Technologies

Principal Author: Seoul National University AI Policy Initiative

Focus: This paper serves as a supplement to paper 3-1.a, providing an overview of how various agencies and institutions in the UN family have reacted, consistently with their respective mandates, to the advent of AI and other NETs. It traces the emerging consensus that human rights must serve as the anchor to ground any effort to deploy NETs in service of sustainable development, global equity, and national security, and positions the Human Rights Council and the Office of the High Commissioner for Human Rights (OHCHR) as central players in that anchoring process.

### Paper 3-2: The HRBA@Tech diagnostic tool and management consultant's toolbox: an introduction to the approach

Principal Author: Seoul National University AI Policy Initiative

Focus: As an illustration of how a normative framework like the 2022 HRBA@Tech model can translate into concrete capacity building

efforts based on that normative guidance, this diagnostic tool and management consultant's toolbox outlines a methodology for a third party facilitator to work with a range of different organizational clients to design a bespoke human rights-based approach to NETs.

# Discussion paper 3-1.a

# Establishment of a Human Rights Council Working Group on New and Emerging Digital Technologies and Human Rights

This paper was authored independently by the Universal Rights Group (URG) as part of an ongoing collaboration between URG, the Seoul National University Artificial Intelligence Policy Initiative (SAPI), and the Permanent Mission of the Republic of Korea to Geneva. A preliminary draft of this concept note was presented for discussion on October 4, 2024 at Policy Dialogue organized at the Permanent Mission of the Republic of Korea in Geneva and subsequently updated based on feedback and inputs from various sources.

Establishment of a Human Rights Council Working Group on new and emerging digital technologies and human rights

Background paper

The advent of new digital technologies, such as artificial intelligence (AI), is ushering in transformative changes across multiple sectors. These changes have the potential to accelerate the achievement of the Sustainable Development Goals (SDGs), and progress towards the realisation of human rights around the world. However, these same technologies also pose novel risks, including to the human rights agenda, a point acknowledged in the new Global Digital Compact, which underscores 'the need to identify and mitigate risks and to ensure human oversight of technology in ways that advance sustainable development and the full enjoyment of human rights.'[1] Often referenced in this regard are the risks posed by artificial intelligence (AI) to the rights to non-discrimination, privacy, and work, the risk that it may further widen the digital divide, as well as the potential misuse of such technologies to spread disinformation and hate speech, or to undermine elections. In all such discussions, a robust human rights lens is necessary to ensure no one is left behind by these potentially unprecedented technological upheavals.

In other words, digital technology has the potential to provide a significant boost to the enjoyment of human rights, and to sustainable development, yet also carries significant risks, and international human rights law offers a unique framework to help States, and technology companies, navigate this terrain and help ensure that digital technology is placed at the service of human rights.

This point was, again, recognised in the Global Digital Compact (GDC), Objective 3 of which pledges States to 'foster an inclusive, open, safe and secure digital space that respects,

[1] Global Digital Compact, A/79/L2, 22 September 2024, paragraph 3

protects and promotes human rights.'[2]

Under the heading 'Human rights,' States commit 'to respect, protect and promote human rights in the digital space. We will uphold international human rights law throughout the life cycle of digital and emerging technologies so that users can safely benefit from digital technologies and are protected from violations, abuses, and all forms of discrimination.' To realise that commitment, States inter alia pledged to ensure that 'the development and implementation of national legislation relevant to digital technologies is compliant with obligations under international law, including international human rights law,'[3] and also recognised 'the responsibilities of all stakeholders in this endeavour [including] the private sector.'[4]

But how to realise these important objectives and commitments in practice?

In June 2024, pursuant to Human Rights Council resolution 53/29, the Office of the High Commissioner for Human Rights (OHCHR) presented a 'Mapping report: human rights and new and emerging digital technologies'[5] to the Council's 56th session. The report aimed to map the existing work of the Council, its mechanisms, and OHCHR, in the area of human rights and digital technologies, identify gaps in the scope and effectiveness of that work, and make recommendations to enhance the impact of the human rights system in the future – both in terms of addressing the challenges and seizing the opportunities presented by digital technology.

The mapping report revealed that the Council, its mechanisms, and OHCHR have been 'immensely productive in responding to the manifold challenges [and opportunities] of the ongoing digitization of societies.'

[2] Global Digital Compact, A/79/L2, 22 September 2024, paragraph 22
[3] Global Digital Compact, A/79/L2, 22 September 2024, paragraph 23
[4] Global Digital Compact, A/79/L2, 22 September 2024, paragraph 22
[5] Mapping report: human rights and new and emerging digital technologies, Report of the Office of the United Nations High Commissioner for Human Rights, A/HRC/56/45, 24 June 2024

For its part, the Council, from its 41st session onwards, has adopted a series of landmark resolutions on new and emerging digital technologies (resolutions 41/11, 47/23, 53/29), which have sought to clarify the normative relationship between human rights and digital technologies, identify the human rights challenges and opportunities presented by such technologies, begin to consider what a human rights-based approach to the conception, design, and roll-out of digital technology might look like, and sought to consider the human rights dimension of specific areas of digital technology, especially AI. The Council has also held panel debates on digital technology and human rights, and myriad informal meetings have been convened on the side-lines of the Council.[6] For example, in November 2021, the Universal Rights Group (URG), Facebook, Norway, Switzerland, and others organised a meeting in Montreux, Switzerland, on how to 'Place digital technology at the service of human rights;'[7] in December 2022 and 2023, the Republic of Korea, the Seoul National University AI Policy Initiative (SAPI), and URG held Council side events to mark the publication of policy reports proposing a universal rights-based approach to digital technologies and human rights, and then seeking to apply that approach to certain sectors (e.g., AI);[8] and in May 2023, the Republic of Korea and URG organised the ninth edition of the Glion Human Rights Dialogue to consider ways to strengthen the work of the Council and its mechanisms to address the challenges, and seize the opportunities presented by digital technology.[9]

The Special Procedures mechanism has also been extremely active in exploring the relationship between human rights and digital technology. For example, the mapping report found that, to date, at least 135 reports by Special Procedures mandate-holders have considered aspects of digitalisation. These reports have provided 'nuanced analysis on topics ranging from internet access to surveillance, online information controls, hate speech, racism embedded in technology, health, worker's protections in the gig economy, education, and the alleviation of poverty, among others.' Human rights questions around

[6] Although these informal meetings were not covered in the mapping report.
[7] https://www.universal-rights.org/urg-policy-reports/placing-digital-technology-and-the-service-of-democracy-and-human-rights-3d2-2/
[8] https://www.universal-rights.org/urg-policy-reports/new-and-emerging-technologies/
[9] https://www.universal-rights.org/urg-policy-reports/glion-ix-placing-new-and-emerging-technologies-at-the-service-of-human-rights-and-democracy/

digital technology have also been raised many times in the UPR, 'although to date with a limited scope and depth.'

Finally, the mapping report highlighted the scope and depth of the work of the Office of the High Commissioner for Human Right's work on digital technologies. This has covered a growing range of topics, from the gender digital divide to data privacy, surveillance, end-to-end encryption, internet shutdowns, the role of technology in the context of peaceful assemblies, technical standards, governance of content on internet platforms, and border governance. The extent of this engagement is also recognised in the newly adopted Global Digital Compact, which acknowledges 'the Office of the United Nations High Commissioner for Human Rights' ongoing efforts to provide, through an advisory service on human rights in the digital space, upon request and within existing mandate and with voluntary resources, expert advice and practical guidance on human rights and technology issues.'[10]

The mapping report underscores the importance and value of these interventions by the UN human rights ecosystem, and argues that they provide clear evidence of 'the relevance and necessity of using the international human rights framework to govern the development and use of digital technologies.' As recognised in the GDC, international human rights law provides the guardrails required to maximise the benefits and added value of digital technologies, while reducing and containing their potential detrimental human rights impacts.

Notwithstanding this positive picture, the mapping report also identifies some important challenges. For example, the report concedes that 'the large number of actors and processes in the human rights system addressing issues concerning digital technologies can lead to overlap and at times tensions between outcomes.' The mapping report references the large number of resolutions and Special Procedures reports on AI as an example. Importantly, the report highlights the problems and risks posed by this diffused approach to digital technologies at the Council, such as 'duplication of effort, thinning [of] already

[10] Global Digital Compact, A/79/L2, 22 September 2024, paragraph 24

sparse resources, and degrees of ambiguity or lack of clarity, with separate initiatives approaching related issues in different ways.' 'Given the engagement of an array of actors on a broad range of topics,' the report continues, 'the potential for inconsistencies and contradiction exists, particularly if new outputs do not adequately consider existing work.'

This normative challenge is extremely important. In order to be readily taken up by governments and technology companies (which must necessarily be the ultimate objective of the Council's work in this area), human rights normative guidance or a 'rights-based approach' to digital technologies must be clear, universal, unambiguous, and easily accessible. It is simply not possible, at present, for States, especially developing countries, or for technology companies, especially SMEs, to track, understand, and reconcile the diffuse normative work of the above-mentioned diverse set of actors, especially when those outputs may be duplicative or incoherent - and even, in some cases, contradictory.

A further challenge identified in the mapping report, is to translate any universal normative guidance framework, developed at the Council, into real-world change. 'More needs to be done,' the report says, 'to ensure that the recommendations from the human rights system are implemented by decision-makers 'on the ground.'' At one level, as noted above, this means providing clear and accessible 'guidance for the implementation of human rights obligations and responsibilities of States and businesses in the context of digital technologies.' But it also means providing counsel, technical assistance, and capacity-building support to States (especially developing States, and in particular to LDCs and SIDS), and to technology companies (especially SMEs and start-ups), so that human rights obligations, commitments, and principles can inform the regulation, conception, design, development, and operation of new and emerging digital technologies.

Linked with this point, the international human rights system must do far more to ensure that a human rights-based approach is leveraged to help bridge (and that new and emerging digital technologies such as AI do not serve to further widen) the digital divide between developed and developing countries. In a similar vein, the Council and its mechanisms

should further explore and promote the opportunities presented by digital technologies to empower individuals to fully enjoy their human rights, for example by improving their digital, media, and information literacy, including AI literacy.

A further important challenge (and opportunity) is the need to break down silos with the related work of other parts of the UN, 'for example those dealing with trade and e-commerce, intellectual property, technical standard-setting, and peace and security.' It is of critical importance for the Council and/or its mechanisms, to use their convening power to bring these actors together, both to inform and contribute to the development of any universal human rights normative guidance framework on digital technology, and to ensure that a human rights-based approach is integrated into their work and decision-making. Ways should therefore be found to institutionalise multistakeholder dialogue and collaboration.

With the foregoing in mind, the question becomes: how can the international human rights system (the Council, its - existing or future – mechanisms, and OHCHR) best meet these challenges and seize these opportunities?

## What next?

In answer to this question, OHCHR's mapping report comes out in favour of a new service, 'established by OHCHR,' and which would 'provide expert advice on human rights and technology issues to support member States and stakeholders in integrating human rights into the design, development, operation, use and regulation of digital technologies, as suggested by the Secretary-General.' As noted above, the importance of this 'advisory service on human rights in the digital space' has also now been acknowledged in the Global Digital Compact.

This service is indeed extremely welcome and will play a key role in helping States and other actors (e.g., the private sector) ensure that human rights are respected throughout the life cycle of digital and emerging technologies. All possible support should be extended to

OHCHR in that regard.

However, as the Council, like the wider UN, is an intergovernmental body/organisation, and because States are the duty-bearers of the international system (and thus must take a lead role in the development of, and consequently feel ownership of, international human rights policy as it relates to digital technologies), the Council and its mechanisms must necessarily continue to play a critical role. This last point is extremely important. The Council and its mechanisms, and OHCHR, play complementary and mutually reinforcing roles in the UN human rights system, and this understanding should be central to any initiative on digital technology and human rights.

For example, the Council and its (new or existing) mechanisms (e.g., Special Procedures) play a unique role in clarifying and setting human rights norms as they relate to new or emerging challenges, and in encouraging States (as well as, where appropriate, private actors such as businesses) to apply those norms at national level. Yet many States, especially developing States, will require technical assistance and capacity-building support to realise such a rights-based approach to the development and regulation of new and emerging digital technologies. OHCHR, as the primary UN agency/programme responsible for human rights must necessarily play a central role in that regard.

Linked with the foregoing, the Council (perhaps through a new mechanism) must necessarily play an important policy coordination role across the UN human rights system. As noted above and recognised in OHCHR's mapping report, different parts of that system have been extremely active in exploring the relationship between different human rights and digital technologies, including AI, and in clarifying human rights norms as they relate to those technologies. While welcome, this has created a complex web of resolutions, reports, principles, general comments, concluding observations, etc. There is therefore a need to collate and distil this rich body of work into some form of guidance framework that is simple, comprehensive, and – importantly – accessible to both States (including LDCs and SIDS), and technology companies (including SMEs). It is important to stress that this would not be

a new norm-shaping exercise, but rather a distillation and clarification of existing normative work.

With the foregoing in mind, and following years of informal consultations between States, civil society actors, technology companies (including multinationals and SMEs), academia, and UN experts, it is our considered view that the Human Rights Council should establish a mechanism for new and emerging digital technologies and human rights. While the exact form of such a mechanism is a matter for further discussion, a number of considerations remain valid:

- It should be composed of independent experts.
- It should be cross-regional and multistakeholder, with the participation of individuals with, for example, government experience, academic/civil society experience, and private sector experience (especially in the digital technology sector). This multistakeholder approach is vital to help ensure the relevance of normative guidance to technology companies, and to help provide tailored guidance on the take-up of those norms.
- Its core mandate should be to collate, analyse, and distil the existing work of the UN human rights system into some form of voluntary guidance framework that is accessible to both States (including LDCs and SIDS), and technology companies (including SMEs).
- This mandate should be pursued in consultation with all relevant stakeholders, including States, other relevant UN entities, civil society, and the private sector.
- The mechanism should work in close cooperation with OHCHR – especially important considering the latter's mandate to provide 'advisory service on human rights in the digital space' to both States and technology companies, in other words, to practically support the implementation of any universal normative guidance framework in the field.

While, as discussed, the exact form of a new mechanism is a matter for further reflection,

there are a number of options and precedents. These include: a grouping of existing thematic Special Procedures mandates; a group of multi-stakeholder experts (based on the precedent provided by Council resolution 38/18 on the contribution of the Council to the Prevention of Human Rights Violations); or a new normative-focused Special Procedures mandate (i.e., a thematic Independent Expert mandate, or a cross-regional and multistakeholder Working Group[11]).

Whatever its final form and mandate, such a mechanism would be best placed to respond to the opportunities and challenges identified in OHCHR's mapping report. It would provide a single, central focal point and repository of information and normative guidance, thereby addressing the aforementioned problems of diffusion, confusion, duplication, and incoherence. This is crucial if the UN is to provide human rights normative guidance to governments and technology companies that is informed and backed by States (i.e., via a Council mandated mechanism), and that is universal, coherent, and easily comprehensible, accessible, and relevant to governments (including in developing countries) and technology companies (including SMEs), so that it can be taken up and readily applied in the regulation, design, development, and operation of digital technologies. As noted above, this norm-clarifying role would also help complement, provide a normative basis for, and support the delivery of technical assistance by OHCHR, other UN agencies and programmes (e.g., ITU), as well as UN Country Teams, through the provision of coherent interpretations of existing human rights law and standards.

Such a mechanism would also be well-placed to work in cooperation with OHCHR to support governments and technology companies to take up, implement, and apply a human rights-based approach to the regulation, design, development, and operation of digital technologies, through expert advice, technical assistance, and capacity-building support. Such work would emphasise cooperation and collaboration, and have the goal of helping national stakeholders apply the aforementioned universal normative guidance framework,

[11] There is no existing precedent for such a normative-focused thematic Special Procedures Working Group. Working Groups normally also undertake implementation and compliance work, in addition to norm-setting.

so that digital technologies are placed at the service of, and do not harm, human rights and human dignity.

A particular focus of such capacity-building and technical assistance should be to help developing country governments regulate digital technologies, and technology companies design and roll-out digital technology products, in a manner that promotes and does not harm human rights, borrowing from emerging international best practice, and applying universal human rights norms. Such technical assistance might also include, for example, helping developing country governments develop policies, with international support, to enhance digital literacy. This steps would form an important contribution to bridging the digital divide.

Such a mechanism would also play a valuable role in reaching out to other relevant parts of the UN system, especially those working on the governance of digital technologies, including AI, and in convening governments, technology companies, civil society, OHCHR, ITU, and other UN agencies, to consider important human rights challenges and opportunities presented by digital technologies.

### Efficiency questions at the Human Rights Council

There has long been concern at the Council that there are too many mechanisms, especially too many thematic Special Procedures mandates. Even though, at the time of the Council's establishment in 2006, the new body was asked by the General Assembly to review, rationalise, and improve all mandates inherited from the former Commission on Human Rights, since that time, member States have regularly added new Special Procedures mandates. As of the end of 2023, there were 46 thematic mandates.

Yet while concerns about the expansion of Council mechanisms are understandable, it has never been the intent of repeated efficiency and rationalisation drives at the Council to prevent the body from addressing new and important emerging human rights concerns,

including through the establishment of new Special Procedures mandates. Indeed, one of the Council's greatest strengths and achievements over the past eighteen years has been its ability to respond quickly to emerging global challenges (e.g., climate change, environmental harm, health pandemics, poverty, the effects of foreign debt). The core question of rationalisation, rather, is the fact that since the first thematic Special Procedures mandate was established in 1980, no mandate has ever been discontinued (though two were merged at the turn of the century), even though some no longer serve a useful purpose. It is important, in order for the Council to remain relevant and responsive, that States and other stakeholders focus on this issue, rather than argue against the establishment of new mechanisms or mandates to help the international community address demonstrably critical new human rights concerns (such as digital technology/AI).

A further consideration in this regard is that it is arguably *inefficient* for the Council to continue to address an issue as important and as complex as the relationship between digital technologies and human rights via myriad different initiatives, resolutions, and mechanisms.

As already noted in this paper, such an approach has created a situation marked by diffusion, confusion, duplication, and incoherence. Following this logic, it would be more efficient for the Council to establish a single, central focal point, and repository of information and normative guidance – especially a Working Group that would bring together individuals from different backgrounds and with different areas of particular expertise. A Working Group would also allow the Council to pursue a more coherent approach to the challenges and opportunities posed by digital technologies, by coordinating and cooperating with existing Special Procedures mandates relevant to the issue, as well as with interested States, OHCHR, and other important UN agencies and programmes (e.g., ITU, UNESCO).

## Discussion paper 3-1.b

# The historical and disciplinary backdrop to current efforts to enhance the UN Human Rights Council's capacity to promote a human rights-based approach to AI and other New and Emerging Digital Technologies

This paper was authored independently by Seoul National University Artificial Intelligence Policy Initiative (SAPI) as part of an ongoing collaboration between SAPI, the Universal Rights Group (URG), and the Permanent Mission of the Republic of Korea to Geneva. It was presented for discussion on June 18, 2023 at a Side Event to the 56th Session of the Human Rights Council and subsequently updated based on feedback and inputs from various sources.

# Table of Contents

# Executive Summary

This paper provides a contextual backdrop to the policy discussion and recommendations which constitute the first part of this ongoing paper series. An initial draft of this paper was presented during a Side Event to the 56th Session of the Human Rights Council, which was hosted jointly by the diplomatic missions to the United Nations in Geneva of the Republic of Korea, Gambia, Austria, Denmark, the Grand Duchy of Luxembourg, and the Kingdom of Morocco. It is intended as a primer for readers who may be less familiar with the history of efforts at the international level to engage with new and emerging technologies (NETs), as well as a supplement to the policy recommendations that immediately precede this paper (Paper 3-1.a).

This paper was authored by the Seoul National University Artificial Intelligence Policy Initiative (SAPI), with the active contribution of the Universal Rights Group (URG) as well as a variety of research assistants from Wesleyan and Yale Universities in the United States. These individual contributions are noted with deep appreciation in the acknowledgements below.

SAPI and URG in 2022 proposed a Human Rights-Based Approach to New and Emerging Technologies (HRBA@Tech) Model, and in 2023 applied that approach to small and medium-sized enterprises (SMEs) working to develop AI technologies. In those previous papers, the authors hinted at reforms that would make the Human Rights Council (HRC) a more central player in any emerging efforts to govern the development and deployment of AI at the global level. Those discussions took on much greater urgency in 2024, in large part due to the developments detailed in this paper. Readers unfamiliar with that progression should take advantage of the conceptual framework presented in this paper to contextualize those discussions as well as the recommendations put forward by our colleagues at URG in their contribution to this year's installment of the paper series.

The paper begins with the premise that organizations like the United Nations can no longer pursue centralized command-and-control governance models. This is especially true in the context of AI, where it is apparent that numerous stakeholders – not all of them beholden to traditional international legal constraints – are today playing a central role in the development and deployment of technologies that simultaneously pose significant and at times novel threats to certain human rights protections while also potentially promising to help advance the human rights agenda in other—equally seismic—ways. The question of how to harness those considerable upsides while simultaneously guarding against the potential downsides of AI and other NETs is the central challenge facing policy makers focusing on AI

at the United Nations today, and the central question animating the development of the HRBA@Tech model that still lies at the heart of this paper series.

The paper posits a fundamental choice between 5 different, and non-exclusive, "disciplinary anchors" – loosely defined philosophical or disciplinary reference points for any effort to govern AI and NETs.

| | | |
|---|---|---|
| (1) | The **human rights lens** (see below, p.39) | Does a given technology advance or undermine **human rights protections**? |
| (2) | The **sustainable development lens** (see below, p.51) | Does a given technology contribute to the achievement of the **Sustainable Development Goals (SDGs)**? |
| (3) | The **south-south and triangular development lens** (see below, p.61) | What can be done to harness the power of AI to **"decolonize development"** and promote the interests of nations in the "Global South"? |
| (4) | The **technical lens** (see below, p.65) | Is it possible to reach a global consensus on **technical standards** to govern AI and other NETs? |
| (5) | The **national security lens** (see below, p.67) | Does a given technology potentially advance or undermine **national security**? |

The paper surveys the efforts of several key UN Agencies and Programmes to contribute (through their respective mandates) to the governance of AI, roughly categorizing them according to these five "disciplinary anchors." This is by no means an exhaustive survey,[12] but ideally it serves the purpose of highlighting the key schools of thought that simultaneously influence the way AI is discussed within policy circles at the United Nations.

[12] *See e.g.,* Chief Executives Board for Coordination | High-Level Committee on Programmes (HLCP) | Inter-Agency Working Group on Artificial Intelligence (IAWG-AI), United Nations System White Paper on AI Governance: An analysis of the UN system's institutional models, functions, and existing international normative frameworks applicable to AI governance (advance unedited version), May 2, 2024, *available at* https://unsceb.org/united-nations-system-white-paper-ai-governance.]

Surveying four recent pronouncements on AI governance, specifically, an Interim Report on Governing AI for Humanity published by the High-Level Advisory Body on AI (Dec. 2023), the UN General Assembly's resolution on AI (March 21, 2024), the Inter-Agency Working Group on AI White Paper on AI Governance (May 2, 2024), and the Global Digital Compact (September 22, 2024), the analysis points to an emerging consensus at the highest level of the United Nations, namely that:

1. AI needs to be regulated at the global level;
2. The United Nations is the proper venue for such international governance efforts;
3. AI governance efforts need to be centrally tethered to human rights as the ultimate disciplinary "lens;"
4. AI governance efforts cannot focus only on the downsides of AI, but must also promote the harnessing of AI's potential to amplify social well-being; and
5. Any process to develop a governance model needs to be intensely participatory and consensus-oriented.

The analysis concludes that the UN HRC is optimally situated to play a key role in this governance model. Strengthening the capacity of the HRC to speak to the human rights impacts of AI and other NETs would not come at the expense of any other institutional players already engaged on the topic of AI. Quite to the contrary, the HRC can and should serve a facilitating role in these various discussions, gently ensuring that these various positive efforts remain coordinated and shielded from the counterproductive forces of institutional capture, budgetary competition, and siloed disciplinary thinking. The analysis also highlights the crucial synergies between the OHCHR and the HRC.

# Overview and Background

**This research paper argues for the United Nations Human Rights Council to expand its capacity to progressively develop and subsequently promote a human rights-based approach to new and emerging digital technologies.**

The paper contributes to ongoing efforts to promote a focus within the United Nations system on the threats as well as the opportunities inherent in the development and proliferation of new and emerging technologies (NETs). Since 2022, the Republic of Korea's Permanent Missions in Geneva (*hereinafter* ROK Geneva Mission) has funded an annual paper series bringing together academic and civil society experts to iteratively develop a "human-rights-based approach to new and emerging technologies." This paper is part of that ongoing initiative and draws heavily on arguments first described in greater detail in the previous

working papers.[13,14] It also draws on research conducted in collaboration with the University Network on Human Rights at Wesleyan and Yale Universities in Connecticut, USA during the first half of 2024.[15]

## New and Emerging Digital Technologies in an Era of Polycentric Global Governance

Most people engaged in the discussion over how to regulate AI and other NETs at the global level neglect to ask and answer an even more fundamental question, which is why we ought to be doing so in the first place. After all, sovereign nation-states—armed with their authority to enforce binding legal and regulatory regimes—are, in many ways, optimally placed to regulate private conduct, not to mention more directly accountable to local cultural and social preferences than would be the case for international organizations. At the same time, universal norms—principal among them the human rights framework—tend to lose their coherence without a strong global voice to ensure their universal application across myriad national, social, and cultural contexts.

Considering this tension, how should the international community approach the challenge of harnessing AI's positive potential while also guarding against its use in ways that jeopardize existing human rights protections? Should the international community push for a new

[13] Perm. Mission of the Republic of Korea to Geneva, SNU AI Policy Initiative, and Universal Rights Group. (2022) Towards a Human Rights-Based Approach to New and Emerging Technologies.
[14] Perm. Mission of the Republic of Korea to Geneva, SNU AI Policy Initiative, and Universal Rights Group. (2023) HRBA@Tech: AI Tech Startups, Climate Change, and Global Normative Governance.
[15] This paper was authored by Stephan Sonnenberg, Associate Professor at Seoul National University in South Korea. The author extends his sincere gratitude to the many thoughtful diplomats in Geneva who have generously shared with us their thoughtful inputs and comments. Of particular note are H.E. Ambassador Seong Deok Yun, Secretary Youngmin Kwon, and Secretary Woohyun Kang, all from the Permanent Mission of the Republic of Korea to the United Nations Office and other international organizations in Geneva. The author would also like to express his sincere gratitude to the representatives of the governments of Austria, Canada, the Czech Republic, Denmark, France, Gambia, Luxembourg, Malaysia, Malta, Morocco, Pakistan, The Permanent Observer Mission of the State of Palestine, Singapore, Switzerland, Ukraine, and the United States for their generous and deeply insightful comments on earlier iterations of this paper. The author also wishes to recognize the invaluable contributions of the Office of the High Commissioner for Human Rights (OHCHR), in particular Ms. Peggy Hicks, Ms. Isabel Laura Ebert, Mr. Tim Engelhardt, and Mr. Scott Campbell – all of whom contributed significantly to the insights in this paper. The author would like to thank Lily Ahluwalia, Akram Elkouraichi, Elizabeth Littell, Fiona O'Reilly, Ashleigh Parlman, Eden Mae Richman, Ziyad Rahman, Sage Saling, Jamie Sloves, and Helen Xie from Wesleyan University in Middletown, Connecticut (USA) for their excellent research contributions to this paper, as well as Aly Moosa from Yale University in New Haven, Connecticut (USA), for his research contributions, as well as his excellent support commenting on this draft. Finally, the author would like to express his deep appreciation and admiration for Louis Mason from the Universal Rights Group for his invaluable editorial and substantive inputs, as well as his leadership in earlier installments of this paper series. The arguments set forth in this paper are the result of independent research by the author and SNU's AI Policy Initiative, and do not represent or reflect the views of the Perm. Mission of the Republic of Korea to Geneva or any of the other generous contributors to this effort.

international treaty on AI and human rights? Should it instead rely on voluntary codes of conduct or non-binding declarations? Should new institutions be built resembling those that have already been created in the past? Or should there be an altogether different model of global governance – one that re-evaluates traditional models in light of new theories of governance in today's multipolar, multi-stakeholder, and multi-modal world?

The late John Ruggie, who served as the Special Rapporteur on Business and Human Rights from 2005 to 2011, reflected on this subject in 2014. Ruggie defined global governance as "an instance of govern*ance* in the absence of govern*ment* "[16] (emphasis in the original). Ruggie defined his approach to global governance to contrast with an "old governance model", whereby states negotiated comprehensive binding treaties to harmonize universal policies and norms. Pointing to the "already weak" [… and] "increasingly unattainable" ideal of such an integrated global governance regime,[17] Ruggie claimed that continued adherence to the ideals and principles of this old model held only "limited utility."[18] An alternative approach, which he noted was still controversial among (some) traditionalist human rights groups and scholars, was premised on the observable reality that States today are unable to solve many "pressing societal challenges" on their own, and that other governance approaches must be harnessed to supplement the capabilities of this "old" state-centric approach to governance.[19]

This new governance theory, Ruggie claimed, relied on a **polycentric model of governance,** both in substance *and* in process. The UNGPs that came together as a result of Ruggie's tireless efforts focus on state-centric governance models, corporate obligations to conduct due diligence, and the integrity of grievance processes designed to empower individuals affected by corporate rights abuse. Not just substantively, but also *procedurally*, Ruggie noted that the process of concretizing the UNGPs "was informed by practical engagement with participants from [ ] various stakeholder groups."[20] "The [UNGPs] do not merely advocate a theory of polycentric governance," Ruggie writes, "in part, they were *produced* through such means (emphasis added)." This procedural embrace of polycentric governance, which Ruggie considered to be equally as important as the actual substantive merit of the principles, led to what Ruggie termed to be a "**thick stakeholder consensus**." Such a consensus, Ruggie

[16] John G. Ruggie, 'New Governance Theory': Lessons from Business and Human Rights. Global Governance 20 (2014), 5–17, 5.
[17] Id., *at* 5-6.
[18] Id., *at* 8.
[19] Id., *at* 8-9 (For Ruggie, this meant focusing not just the traditional "old governance" levers of public law and governance at the domestic and international levels but also a 'civil governance system' comprised of various non-adjudicative and civil society or consumer-driven mechanisms to influence corporate behavior, as well as the various corporate governance models operating within large multinational enterprises themselves.)
[20] Id., *at* 10.

claimed, was qualitatively superior to the traditional model of state-monopolized decision-making, and the reason for the unanimous Human Rights Council endorsement that ultimately gave the UNGPs such a powerful normative weight within the contemporary global human rights normative scaffolding.[21]

John Ruggie's approach to the development of global governance norms pertaining to businesses and human rights still represents the 'best practice' model for the international community as it once again grapples with how to develop a human rights-based approach to NETs in today's even more polycentric global governance context.

[21] Id.

# A Choice Between "Disciplinary Lenses"

In its efforts to develop a global governance regime for AI and other NETs, the international community has a choice between various non-exclusive **disciplinary lenses** that it can draw upon to help it manage the inevitable tradeoffs inherent in any governance regime.

At least five such disciplinary lenses can be identified:

| | | |
|---|---|---|
| (1) | The **human rights lens** | Does a given technology advance or undermine **human rights protections**? |
| (2) | The **sustainable development lens** | Does a given technology contribute to the achievement of the **Sustainable Development Goals (SDGs)**? |
| (3) | The **south-south and triangular development lens** | What can be done to harness the power of AI to **"decolonize development"** and promote the interests of nations in the "Global South"? |
| (4) | The **technical lens** | Is it possible to reach a global consensus on **technical standards** to govern AI and other NETs? |
| (5) | The **national security lens** | Does a given technology potentially advance or undermine **national security**? |

This paper assumes that the human rights lens should take priority over the other four possible lenses. Such a bias is, as a matter of necessity, inherent to any governance approach claiming to be "human rights-based."

We are not, however, alone in embracing this bias in favor of the human rights disciplinary lens as the "lodestar"[22] for any attempts to design a global approach towards AI governance. Three recent high-level reports, not to mention countless other policy papers, conferences, and policy initiatives, suggest the same. Of particular note are (1) an Interim Report by the

[22] Quoting from the Interim Report of the High Level Advisory Body on Artificial Intelligence, see *infra*, note 23, *at* 1.

High-Level Advisory Body on Artificial Intelligence published in December of 2023, (2) a UN General Assembly Resolution passed in March of 2024, and (3) a White Paper by the Inter-Agency Working Group on Artificial Intelligence published in May of 2024. Each of these documents is discussed below. Each of these three publications was likely produced in anticipation of (and as a contribution towards) the release of the Global Digital Compact, which was finalized on September 22, 2024 during the Summit of the Future in New York City.

Despite this emerging consensus, this paper still considers it to be a matter of open debate whether the human-rights-based approach should indeed predominate the four alternative disciplinary lenses. Realistically speaking, the global governance of AI, especially in a polycentric world, will inevitably involve a compromise between competing policy lenses. Different actors will undoubtedly advocate for solutions premised on different disciplinary lenses.

The remainder of this section highlights the emerging consensus within the United Nations about the centrality of the human rights disciplinary lens over other possible disciplinary lenses in any discussions about the design of a global approach to AI governance.

## High-Level Advisory Body on Artificial Intelligence

In light of growing calls for the UN to take a proactive role in the governance of AI and other NETs, the UN Secretary-General in October 2023 formed the High-Level Advisory Body on Artificial Intelligence (HLAB-AI). This ad-hoc body brings together 39 experts from a variety of disciplines and backgrounds for a time-limited engagement, culminating in the **Summit of the Future** in September 2024, on which occasion the **Global Digital Compact** (GDC) was announced. The HLAB-AI was the body that shepherded the GDC towards its final form.

**Interim Report (December 2023):**

In December 2023, the HLAB-AI published an Interim Report on Governing AI for Humanity.[23] This report was published to elicit feedback from interested stakeholders, and gave a provisional sense of the broad direction of the emerging consensus that would ultimately inform the GDC.[24]

[23] United Nations: Advisory Body on Artificial Intelligence, Interim Report: Governing AI for Humanity, December 2023, *available at* https://www.un.org/en/ai-advisory-body.
[24] Id., *at* 3.

The Interim Report argued that there exists a clear need for the global governance of AI, and that this global governance role should be housed within the United Nations. The Report's authors claim that "[AI] cries out for governance, not merely to address the challenges and risks but to ensure we harness its potential in ways that leave no one behind."[25] Beyond the need to develop an appropriate universal normative framework to govern the design and deployment of AI globally, the report also notes that "the very nature of the technology itself — AI systems being transboundary in structure, function, application, and use by a wide range of actors — necessitates a global approach."[26] Further, the report argued that the United Nations is uniquely mandated to develop this normative governance framework as it is the only "truly global forum founded on international law, in the service of peace and security, human rights, and sustainable development.[27]

The Report proposes **five overarching principles** that should apply to any global AI governance regime:

| | |
|---|---|
| **(1)** | AI should be **governed inclusively**, by and for the benefit of all. |
| **(2)** | AI must be governed in the **public interest** |
| **(3)** | AI governance should be **built in step with data governance and the promotion of data commons**. |
| **(4)** | AI governance must be **universal**, **networked**, and **rooted in adaptive multi-stakeholder collaboration**; and |
| **(5)** | AI governance should be **anchored in the UN Charter**, **International Human Rights Law**, and **other agreed international commitments such as the Sustainable Development Goals**. |

The AI Advisory Body reiterated Ruggie's notion that the global governance of AI would have to proceed within a polycentric model of global decision-making.

[25] Id., *at* 1.
[26] Id., *at* 4.
[27] Id., *at* 4.

> "At the global level, international organizations, governments, and private sector would bear primary responsibility for these functions. Civil society, including academia and independent scientists, would play key roles in building evidence for policy, assessing impact, and holding key actors to account during implementation. Each set of functions would have different loci of responsibility at different layers of governance — private sector, government, and international organizations."[28]

The Interim report highlighted the following **seven functions** of an effective global AI governance regime,[29] stating that it should be:

| | |
|---|---|
| **(1)** | **nimble and responsive** to the rapidly evolving pace of technological innovation inherent to AI and other NETs. |
| **(2)** | optimized to **bridge the gap between national and global AI governance efforts** and primed to reinforce the centrality of international norms at all levels; |
| **(3)** | mandated to **develop and harmonize standards, safety, and risk management frameworks**…; |
| **(4)** | …while simultaneously enabled to **facilitate the development, deployment, and use of AI for economic and societal benefit** through international multi-stakeholder cooperation; |
| **(5)** | **adequately resourced** to promote international collaboration on talent development, access to computing infrastructure, building of diverse high-quality datasets, responsible sharing of open-source models, and AI-enabled public goods for the SDGs; |
| **(6)** | empowered to **monitor risks**, **report incidents**, and **coordinate emergency response**…; and |
| **(7)** | …designed to **incentivize compliance and accountability based on norms**. |

These principles are reflected prominently in the finalized GDC, which is discussed below (see p.34)

[28] Id., *at* 15.
[29] Id., *at* 15-19. These descriptions have been paraphrased to be more accessible to non-technical audiences.

## UN General Assembly

The UN General Assembly is the United Nations' main policy-making organ. It serves as the supreme venue for any discussions that fall within the scope of the United Nations Charter.[30] The United Nations Charter mandates the General Assembly, among other tasks, to "initiate studies and make recommendations for the purpose of: [ . . . . ] encouraging the progressive development of international law and its codification; [. . . . ] and assisting in the realization of human rights and fundamental freedoms for all without distinction as to race, sex, language, or religion."[31]

**UN General Assembly Resolution on AI**

On 21 March, 2024, the UN General Assembly (GA) adopted the UN's first-ever resolution on AI.[32] In this Resolution, the GA

> "Acknowledges that the United Nations system, consistent with its mandate, *uniquely contributes to reaching global consensus on safe, secure and trustworthy artificial intelligence systems*…" (emphasis added).[33]

In doing so, the UN Member States, speaking through the GA, underscored their intention to place the UN at the center of international governance efforts to promote safe, secure, and trustworthy AI and other NETs, approving the already active landscape of various UN Agencies engaged in such efforts.

The Resolution's preambular language indicates that the members of the General Assembly give primacy to human rights as the disciplinary "lodestar" of any AI governance model. In the first paragraph of the Resolution, the GA "Reaffirms international law, in particular the Charter of the United Nations, and [ ] the Universal Declaration of Human Rights."[34] The GA advances an idealistic (tech-optimist) vision for the role of AI and other NETs in society:

[30] United Nations Charter (June 26, 1945), Article 10, subject to the provisions of Article 12 whereby the Security Council can prevent the General Assembly from speaking on any issue "relative to the maintenance of international peace and security."
[31] *Id.,* Articles 13 and 55.
[32] United Nations General Assembly (2024) Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development, March 11, 2024, *available at* https://digitallibrary.un.org/record/4040897?v=pdf&ln=en#files.
[33] Id., Article 13.
[34] Id.

> "[S]afe, secure and trustworthy artificial intelligence systems – which, for the purpose of this resolution, refers to artificial intelligence systems in the non-military domain [ . . . . ] that [ . . . . ] are human-centric, reliable, explainable, ethical, inclusive, in full respect, promotion and protection of human rights and international law, privacy preserving, sustainable development oriented, and responsible – have the potential to accelerate and enable progress towards the achievement of all 17 Sustainable Development Goals and sustainable development in its three dimensions – economic, social and environmental – in a balanced and integrated manner; promote digital transformation; promote peace; overcome digital divides between and within countries; and promote and protect the enjoyment of human rights and fundamental freedoms for all, while keeping the human person at the centre."

This language underscores the potential of AI and other NETs to help society achieve several socially desirable outcomes, including the fulfillment of the SDGs, the bridging of the digital divide, the promotion of human rights, and the achievement of a more sustainable peace. At the same time, the UN General Assembly resolution also insisted that **international human rights law must serve as the disciplinary lens for managing potential tensions between those noble objectives and potential threats to human dignity**.

The resolution articulated several policy priorities directed in turn towards the United Nations and its constituent Agencies (§§ 1,2,8,10, and 11), UN Member States (§§ 3-7), and the private sector (§§ 5 and 9). Those policy priorities are focused on the following priorities:

(1) bridging the AI and digital divides between countries (§1);
(2) promoting safe, secure, and trustworthy AI systems in furtherance of the SDGs and "the world's greatest challenges" (§§2,6 and 10)
(3) encouraging multi-stakeholder efforts to develop effective regulatory and governance solutions to ensure that AI systems are safe, secure, and trustworthy (§3);
(4) encouraging development initiatives designed to help developing countries also harness the benefits of safe, secure, and trustworthy AI systems (§4);
(5) promoting good data generation and management practices as an integral part of the development of safe, secure, and trustworthy AI (§7);
(6) promoting the central role of the UN as a forum for ongoing research and discussions about the use of AI for good in light of the constantly evolving and rapidly accelerating pace of AI and NEDT innovations (§8); and

(7) reiterating the obligation for States as well as other private actors to "refrain from or cease the use of [AI] systems that are impossible to operate in compliance with international human rights law or that pose undue risks to the enjoyment of human rights, especially of those who are in vulnerable situations, and reaffirms that the same rights that people have offline must also be protected online, including throughout the life cycle of artificial intelligence systems," (§5) and emphasizing the continued validity of the UN Guiding Principles on Business and Human Rights (see below) for the private sector as it develops safe, secure, and trustworthy AI solutions (§9).

# An Overview of the Existing Landscape of AI-Related Work at the United Nations

**"The UN is a management consultant's worst nightmare."**[35]

The United Nations is a uniquely complex organization. The UN's complexity should be understood in light of the organizations' substantive, historical, diplomatic, and institutional history. Any recommendations made in this paper must therefore consider (and respect) the UN's spectacular institutional complexity, as well as the many efforts *already underway* within that system to address the human rights impacts of AI and other NETs. This section provides a brief overview of some of those efforts, categorized them roughly in terms of the disciplinary lenses most prominently advanced by each of these efforts.

As we survey the UN's landscape, it makes sense to start at the very top, with the UN's "longest-standing and highest-level coordination forum": the Chief Executives Board for Coordination.

## UN System Chief Executives Board for Coordination (CEB) and its subordinate High-Level Committees on Programmes (HLCP) and Management (HLCM)

The CEB is a 31-member body chaired by the UN Secretary-General that meets twice a year. It brings together top officials from several prominent UN Agencies.[36] The OHCHR and the Human Rights Council are not represented on the CEB.[37]

[35] Kelly Lee (2005) "Is the UN broken, and how can we fix it?" Vol.331:7516 BMJ 525.
[36] UN System Chief Executives Board for Coordination: Board Members (website, last accessed June 2, 2024), *available at* https://unsceb.org/board-members.
[37] The OHCHR administratively falls within the UN Secretariat, and is therefore represented on the CEB only indirectly by the UN Secretary-General. The Human Rights Council is not an "Agency" of the UN but rather a deliberative body, and therefore also goes without direct representation at the CEB.

The CEB's mandate is to "promote coherence and cooperation on a range of programmatic, policy and management issues faced by UN system organizations."[38] To help it achieve its mandate, the CEB is supported by the High-Level Committee on Programmes (HLCP) and the High-Level Committee on Management (HLCM). The mandate of the former is to

"foster[ ] coherence, cooperation and coordination on the programme dimensions of strategic issues facing the United Nations system",[39] while the latter is to "identif[y] and analyze[ ] administrative management reforms with the aim of improving efficiency and simplifying business practices."[40]

The CEB, with the support of the HLCP and HLCM, in other words, is essentially a body of in-house "**management consultants**" tasked with maintaining the coherence of an inherently multi-disciplinary, multi-mandate, and fragmented institutional landscape.

**Relevant Standards and Institutional Capacities:**

The CEB and its two High-Level Committees are tasked with thinking structurally about how to advance certain inter-agency priorities and initiatives across the vast institutional landscape of the UN. One of those topics is AI.[41] The CEB began to think about AI as a "frontier issue" facing the UN in 2017. In 2020, the HLCP created the Inter-Agency Working Group on AI (IAWG-AI) "to bring together [UN] system expertise on artificial intelligence in [a way that integrates] both normative and programmatic dimensions."[42] The IAWG-AI was co-chaired by the International Telecommunications Union (ITU, see below, p.52) and the United Nations Educational, Scientific, and Cultural Organization (UNESCO, see below, p.40).

[38] UN System Chief Executives Board for Coordination: About (website, last accessed June 2, 2024), *available at* https://unsceb.org/about.
[39] UN System Chief Executives Board for Coordination: High Level Committee on Programmes (HLCP) (website, last accessed June 2, 2024), *available at* https://unsceb.org/structure.
[40] UN System Chief Executives Board for Coordination: High Level Committee on Management (HLCM) (website, last accessed June 2, 2024), *available at* https://unsceb.org/high-level-committee-management-hlcm.
[41] UN System Chief Executives Board for Coordination: Artificial Intelligence (website, last accessed June 2, 2024), *available at* https://unsceb.org/topics/artificial-intelligence.
[42] Id.

On May 2, 2024, the IAWG-AI published its **United Nations System White Paper on AI Governance.**[43] We refer the reader to this high-level white paper for a comprehensive analysis that goes beyond that in this paper of the AI-related activities currently unfolding at the UN. The IAWG-AI White Paper surveys over 50 laws and instruments, some of them binding and others non-binding, "that are either directly applicable to AI or are applied in inter-related areas like ethics, data, cybersecurity, copyrights, patents, information integrity, disarmament, human rights, international labour standards and codes of practice, international humanitarian law, and others."[44] The paper's analysis finds that "[i]nternational law, including the UN Charter and international human rights law, provide the fundamental frameworks that should underpin the design, implementation and operation of [global AI] governance instruments, mechanisms, institutions and processes."[45] This finding underscores the primacy of the human rights disciplinary lens in the context of any effort to create a global AI governance strategy.[46]

Echoing John Ruggie's insights described at the outset of this paper, the document also highlights the crucial importance of "[i]ncluding key stakeholders from the beginning of the process, including relying on their support in piloting the frameworks under development." Doing so, the White Paper claims, "brings legitimacy, helps demonstrate early results, and improves the adoption rate."[47] The IAWG-AI specifically mentions the importance of adopting a "human rights-based approach[ ]" that reaches out to "vulnerable groups" and also proactively seeks the views of "developing countries,"[48] as well as the importance of involving the private sector in any such consultation processes.[49] The White Paper concludes that "for a cross-cutting and transversal topic like AI, the various governance functions [across the UN] are expected to be distributed across multiple entities."[50]

The IAWG-AI working group highlights two priority areas where the Human Rights Council can play a significant role:[51]

---

[43] *Supra* note 12.
[44] Id., *at* 12.
[45] Id., *at* 16.
[46] Id. (The white paper by no means suggests that human rights should be the *exclusive* consideration in designing such a governance strategy, just that it should be the most important doctrinal consideration. The White paper also specifically mentions the sustainable development anchor and the technical standards anchor as other relevant lenses to incorporate into an overall governance approach).
[47] Id., *at* 18.
[48] Id., *at* 29-30.
[49] Id., *at* 30.
[50] Id., *at* 28.
[51] For a full list of the IAWG-AI's recommendations, many of which are also technical or administrative in nature, see pages 34-36 of the White Paper.

**(1) Compliance, monitoring, and enforcement of systems relying on AI and other NETs.**

Drawing on the experience of the UN as it dealt in the past with the threats of nuclear proliferation and global health threats (among other challenges), the IAWG-AI White Paper finds that "[t]he implementation of normative instruments governing global public goods provides important lessons for **transparency, accountability, and redress mechanisms, which are essential for AI governance efforts**. Currently, except for voluntary efforts to monitor AI incidents, there are no internationally coordinated avenues specifically aimed at redress mechanisms for AI harms once they have been reported and recorded." (emphasis added). The White Paper highlights several potentially valuable human-rights reporting and accountability mechanisms, most of which fall under the direct purview of the Human Rights Council.[52] Specifically, the White Paper recommends that:

> **"the capacity needs of [ . . . . ] existing mechanisms to address human rights risks from AI could be supported and enhanced"**

**(2) Building the capacity of the United Nations as a whole to more effectively support Member States' capacity to promote safe, secure and trustworthy AI and other NETs.**

The IAWG-AI White Paper argues that the UN should do a better job building the capacity of Member States to effectively regulate AI.[53]

> Capacity development is key to supporting the implementation of relevant instruments. In this regard, the UN System has a twofold role:
>
> (i) **Develop technical guidance and tools that assist Member States** in translating instruments into national/sub-national legislation, and
>
> (ii) **Provide capacity development support for legislative and enforcement capacities** through development and capacity building programs, training and other avenues."

## The Global Digital Compact

Possibly the defining pronouncement on the global governance of AI and other NETs came in September of 2024 in the form of the Global Digital Compact (GDC). The GDC draws on many of the themes that emerged in previous high-level UN statements on AI, and is likely to bring

[52] Id., *at* 19.
[53] Id., *at* 21.

greater coherence to the many debates and nuances that characterized the discussion over AI governance in recent years.

The idea to formulate the GDC arose in 2023 on the occasion of a global stock-taking of emerging global governance challenges on the 75th anniversary of the UN.[54] It was finalized at the Summit of the Future in September 2024,[55] after several rounds of revision and feedback, and commentary from civil society, activists, numerous experts, and others. Its framers released the GDC with the hope that it could achieve lofty objectives: closing digital divides, promoting human rights, and ensuring responsible AI development. The GDC's authors intend for it to serve as a universal normative framework that can guide the "multi-stakeholder action required to overcome digital, data and innovation divides and innovation divides and to achieve the governance required for a sustainable future."[56] It does so by "[articulating] principles, objectives and actions for advancing an open, free, secure and human-centred digital future, one that is anchored in universal human rights and that enables the attainment of the Sustainable Development Goals."[57] In line with the idea of polycentric governance articulated by Ruggie (above), it proposes a non-exclusively state-centric call to action, addressing not just governments but also corporations, civil society and other stakeholders directly.

The GDC sets forth five key objectives for this multi-stakeholder action agenda,[58] namely to:

**(1) Close all digital divides and accelerate progress across the Sustainable Development Goals;**

**(2) Expand inclusion in and benefits from the digital economy for all;**

**(3) Foster an inclusive, open, safe and secure digital space that respects, protects and promote human rights;**

**(4) Advance responsible, equitable and interoperable data governance approaches; and**

**(5) Enhance international governance of artificial intelligence for the benefit of humanity.**

[54] UNGA, "Declaration on the commemoration of the seventy-fifth anniversary of the United Nations," UN Doc No. A/Res/75/1.
[55] United Nations Office of the Secretary General's Envoy on Technology, The Global Digital Compact (22 September, 2024), https://www.un.org/techenvoy/global-digital-compact.
[56] United Nations, Our Common Agenda Policy Brief 5: A Global Digital Compact – an Open, Free and Secure Digital Future for All (May 2023), https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-gobal-digi-compact-en.pdf.
[57] Id.
[58] *Supra,* note 55, par. 7.

As with so many other frameworks for action, the GDC begins with a description of principles. Noteworthy is the order of the first three principles, prioritizing:

**(1) A commitment to closing the digital divide, primarily by means of "the inclusive participation of all States and other stakeholders;"**

**(2) A rootedness in the 2030 development agenda, with a particular commitment to "address the needs of developing countries, in particular the least developed countries, landlocked developing countries and small island developing States, as well as the specific challenges facing middle-income countries", and**

**(3) An anchor in international law, including international human rights law.**

This order is telling. Human rights serve as the normative anchor for the GDC, whereas the need to close the digital divide and achieve the SDGs serves as the motivation for deploying NETs. Text of the GDC makes specific references to individual SDG goals throughout, positioning the GDC essentially as an "annotation" to the implications of the SDG framework in the digital realm.

The GDC's commitment to inclusivity is one of its strongest aspects. The Compact aims to expand digital access for all, with a specific focus on vulnerable groups who have not enjoyed equitable access to safe digital spaces in the past. The GDC aims to connect the remaining 2.6 billion people globally without access to the internet by 2030, emphasizing affordability and accessibility for all. Human rights groups and others concerned about the persistent and growing digital divide would see this as a significant step forward.

The GDC commits adherents to the following concrete action agendas:

- **To connect all persons to the internet;**[59]
- **To cultivate digital skills and guarantee lifelong access to digital learning opportunities, taking into account the specific social, cultural and linguistic needs of each society and persons of all ages and backgrounds;** [60]
- **To empower individuals and communities with universal access to digital public goods;** [61]

[59] Id., par. 7.
[60] Id., par. 12.
[61] Id., pars. 14-16.

- **To promulgate and respect effective policy, legal and regulatory frameworks that support innovation, protect consumer rights, nurture digital talent and skills, promote fair competition and digital entrepreneurship, and enhance consumer confidence and trust in the digital economy;** [62]
- **To uphold international human rights law throughout the life cycle of NETs so that users can safely benefit from them and are protected from violations, abuses and all forms of discrimination;**[63]
- **To support a multistakeholder internet governance regime that continues to guarantee an open, global, interoperable, stable and secure Internet;**[64]
- **To promote digital trust and safety by "urgently counter[ing] and address[ing] all forms of violence, including sexual and gender-based violence, which occurs through or is amplified by the use of technology, all forms of hate speech and discrimination, misinformation and disinformation, cyberbullying and child sexual exploitation and abuse" and "establish and maintain robust risk mitigation and redress measures that also protect privacy and freedom of expression;"** [65]
- **To guarantee access to "relevant, reliable and accurate information;"** [66]
- **To foster the equitable "development and implementation of data governance frameworks at all levels that maximize the benefits of data use while protecting privacy and securing data;"** [67]
- **To promote the equitable and safe sharing of data across the digital divide and for purposes of achieving the SDGs;** [68]
- **To promote cross-border data flows and data governance regimes;** [69] **and finally**
- **To put in place "a balanced, inclusive and risk-based approach to the governance of [AI], with the full and equal representation of all countries, especially developing countries, and the meaningful participation of all stakeholders."** [70]

The GDC promises to "[initiate], within the United Nations, a Global Dialogue on AI Governance involving Governments and all relevant stakeholders which will take place in the

[62] Id., par. 19.
[63] Id., par. 22.
[64] Id., par. 26-27.
[65] Id., par. 30.
[66] Id., par. 33.
[67] Id., par. 38.
[68] Id., par. 40-43.
[69] Id., par. 46-49.
[70] Id., par. 50.

margins of existing relevant United Nations conferences and meetings."[71] The contours of this Global Dialogue on AI Governance were not yet concretized at the time this paper goes to press, but certainly the Human Rights Council – given its prominence as one of the key "UN conferences and meetings" – stands to play a potential role in hosting such dialogues.

The remainder of this section surveys the most prominent actors currently engaged in efforts to develop a global governance regime for AI and other NETs. This discussion is less comprehensive than that found in the IAWG-AI's white paper but also more customized to the question of whether—and if so, how—to enhance the Human Rights Council's capacity to engage in this debate. The discussion is organized according to the principle "disciplinary lens" that each of these actors espouses as it engages with the question of AI governance.

The first-movers in this discussion so far have been the United Nations Educational, Scientific, and Cultural Organization (UNESCO) and the International Telecommunications Union (ITU). These two organizations, however, are increasingly sharing the mandate to engage with AI and other NETs with a host of other UN Agencies and Organizations. All of these institutional actors, however, are likely gradually to realign their activities in light of the themes described above, most notably the publication of the GDC.

[71] Id., par. 55(b).

## The Human Rights Lens

The 2024 GA Resolution, the IAWG-AI White Paper, and the GDC all reaffirm the primacy of the Human Rights disciplinary lens as the normative anchor when thinking about AI and other NETs. The existing human rights regime may not have been designed with AI and other NETs in mind, but it certainly does not lack answers to how these technologies should be considered, especially if and when they do jeopardize existing human rights.

The **International Bill of Human Rights** is often thought of as the globally applicable human rights "constitution." This doctrine is not codified in any one unified document, however. In its narrowest sense, this doctrine can be thought of as the sum of the non-binding Universal Declaration of Human Rights (UDHR) and the two binding Covenants on Civil and Political Rights (ICCPR) and Economic, Social and Cultural Rights (ICESCR). In its most expansive definition, one might also include the seven additional core human rights treaties as well as a range of optional protocols to those treaties giving rise to monitoring bodies mandated to elaborate upon and oversee for those treaties.[72] An even more expansive definition of the core human rights doctrine might also include the provisions of international labor law promulgated by the International Labour Organization, which directly addresses many of the most commonly-raised issues concerns raised in the context of AI and NETs, notably those pertaining to the right to be free from discrimination, the right to privacy, the freedom of speech, the right to work, freedom of assembly, the rights to health care and education, and the right to enjoy the fruits of scientific progress.

Under this corpus of treaties and oversight mechanisms, states (especially signatory states) have duties to respect, protect, and fulfill a range of human rights.

**Respect:** States cannot take actions that would violate human rights. Regarding AI and other NETs, for example, states could not use AI to discriminate against certain categories of individuals as they seek access to services, for example, vulnerable communities at a border seeking asylum or access to health care, education, or housing. It would also apply to the use of AI in a national security context, supplemented in some instances by provisions of international humanitarian law. This duty speaks most directly to instances where the State itself deploys AI or NETs as part of its governance strategy.

[72] UN Office of the High Commissioner for Human Rights (2016), The Core International Human Rights Treaties, Rev. 1, (New York and Geneva, United Nations, ST/HR/3/Rev.1).

**Protect:** International human rights law also recognizes that States are not the only actors that threaten the enjoyment of human rights. Private actors (corporations, for example) can also jeopardize the universal enjoyment of human rights. In such instances, States are obligated to not passively stand by whilst other actors undermine human rights, but rather to intervene with regulatory and judicial measures to safeguard those rights. This duty requires States to institute realistic and effective regulatory and judicial measures to ensure that private uses of AI systems and those based on other NETs do not infringe upon human rights.

**Fulfill:** Not all rights are instantly realized the minute they are declared. Many rights, specifically economic, social, and cultural rights, are subject to a State's duty to progressively take measures to fulfill those rights. States could, for example, deploy AI and other NETs in ways that *positively* work towards the realization of full human rights, perhaps in the context of progressively realizing the human rights to health care, social security, or universal access to education and justice.

States can also look to a range of soft or 'softer-law' instruments and regulations – often 'owned' by one of the UN sub-agencies – for guidance more specific to AI and other NETs. UNESCO, for example, has taken an important leadership role as the chief advocate for human rights within many discussions over the global governance of AI. It has also convened the process of formulating the first UN-level ethics principles specifically geared towards Artificial Intelligence. Other actors also have an important role to play in the future governance of AI. Of note are the UN Office of the High Commissioner for Human Rights, the International Labour Organization, and, of course, the Human Rights Council.

## United Nations Educational, Scientific, and Cultural Organization (UNESCO)

UNESCO is currently the principal institutional actor advocating for a norms-based approach to the governance of AI. UNESCO was established in 1945 as a stand-alone international organization. Subsequent to the 1947 Agreement specifying the relationship between the UN and UNESCO, it operates as a Specialized Agency of the United Nations.[73] The activities of UNESCO and the broader UN system are coordinated through the Economic and Social Council (ECOSOC).

[73] Protocol concerning the entry into force of the Agreement between the United Nations and the United Nations Educational, Scientific and Cultural Organization (Feb. 3, 1947) UN Treaty Series 11, 234.

During its 2019 General Conference, the UNESCO Member States mandated the organization to formulate non-binding AI ethics recommendations.[74] The Director-General of UNESCO convened an ad-hoc expert group to formulate a first draft of a future AI Ethics Code and then received supplemental support and input from other UN Agencies as intermediated and facilitated by the HLCP (described above, p.31).[75]

**Relevant Standards and Institutional Capacities:**

In November of 2021, the 193 member states of UNESCO adopted the **UNESCO Recommendations on the Ethics of Artificial Intelligence** (UNESCO AI Recommendations). These recommendations "pay[ ] specific attention to the broader ethical implications of AI systems in relation to the central domains of UNESCO: education, science, culture, and communication and information."[76] The recommendations embrace the need to respect, protect, and promote human rights, fundamental freedoms, and human dignity as one of the four core values of the document. The recommendations begin with a set of 10 "principles" (for example, "proportionality and do-no harm" or "safety and security") and then 'operationalize' those principles into various policy recommendations, organized into 11 policy domains. These policy recommendations, which technically apply to all UNESCO Member States (albeit on a voluntary basis), focus on some policy domains that fall within UNESCO's mandate, specifically: (1) Culture and (2) Education and Research, but also many that fall well outside of UNESCO's traditional area of expertise, specifically (3) Health and Social Well-Being, (4) Economy and Labour, (5) Communication and Information, (6) Gender, (7) Environment and Ecosystems, (8) Development and International Cooperation, (9) Data Policy, (10) Ethical Governance and Stewardship, and (11) Ethical Impact Assessments.

The UNESCO AI Recommendations mandate UNESCO to develop various diagnostic and monitoring tools to help them implement the Recommendations.[77] In line with that mandate, UNESCO produced a **Readiness Assessment Methodology** (RAM)[78] in 2023 to help countries conduct a structured macro-analysis of their institutional readiness to promote and regulate AI ethically and responsibly. Similarly, it published an **Ethical**

[74] UNESCO (2020), Records of the General Conference, 40th session, Paris, 12 November-27 November 2019, volume 1: Resolutions, *at* 35, *available at* https://unesdoc.unesco.org/ark:/48223/pf0000372579.
[75] *Supra,* note 41.
[76] UNESCO, Recommendation on the Ethics of Artificial Intelligence, Adopted 23.11.2021, ¶ 3.
[77] Id., ¶ 131.
[78] UNESCO (2023), Readiness Assessment Methodology: A Tool of the Recommendation on the Ethics of Artificial Intelligence (Paris, UNESCO), *available at* https://unesdoc.unesco.org/ark:/48223/pf0000385198.

**Impact Assessment** (EIA)[79] tool in 2023 to help government agencies and private corporations determine whether a specific AI application aligns with the principles and norms set forth in the UNESCO AI Recommendations. UNESCO has also created an online **Global AI Ethics and Governance Observatory**, which brings together key resources for policymakers on AI governance, including the Country Reports of countries that have completed an assessment of their institutional readiness using the aforementioned RAM, a case study and research repository, and various expert networks.[80]

The Recommendations were adapted by the IAWG-AI to guide the use of AI by all UN actors.[81]

## International Labour Organization (ILO):

The ILO was created after the ruins of the First World War to "promot[e] social justice and internationally recognized human and labour rights."[82] The ILO's formation was motivated by a concern that a failure to address the grievances of the world's workers could "produce unrest so great that the peace and harmony of the world [could be] imperilled."[83] Consequently, the ILO's constitution mandates it to seek "an improvement of [labor conditions], for example, by [ . . . . ], the regulation of the labour supply, the prevention of unemployment, the provision of an adequate living wage, [ . . . . ].[84] In 1946, the ILO became a UN Specialized Agency.[85] From its outset, the ILO pursued its mandate via its unique "tripartite" governance model that brings together member states, industry representatives, and labor representatives in everything the organization does. This governance model foreshadows by almost a century the idea of a "thick stakeholder consensus" described by Ruggie above.

---

[79] UNESCO (2023), Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence (Paris, UNESCO), *available at* https://unesdoc.unesco.org/ark:/48223/pf0000386276.
[80] UNESCO, Global AI Ethics and Governance Observatory (website, last accessed June 2, 2024), *available at* https://www.unesco.org/ethics-ai/en.
[81] United Nations Chief Executives Board for Coordination | High-Level Committee on Programmes | Inter-Agency Working Group on Artificial Intelligence, Principles for the Ethical Use of Artificial Intelligence in the United Nations System, Sept. 20, 2022, *available at* https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf.
[82] International Labour Organization, About the ILO (website, accessed June 5, 2024), https://www.ilo.org/about-ilo.
[83] ILO, Constitution (accessed June 5, 2024), https://www.ilo.org/about-ilo.
[84] Id. This list was shortened to focus in particular on those conditions of labor most at risk due to the rise of AI. The full list also includes: "the regulation of the hours of work, including the establishment of a maximum working day and week, [ . . . . ], the protection of the worker against sickness, disease and injury arising out of his employment, the protection of children, young persons and women, provision for old age and injury, protection of the interests of workers when employed in countries other than their own, recognition of the principle of equal remuneration for work of equal value, recognition of the principle of freedom of association, the organization of vocational and technical education and other measures."
[85] Protocol concerning the entry into force of the Agreement between the United Nations and the International Labour Association (Dec. 19, 1946) UN Treaty Series 1, 183.

**Relevant Standards and Institutional Capacities:**

The ILO's mandate includes the promotion of full employment.[86] This was articulated in the Philadelphia Declaration of 1944 which articulated the ILO's mandate and eventually merged with its Constitution in 1946. The Philadelphia Declaration was reaffirmed in the ILO Declaration for Social Justice for a Fair Globalization (2008 and updated in 2022)[87] and the ILO Centenary Declaration for the Future of Work (2019).[88] As AI and other NETs continue to evolve, gradually matching or exceeding human capabilities for certain tasks, increasing concern has focused on the risks of mass unemployment as a result of AI technologies. The human rights aspects of this threat fall squarely within the ILO's mandate. So far, however, the ILO has taken only limited steps in pursuit of its mandate to protect workers from AI-fueled unemployment. It has undertaken several research projects exploring various aspects of AI and labor,[89] but it could still take a significantly more prominent advocacy and norm generation role than it has to date.

## Office of the High Commissioner for Human Rights (OHCHR):

The OHCHR, in recent years, has assumed a much more prominent role in the discussions about the governance of AI, especially regarding the private sector. The OHCHR was established as an office of the UN Secretariat in 1993. Reporting annually to the Human Rights Council (HRC),[90] its mandate overlaps in part with that of the HRC.[91] The Office as a whole remains subordinate to the overarching mandate of the Secretariat and accountable to the HRC and, by extension, that of the UN General Assembly.[92]

**Relevant Standards and Institutional Capacities:**

[86] ILO (2019) Centenary Declaration for the Future of Work, https://www.ilo.org/resource/ilc/108/ilo-centenary-declaration-future-work. *See also* Id., (Annex: Philadelphia Declaration, 1944).
[87] ILO, ILO Declaration on Social Justice for a Fair Globalization (Aug. 13, 2008), https://www.ilo.org/resource/ilo-declaration-social-justice-fair-globalization.
[88] Supra, note 86.
[89] *See* AI for Good, International Labour Organization, (Website, last accessed June 5, 2024), https://aiforgood.itu.int/about-ai-for-good/un-ai-actions/ilo/.
[90] UN General Assembly Resolution on the High Commissioner for the Promotion and Protection of all Human Rights, A/RES/48/141, Dec. 20, 1993 (hereinafter UNGA Res 48-141), and UN General Assembly Resolution on the Human Rights Council, A/RES/60/251, Mar. 15, 2006 (hereinafter UNGA Res 60-251).
[91] See the comparison table below, p.20.
[92] UNGA Res 48-141, *id.* Article 4 "The [OHCHR's] responsibilities shall be [ . . . . ] (b) [t]o carry out the tasks assigned to him/her by the competent bodies of the United Nations system in the field of human rights and to make recommendations to them with a view to improving the promotion and protection of all human rights") (c) To promote and.")

The OHCHR's mandate flows from the constituent treaties that collectively comprise the International Bill of Human Rights.[93] In addition, the OHCHR periodically produces a range of soft-law standards derived from the International Bill of Human Rights.

The most salient normative anchor for the OHCHR's engagement with private sector actors working to develop and deploy AI is the **UN Guiding Principles** (UNGPs).[94] The Human Rights Council imbued the UNGPs with their contemporary authoritative standing by endorsing them by consensus in 2011. While technically still considered to be a "soft" (i.e., non-binding) source of international law, they have been increasingly "embraced by regulators, investors and standard setters around the world,"[95] and are today considered to be "the leading global framework focused on business responsibility and accountability for human rights."[96]

The UNGPs offer "a compelling starting point for companies and States seeking to enhance the positive impact and opportunities of technological innovation by effectively managing associated risks to people."[97] They require corporations to put in place clear corporate policies designed to protect human rights, conduct due diligence about the human rights impact(s) of their various operations, and put in place credible grievance and remediation mechanisms for impacted stakeholders to use if they feel their human rights to have been violated. These standards obviously apply to corporations as they consider working with AI or AI-enhanced products and services.

The OHCHR launched the **B-Tech Project** in 2019 to progressively clarify and elaborate upon the relevance of the UNGPs to any efforts by businesses to develop AI and other NETs. The B-Tech project aims to (1) address the human rights risks inherent in various tech-related business models; (2) help businesses know how to conduct human rights due diligence and end-use assessments; (3) promote accountability and remedies for victims of NETs, and (4) elaborate effective regulatory and policy responses to the address the human rights challenges linked to NETs.[98]

---

[93] *Supra* note 72, Introduction.
[94] United Nations Office of the High Commissioner for Human Rights (2011), Guiding Principles on Business and Human Rights (New York and Geneva, United Nations, HR/PUB/11/04).
[95] United Nations Office of the High Commissioner for Human Rights (2020), An Introduction to the UN Guiding Principles in the Age of Technology: B-Tech Foundational Paper, 5, *available at* https://www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/introduction-ungp-age-technology.pdf.
[96] Id., *at* 3.
[97] Id.
[98] United Nations Office of the High Commissioner for Human Rights (2024), B-Tech Project (website, last accessed June 2, 2024), https://www.ohchr.org/en/business-and-human-rights/b-tech-project.

## Human Rights Council (HRC):

The Human Rights Council dates to 2006, when the UNGA issued a resolution mandating its formation. The HRC is structured as a multilateral body comprised of 47 elected Member States, deliberating on matters of human rights under the umbrella authority of the UN General Assembly. The HRC "is responsible for strengthening the promotion and protection of human rights around the globe."[99] It is also mandated to "address human rights violations and country situations" [respond] "to human rights emergencies and make[ ] recommendations on how to better implement human rights on the ground."[100] In so doing, it "benefits from substantive, technical, and secretariat support from the Office of the High Commissioner for Human Rights (OHCHR)." [101]

[99] OHCHR, About the HRC (website, accessed Jun 5, 2024), https://www.ohchr.org/en/hr-bodies/hrc/about-council.
[100] Id.
[101] Id.

The HRC's mandate and that of OHCHR overlap in many ways, as illustrated by this comparison chart below.

| | **GA Res. 48/141**<br><br>(Creation of the Office of the High Commissioner for Human Rights) | **GA Res. 60/251**<br><br>(Creation of the Human Rights Council) |
|---|---|---|
| Reporting Requirement | Annually (to Human Rights Council and ECOSOC) [102] | Annually (to UNGA) [103] |
| **Promote universal respect for human rights**. | Articles 4.a, 4.f, and 4.h [104] | Articles 2, 4 and 5.i [105] |
| Promote and protect the realization of the **right to development**. | Article 4.c [106] | Article 4 [107] |
| **Engage in dialogue with governments** on the implementation of human rights standards. | Article 4.g [108] | Articles 5.d, 5.f and 5.h [109] |

[102] UNGA Res 48-141, *supra*, note 90 (Article 5, as amended by UNGA Res. 60-251, Article 4.g)
[103] UNGA Res 60-251, *supra*, note 90 (Article 5.j)
[104] UNGA Res 48-141, *supra*, note 90 (Article 4 "The [OHCHR's] responsibilities shall be [ . . . . ] (a) [t]o promote and protect the effective enjoyment by all of all civil, cultural, economic, political and social rights, (f) [t]o play an active role in removing the current obstacles and in meeting the challenges to the full realization of all human rights and in preventing the continuation of human rights violations throughout the world, as reflected in the Vienna Declaration and Programme of Action; [and] (h) [t]o enhance international cooperation for the promotion and protection of all human rights.")
[105] UNGA Res 60-251, *supra*, note 90 (Article 2: "[T]he Council shall be responsible for promoting universal respect for the protection of all human rights and fundamental freedoms for all, without distinction of any kind and in a fair and equal manner"; Article 4: "[T]he work of the Council shall be [ . . . . ] with a view to enhancing the promotion and protection of all human rights, civil, political, economic, social and cultural rights [ . . . . ]; and Article 5: "[T]he Council shall [ . . . . ] (i) [m]ake recommendations with regard to the promotion and protection of human rights.")
[106] UNGA Res 48-141, *supra*, note 90 (Article 4 "The [OHCHR's] responsibilities shall be [ . . . . ] (c) [t]o promote and protect the realization of the right to development and to enhance support from relevant bodies of the United Nations system for this purpose.")
[107] UNGA Res 60-251, *supra*, note 90 (Article 4: "[T]he work of the Council shall be [ . . . . ] with a view to enhancing the promotion and protection of all human rights [ . . . . ] including the right to development.")
[108] UNGA Res 48-141, *supra*, note 90 (Article 4 "The [OHCHR's] responsibilities shall be [ . . . . ] (g) [t]o engage in a dialogue with all Governments in the implementation of [the OHCHR's] mandate with a view to securing respect for all human rights.")
[109] UNGA Res 60-251, *supra*, note 90 (Article 5: "[T]he Council shall [ . . . . ] (d) [p]romote the full implementation of human rights obligations undertaken by States [ . . . . and] (f) [c]ontribute, through dialogue and cooperation, towards the prevention of human rights violations and respond promptly to human rights emergencies [ . . . . and] (h) [w]ork in close cooperation in the field of human rights with Governments [ . . . . ].")

| | | |
|---|---|---|
| **Coordinate human rights activities across the UN** institutional landscape and make recommend ways to better streamline the system. | Article 4.i and 4.j [110] | Articles 3 [111] |
| **Coordinate with regional human rights organizations, national human rights organizations, and civil society**. | | Article 5.h [112] |
| Play an **educative role** with regards to human rights protections, including **capacity building**. | Article 4.d and 4.e [113] | Article 5.a [114] |
| **Address situations of violations of human rights, including gross and systematic violations, and make recommendations thereon**. | | Articles 3 [115] |
| To **serve as a forum for dialogue** on thematic issues. | | Article 5.b [116] |
| To make recommendations to the UNGA for further **development of international law in the field of human rights**. | | Article 5.c [117] |

[110] UNGA Res 48-141, *supra*, note 90 (Article 4 "The [OHCHR's] responsibilities shall be [ . . . . ] (i) [t]o coordinate the human rights promotion and protection activities throughout the United Nations system; [and] (j) [t]o rationalize, adapt, strengthen and streamline the United Nations machinery in the field of human rights with a view to improving its efficiency and effectiveness.")
[111] UNGA Res 60-251, *supra*, note 90 (Article 3: "[T]he Council should [ . . . . ] promote the effective coordination and the mainstreaming of human rights within the United Nations system.")
[112] UNGA Res 60-251, *supra*, note 90 (Article 5: "[T]he Council shall [ . . . . ] (h) [w]ork in close cooperation in the field of human rights with [ . . . . ] regional organizations, national human rights institutions and civil society.")
[113] UNGA Res 48-141, *supra*, note 90 (Article 4 "The [OHCHR's] responsibilities shall be [ . . . . ] (d) [t]o provide [ . . . . ] advisory services and technical and financial assistance, at the request of the State concerned and, where appropriate, the regional human rights organizations, with a view to supporting actions and programmes in the field of human rights [and] (e) [t]o coordinate relevant United Nations education and public information programmes in the field of human rights.")
[114] UNGA Res 60-251, *supra*, note 90 (Article 5: "[T]he Council shall [ . . . . ] (a) [p]romote human rights education and learning as well as advisory services, technical assistance and capacity-building, to be provided in consultation with and with the consent of Member States concerned.")
[115] UNGA Res 60-251, *supra*, note 90 (Article 3: "[T]he Council address situations of violations of human rights, including gross and systematic violations, and make recommendations thereon [ . . . . ].")
[116] UNGA Res 60-251, *supra*, note 90 (Article 5: "[T]he Council shall [ . . . . ] (b) [s]erve as a forum for dialogue on thematic issues on all human rights.")
[117] UNGA Res 60-251, *supra*, note 90 (Article 5: "[T]he Council shall [ . . . . ] (c) [m]ake recommendations to the General Assembly for the further development of international law in the field of human rights.")

| | | |
|---|---|---|
| To **follow-up to summits and high-level conferences**. | | Article 5.d [118] |
| To conduct a **Universal Periodic Review** (UPR) | | Article 5.e [119] |

The HRC and OHCHR are completely different bodies, however. The former serves as a "specialized" multilateral deliberative body of the General Assembly, tasked with the creative, generative, diplomatic, and also adjudicative aspects of human rights promotion within the United Nations. The latter is an operational division of the UN Secretariat, and can be thought of more as playing an executive and programmatic role. Given the HRC's direct hand in facilitating the development in international law standards pertaining to human rights, the question of how to establish a human-rights based approach to the global governance of AI and other NETs falls squarely within the HRC's core mandate.

**Relevant Standards and Institutional Capacities:**

The HRC, on its own, does not typically speak on issues of Human Rights. Rather, UN Member States serving on the Human Rights Council propose certain initiatives during the periodic HRC sessions, which the HRC—supported by the OHCHR and other relevant actors—subsequently deliberates and potentially acts upon in the form of resolutions or other similar proclamations calling for concerted effort by other UN agencies, organizations, or member states.

When faced with an emerging thematic issue of great relevance to the global human rights regime, the UN HRC has typically responded by either referring the issue for further study, for example to an ad-hoc **group of experts** or its own in-house **Advisory Committee** for further study and analysis, or by creating various **Special Procedures** tasked with the

[118] UNGA Res 60-251, *supra*, note 90 (Article 5: "[T]he Council shall [ . . . . ] (d) [ . . . . ] follow-up to the goals and commitments related to the promotion and protection of human rights emanating from United Nations conferences and summits.")
[119] UNGA Res 60-251, *supra*, note 90 (Article 5: "[T]he Council shall [ . . . . ] (e) [u]ndertake a universal periodic review, based on objective and reliable information, of the fulfilment by each State of its human rights obligations and commitments in a manner which ensures universality of coverage and equal treatment with respect to all States; the review shall be a cooperative mechanism, based on an interactive dialogue, with the full involvement of the country concerned and with consideration given to its capacity-building needs; such a mechanism shall complement and not duplicate the work of treaty bodies [ . . . . ].")

long-term study, documentation, and normative development of a particular sub-topic of human rights.[120]

In recent years, the "Core Group" of Nations in the Human Rights Council has been pushing for more concerted action on AI and other NETs. In 2019, on the initiative of this Core Group, the Council adopted Resolution 41/11[121] in which it recognized that digital technologies have the potential to accelerate human progress and to facilitate efforts to promote and protect human rights. The resolution requested the HRC's Advisory Committee to prepare a report exploring the human rights implications of new and emerging digital technologies as well as the potential role of international human rights mechanisms in helping to address those issues.

The Advisory Committee presented its subsequent report to the HRC in June 2021.[122] The Advisory Committee's report noted the paradox that new and emerging technologies represent both an opportunity for human rights progress as well as a potential threat to existing protections. It also noted the central role of private actors in the development of these technologies. The report highlighted various conceptual and institutional gaps in the existing human rights system's ability to adequately 'nudge' these NETs in line with human rights priorities. Challenges included the still-unresolved philosophical questions about the applicability of certain human rights to this question, a lack of cooperation or familiarity between the human rights and the tech communities, and a selective preoccupation with only a subset of technologies and human rights harms as opposed to a more holistic approach that also embraced the potential 'upsides' of technological innovation.

[120] OHCHR, Special Procedures of the Human Rights Council (website, last accessed June 5, 2024), https://www.ohchr.org/en/special-procedures-human-rights-council (These special procedures can come in the form of a special rapporteur—usually a term-limited expert appointed to study a particular issue—or a working group—usually a small group of term-limited experts tasked with addressing a particularly complex or multifaceted issue. Special Procedures can be focused on a country situation, for example, the UN Special Rapporteur on the Situation of Human Rights in Myanmar, or a particular sub-theme of human rights, for example, the UN Working Group of Experts on People of African Descent or the UN Special Rapporteur on the Rights to Freedom of Peaceful Assembly and of Association. Currently, there are Special Procedures focusing on 46 thematic and 14 country mandates.)

[121] Human Rights Council Resolution 41/11, "New and emerging digital technologies and human rights," (Jul. 17, 2019), UN Doc. No. A/HRC/RES/41/11.

[122] Human Rights Council Advisory Committee, "Possible impacts, opportunities and challenges of new and emerging digital technologies with regard to the promotion and protection of human rights," (June 2021), UN Doc. No. A/HRC/47/52, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/110/34/PDF/G2111034.pdf.

Building on the Advisory Committee's report, the HRC in July of 2021 adopted Resolution 47/23, in which it requested for the OHCHR to convene expert consultations exploring the links between the human rights impacts of new and emerging technologies and technical standard-setting.

In June 2023, the 'Core Group' narrowed its focus on AI systems. The HRC's Resolution 53/29[123] highlighted AI's potential to "facilitat[e] access to information and participation in public life, strengthen[ ] the efficiency and accessibility of health-care services, enable[ ] greater availability and accessibility of education, advance[e] gender equality and empower[ ] all women and girls, contribut[e] to the full enjoyment of human rights by older persons, persons with disabilities and those in vulnerable situations, strengthen[ ] climate mitigation and adaptation and support[ ] environmental protection", while also recognizing that "certain application of AI present an unacceptable risk to human rights."[124] It also emphasized the importance of a human rights-based approach to new and emerging digital technologies[125] by protecting individuals from harm, notably through human rights due diligence and impact assessments, guarding against discrimination and bias, promoting algorithmic transparency, ensuring that data collection, storage, and use is consistent with human rights obligations, and strengthening oversight and enforcement capacity. The resolution also encouraged multi-stakeholder collaboration and further exploration of ways for the Human Rights Council to promote the human-rights-based approach to AI.

The HRC's Resolution 53/29 also asked OHCHR to conduct an institutional gaps analysis examining "the work and recommendations of the Human Rights Council, the [OHCHR], the treaty bodies and the special procedures of the [HRC] in the field of human rights and new and emerging digital technologies, including artificial intelligence, as well as identifying gaps and challenges and making recommendations on how to address them."[126] After an exhausting analysis highlighting the tremendous volume of work already done by the HRC and its subsidiary special mandate holders, OHCHR published its Mapping Report during the HRC's June-July 2024 56th Session. The key findings of that report are highlighted in URG's parallel paper in this section.

[123] Human Rights Council Resolution 53/29, "New and emerging digital technologies and human rights," (Jul. 14, 2023), UN Doc. No. A/HRC/RES/53/L.27/Rev.1 (as orally revised).
[124] Id., *at* Preamble.
[125] Id., *at* Art. 3.
[126] Id., *at* Art. 5.

Numerous existing Special Mechanisms mandate holders have also weighed in on the topic of AI, each of them using the lens of their mandates to focus their analysis:

- In 2016, the **Independent Expert on the Rights of Older Persons** discussed both the opportunities as well as the challenges associated with the potential and challenges of the use of assistive and robotics technology, AI, and automation for the enjoyment by older persons of their human rights.[127]
- In 2018, the **UN Special Rapporteur on freedom of expression** examined the impact of AI and other NETs on human rights in the information environment, focusing in particular on the rights to freedom of expression, privacy, and non-discrimination.[128]
- In 2021, the **Special Rapporteur on the Rights of Persons with Disabilities** published a report in which he highlighted the potential of AI systems to improve accessibility through assistive and mobility-enhancing technologies, but also potentially lead to discriminatory outcomes.[129]
- In 2019, the **Special Rapporteur on extreme poverty** stressed the huge potential of digital technologies, including artificial intelligence, to "improv[e] the well-being of the less well-off members of society," albeit but not without "deep changes in existing policies [towards a more] genuine commitment to [. . . .] ensure a decent standard of living for everyone in society."[130]
- In 2020, **The Special Rapporteur on racism**, the CERD and the Special Rapporteur on persons with disabilities have identified numerous instances where the use of algorithmic systems have led to discriminatory outcomes, including in access to health care, justice, or employment opportunities.[131]

## The Sustainable Development Lens

The most frequently-discussed alternative to the human rights lens described above is undoubtedly the Sustainable Development Lens. Sustainable development has been a catchphrase since at least the early 1990s, when the United Nations decided at the Rio Earth Summit to amend more neo-conservative notions of *economic* development with a focus also on the impacts of that development on the environment. This focus on the environment was later supplemented by concerns about the social impacts of development, leading to a

[127] Rosa Kornfeld-Matte, Report of the Independent Expert on the enjoyment of all human rights by older persons, Human Rights Council, UN Doc. No. A/HRC/36/48, https://www.ohchr.org/en/documents/thematic-reports/ahrc3648-report-independent-expert-enjoyment-all-human-rights-older.
[128] David Kaye, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, General Assembly, UN Doc. No. A/73/348.
[129] Gerard Quinn, Report of the Special Rapporteur on the rights of persons with disabilities, Human Rights Council, UN Doc. No. A/HRC/49/52, https://www.ohchr.org/en/documents/thematic-reports/ahrc4952-artificial-intelligence-and-rights-persons-disabilities-report.
[130] Philip Alston, Digital welfare states and human rights - Report of the Special Rapporteur on extreme poverty and human rights, General Assembly, UN Doc. No. A/74/48037, https://www.ohchr.org/en/documents/thematic-reports/a74493-digital-welfare-states-and-human-rights-report-special-rapporteur.
[131] E. Tendayi Achiume, Racial discrimination and emerging digital technologies: a human rights analysis, Human Rights Council, UN Doc. No. A/HRC/44/57,

contemporary understanding of *sustainable* development. At the turn of the millennium, this broadened notion of development was captured and defined in terms of the quantifiable metrics of the Millennium Development Goals, and subsequently in terms of the 17 Goals, 169 subordinate targets, and 231 quantifiable and measurable correlated indicators of the **Sustainable Development Goals** (SDGs). The SDGs are meant to guide global development policies between 2015 and 2030.

Numerous recent UN pronouncements on AI and other NETs, including the High-Level Advisory Body on AI's Interim Report (*see above* p. 26), the UN GA Resolution on AI (*see above* p. 29), the IAWG-AI White Paper (*see above* p. 31), and most recently the GDC (*see above* p. 34), all reference the potential for AI and other NETs to help achieve the SDGs by 2030. Each of these documents also adds the crucial caveat, however, that exploring AI in furtherance of sustainable development goals should always be subject to the overriding priority of safeguarding and promoting human rights.

The United Nations Development Programme is typically regarded as the chief UN Agency tasked with coordinating progress towards the gradual achievement of the SDGs. The International Telecommunications Union (ITU) has also contributed substantially to discussions about the potential relevance of AI and other NETs in furthering the SDGs.

## International Telecommunications Union (ITU)

Founded in 1865, the ITU is the oldest agency in the UN. Like UNESCO, it operates as a specialized agency of the United Nations.[132] coordinating its activities with the remainder of the UN system through the Economic and Social Council (ECOSOC).

**Relevant Standards and Institutional Capacities:**

The ITU's mission is to "promote, facilitate and foster affordable and universal access to telecommunication/ information and communication technology networks, services and applications and their use for social, economic and environmentally sustainable development."[133] Within that mandate, the ITU focuses on standardization, not only of

[132] Agreement between the United Nations and the International Telecommunication Union approved by the Plenipotentiary Telecommunication Conference on 4 September 1947 and by the General Assembly of the United Nations on 15 November 1947, and Protocol concerning the entry into force of the said Agreement. Signed on 26 April 1949, UN Treaty Series 30, 316.
[133] ITU, Vision and Mission (website, last accessed June 2, 2024), *available at* https://www.itu.int/en/council/planning/Documents/ITU-Strategic-Plan-2024-2027-Vision%26Mission.pdf.

existing technologies but also future ones.[134] Currently, in addition to AI, the ITU is also developing standards related to the Internet of Things and Quantum Computing.[135]

In addition to its technical recommendations on AI standards, the ITU has also created several policy-focused initiatives meant to highlight the positive applications of AI. The most prominent of these forums is its year-round digital platform called **AI for Good**. This platform, which also routinely holds summits bringing together a broad cross-section of stakeholders from various UN agencies, other international organizations, national government representatives, civil society, industry, and academia, is focused on "[the identification of] practical applications of AI to advance the SDGs and scale those solutions for global impact."[136] The ITU coordinates the activities of over 40 other UN Agencies, partly driven by the annual publication of the **UN Activities on AI Report**, and partly through the co-hosting of the AI for Good platform.

The ITU has also partnered with UNESCO (described above) to informally co-host and coordinate efforts across the UN to engage with AI and NETs. It co-chaired the IAWG-AI, for example, and frequently promotes itself as uniquely qualified to bring its substantial technical expertise into discussions over NETs.

## United Nations Development Program (UNDP)

The UNDP was created in 1965 as a Programme subordinate to the UN General Assembly.[137] The UNDP's mandate was to ensure that "[UN] assistance programmes are designed to support and supplement the national efforts of developing countries in solving the most important problems of their economic development, including industrial development."[138] The UNDP has since grown to be the seventh largest administrative unit of the UN Agency by staff, after the UN Secretariat, UNICEF, UNHCR, WFP, WHO, and the IOM (2022 figures).[139]

[134] ITU, Report by the Secretary General: Report on the Implem,entation of the Strategic Plan and the Activities of the Union, April 2018-June 2022, Plenipotentiary Conference (PP-22) Bucharest, Sept. 26 – Oct. 14, 2022, 10, *available at* https://www.itu.int/en/council/planning/Documents/Report-activities-2018-2022-PP22-doc20-final.pdf.
[135] Id., *at* 13.
[136] ITU, AI for Good: About (website, last accessed June 2, 2024), *available at* https://aiforgood.itu.int/about-ai-for-good/.
[137] UN General Assembly, Consolidation of the Special Fund and the Expanded Programme of Technical Assistance in a United Nations Development Programme, Nov. 22, 1965, UN Doc. A/RES/2029 (XX).
[138] Id.
[139] UN CEB, Personnel by Organization (website, last accessed June 7, 2024), https://unsceb.org/hr-organization.

**Relevant Standards and Institutional Capacities:**

In 2024, the UNDP is poised to become one of the dominant players in the discussion about the use of AI an other NETs in service of sustainable development. Given its size, global field presence, and intensifying ambitious to harness the potential of AI in service of its mandate, the UNDP is already a central player in the debate over global AI governance. In 2023, the ITU and UNPD jointly published an **SDG Digital Acceleration Agenda**.[140] The authors of that study strike an alarmist tone about our collective global ability to achieve the SDGs by the year 2030 *without* the use of AI and other NETs.

> "The world is far behind in achieving the SDGs. Extreme poverty is rising, and progress is stalling with respect to key development targets including those on health, education, and gender. There is an urgent need to re-focus efforts and attention on the SDGs, get back on track, and support people, peace, and prosperity around the world.
>
> Digital technologies are catalysing positive transformation alongside economic and societal growth. Countries with digital foundations may find it easier to meet key development outcomes. Leveraging digital tools, technologies and data can help accelerate progress towards 70 per cent of SDG targets directly, while indirectly supporting the other 30 per cent. As highlighted by the 34 digital solutions presented here, digital transformation are having real and exciting impact and accelerating progress towards the SDGs: from ensuring financial inclusion, to building crucial skills and knowledge and improving the effectiveness of public and private service delivery. Emerging technologies could yet accelerate many of these benefits.[141]

In line with this growing sense of urgency about the need to embrace AI as a central part of a global sustainable development agenda.[142] The UNDP currently "provide[s] tools and strategic advice to help ensure digital transformation is purposefully planned and implemented, and that no one is left behind, [ . . . . ] offer[s] technical support and facilitate[s] knowledge sharing to accelerate [governments' digital capacities, and works] toward universal, meaningful connectivity, digital inclusion and digital rights [ . . . . ]."[143]

---

[140] ITU, UNDP (2023) SDG Digital Acceleration Agenda, *available at* https://www.undp.org/sites/g/files/zskgke326/files/2023-09/SDG%20Digital%20Acceleration%20Agenda_2.pdf.
[141] Id., *at* 82.
[142] *See* UNDP: Digital – Overview (website, *last accessed* June 7, 2024), https://www.undp.org/digital/overview.
[143] Id.

In 2019, the UNDP released its First Digital Strategy in which it outlined its efforts to harness the power of NETs to improve its service delivery to its UNDP partners, as well as to optimize its *internal* capabilities.[144] That initial strategy was replaced in 2022 with a second **Digital Strategy** (2022-2025),[145] which had as its objective to "reimagine development in a digital age."[146]

The Strategy outlines six "guiding principles" that govern the UNDP's engagement with NETs, namely that its efforts will:[147]

| | |
|---|---|
| **(1)** | Put **human rights at the center**; |
| **(2)** | Promote **inclusive and gender-sensitive approaches that leave no one behind**; |
| **(3)** | **Contribute to shared global standards** and frameworks that protect people's rights; |
| **(4)** | **Advocate for open digital standards and open data**; |
| **(5)** | Work to **strengthen local digital ecosystems**; and |
| **(6)** | Leverage **strategic partnerships** to catalyze inclusive approaches to digital development. |

[144] UNDP (2019), Future Forward: UNDP Digital Strategy, *available at* https://digitalstrategy.undp.org/documents/UNDP-digital-strategy-2019.pdf.
[145] UNDP (2022), Digital Strategy 2022-2025, *available at* https://digitalstrategy.undp.org/documents/Digital-Strategy-2022-2025-Full-Document_ENG_Interactive.pdf.
[146] Id., *at* 6.
[147] Id., *at* 12-13.

The strategy also articulates five elements for UNDP's unique "value proposition" in this space:[148]

| | |
|---|---|
| **(1)** | The UNDP's **broad mandate** and **integrator role** in the UN System; |
| **(2)** | Its **longstanding expertise** in supporting governments on digital transformation; |
| **(3)** | Its **rights-based, intentionally inclusive approach**; |
| **(4)** | Its **proactive consideration of the potential risks of digital technology**; and |
| **(5)** | Its unparalleled **country presence**. |

The UNDP will highlight the centrality of digital technologies to its mandate in the **2025 UN Development Report**, which serves as its flagship annual publication, the theme of which will be "Harnessing digital transformation to advance human development."[149]

In line with this focus on the potential of NETs to advance sustainable development, the UNDP has carried out numerous NEDT-related development projects. In the Philippines, for example, UNDP partnered with a digital think tank to develop an AI model that could enhance its efforts to accurately target poverty alleviation interventions.[150] In a similar vein, the UNDP is also experimenting with a Digital Social Vulnerability Index designed "to monitor and understand the exact location, distribution and underlying drivers of social vulnerabilities" that "incorporate[s] a much more comprehensive [social vulnerability] analysis [than previous methods] by integrating numerous data sources and indices into one."[151] In Mexico, the UNDP partnered with local governments to use AI to better understand the impact of government initiatives to combat Sexual- and Gender-Based Violence and government performance metrics.[152] Meanwhile, the UNDP's **Accelerator Labs** is "the world's largest and fastest learning network on wicked sustainable

[148] Id., *at* 14-15.
[149] UNDP, Human Development Reports: News "Announcing the theme of the 2025 Human Development Report: Harnessing digital transformation to advance human development" (June 4, 2024, Press Release, available at https://hdr.undp.org/content/announcing-theme-2025-human-development-report-harnessing-digital-transformation-advance).
[150] Thinking Machines Data Science, "Understanding poverty in the Philippines with artificial intelligence," (website, last accessed June 7, 2024), https://stories.thinkingmachin.es/poverty-mapping-artificial-intelligence/.
[151] UNDP (2024) Digital Social Vulnerability Index Technical Whitepaper (New York: United Nations), *available at* https://www.undp.org/publications/digital-social-vulnerability-index-technical-whitepaper.
[152] UNDP44-47

development challenges. [ . . . . ] cover[ing] 115 countries, and tap[ping] into local innovations to create actionable insights and reimagine sustainable development for the 21st century."[153] The UN also offers numerous services and assessment tools for the use of national governments,[154,155,156,157,158] and private organizations and individuals[159] to help them assess their readiness for digital transformation in a development context.

## World Bank (WB)

The World Bank Group (informally referred to as the World Bank) is composed of five subordinate institutions. The **International Bank for Reconstruction and Development** (IBRD) was created in 1944 to help rebuild European economies after the destruction of the Second World War. It has since become "the largest development bank in the world," mandated to "provid[e] loans, guarantees, risk management products, and advisory services to middle-income and creditworthy low-income countries.[160] The **International Development Association** (IDA) was established in 1960 to "complement" the IBRD by "grant[ing] and low-interest loans help [low-income] countries invest in their futures, improve lives, and create safer, more prosperous communities around the world."[161] The **International Finance Corporation** (IFC) focuses its investments on private sector investment opportunities, "empower[ing] visionary entrepreneurs to bring sustainable solutions to scale."[162] The **Multilateral Investment Guarantee Agency** (MIGA) "provid[es] [ . . . . ] political risk insurance and credit enhancement [ ] to investors and lenders" in order to "promote cross-border investment in developing countries."[163] Finally, the

[153] Berditchevskaia, A., Peach, K., Lucarelli, G., Ebelshaeuser, M. (2021). Collective Intelligence for Sustainable Development: 13 Stories from the UNDP Accelerator Labs, 1. *available at* https://media.nesta.org.uk/documents/UNDP_CI_Report2_final_20210521.pdf.
[154] UNDP, Data to Policy Navigator (website, *last accessed* June 7, 2024), https://www.datatopolicy.org (a tool to "assist government executives and policymakers [grasp] the fundamentals of data-driven decision-making [by] provid[ing them with] a step-by-step guide and a range of practical examples from across the globe on how to integrate data into policy and programme development.")
[155] UNDP, Data Futures Exchange (website, *last accessed* June 7, 2024), https://data.undp.org ("an open-source central hub for data innovation for development impact [that uses] a systems approach and advanced analytics to identify actions to accelerate sustainable development around the world.")
[156] UNDP, Digital Development Compass (website, *last accessed* June 7, 2024), https://www.digitaldevelopmentcompass.org (A tool designed to "[l]everage the pillars of The [UNDP's] Digital Transformation Framework to discover & compare [national] progress across a range of key issues.]
[157] UNDP, Digital X (website, *last accessed* June 7, 2024), https://digitalx.undp.org/ ("A [program] designed to find, match, and connect ready-to-scale digital solutions with the urgent needs of UNDP Country Offices and governments in 170 countries.")
[158] UNDP, IVERIFY (website, *last accessed* June 7, 2024), https://www.undp.org/digital/iverify (An initiative to "[s]upport[ ] actors around the world for the prevention and mitigation of disinformation, misinformation and hate speech.")
[159] UNDP, Digital Guides (website, *last accessed* June 7, 2024), https://digitalguides.undp.org.
[160] World Bank: International Bank for Reconstruction and Development: Who We Are, (website, *last accessed* June 8, 2024), https://www.worldbank.org/en/who-we-are/ibrd.
[161] IDA, What is IDA (website, *last accessed* June 8, 2024), https://ida.worldbank.org/en/what-is-ida.
[162] IFC, Who We Are (website, *last accessed* June 8, 2024), https://www.ifc.org/en/about.
[163] MIGA, About Us (website, *last accessed* June 8, 2024), https://www.miga.org/about-us.

**International Centre for Settlement of International Disputes** (ICSID) specializes in international investment dispute settlement as stipulated in many international investment treaties and numerous national investment laws and/or private contractual agreements.[164] The IBRD,[165] IDA[166] and IFC[167] are Specialized Agencies of the United Nations, whereas MIGA and the ICSID remain unaffiliated with the UN.

**Relevant Standards and Institutional Capacities:**

Like most banks, the World Bank makes its investment decisions based partly on its evaluation of the likely Returns on any Investments it makes (RoI). This reality causes some tension with actors who might prefer other criteria, such as the human rights implications of a certain investment or the objective level of humanitarian need a community might face, as more relevant determinants directing global financial flows. Nonetheless, the World Bank –remains an important actor in the global development landscape. This is also true for AI, where WB funding flows can often catalyze the growth of an AI industry, thus potentially helping to bridge the global divide between the Global North and the Global South.

In 2021 the WB used its flagship annual report (the World Development Report) to focus on the use of data as a driver for development.[168] Taking an economist's perspective on the value of data, the report posits that just like "capital, land, and labor, data are also an input to the development objectives [ . . . . ], [but] unlike capital, land, and labor, using data once does not diminish its value."[169] The report highlights three pathways linking the use of data to development objectives: (1) "the use of data by governments and international organizations to support evidence-based policy making and improved service delivery" (2) "the use of data by civil society to monitor the effects of government policies and by individuals to enable them to monitor and access public and commercial services" and (3) "the use of data by private firms in the production process [ . . . . ]."[170]

[164] ICSID, About ICSID (website, *last accessed* June 8, 2024), https://icsid.worldbank.org/about.
[165] Protocol concerning the entry into force of the Agreement between the United Nations and the International Bank for Reconstruction and Development. (Apr. 15, 1948) UN Treaty Series 16, 341.
[166] Agreement on Relationship between the United Nations and the International Finance Corporation. (Feb. 20, 1957) UN Treaty Series 265, 312.
[167] Protocol concerning the entry into force of the Agreement Approved by the Board of Governors of the International Development Association (Apr. 16, 1961) UN Treaty Series 394, 221.
[168] World Bank (2021) Data for Better Lives, Washington, DC: World Bank, https://www.worldbank.org/en/publication/wdr2021.
[169] Id., *at* 3.
[170] Id., *at* 3.

The WB calls for a new "social contract" for data; "one that enables the use and reuse of data to create economic and social *value*, promotes *equitable* opportunities to benefit from data, and fosters citizens' *trust* that they will not be harmed by misuse of the data they provide" while also not disadvantaging lower-income countries because they lack the adequate infrastructure and skills to fully benefit from such data-driven development.[171]

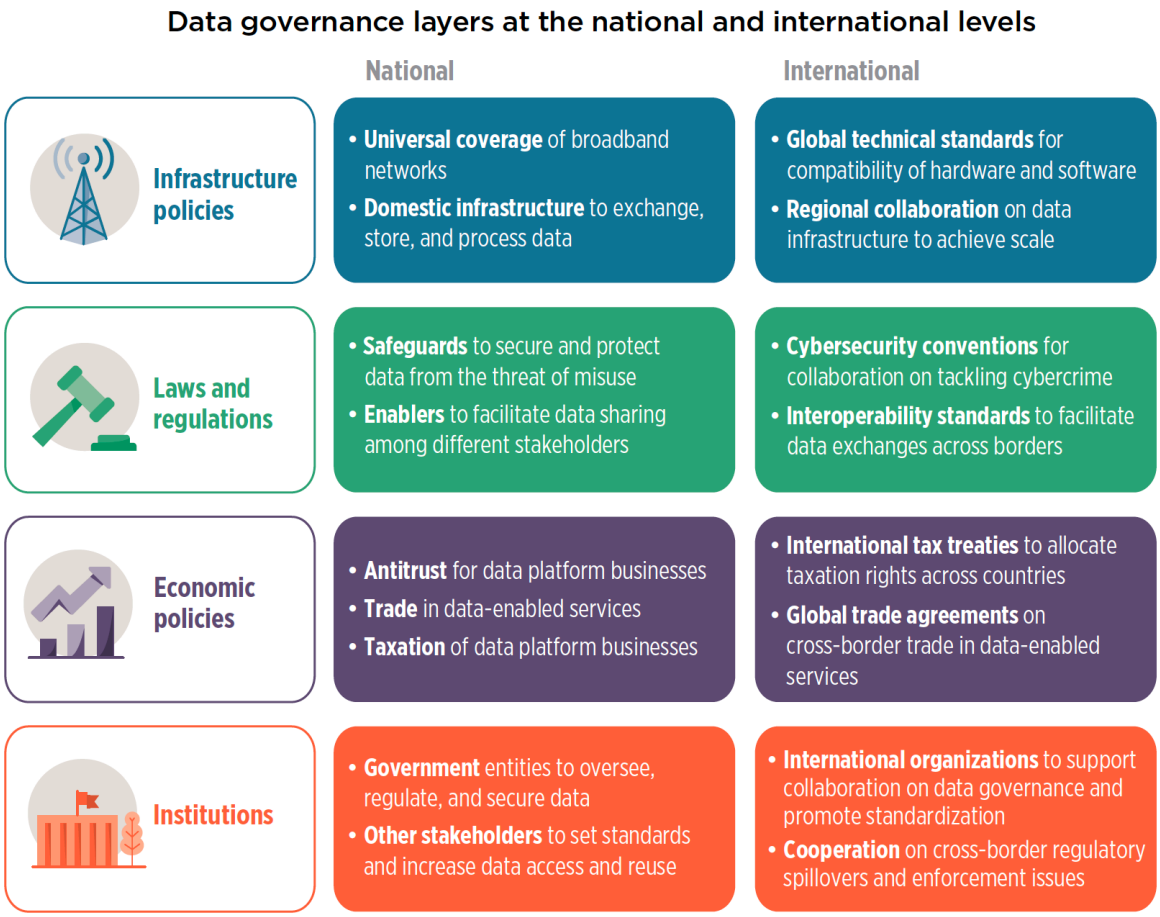
The report presents a data governance framework to further this vision, broken down by whether they are best managed at the national or international level.[172] This governance framework consists of four layers: (1) a "foundational layer [ . . . . ] promot[ing] universal access to internet data services and the policies that ensure that countries have adequate infrastructure to exchange, store, and process data efficiently over the internet;" (2) a legal legal and regulatory layer, grounded in international human rights law,[173] that "creates [the] rules to enable the reuse and sharing of data while safeguarding against their potential abuse and misuse;" (3) a related but distinct economic policy layer that speaks to the "country's ability to harness the economic value of data through competition, trade, and taxation;" and (4) an institutional layer optimized to ensure "ensure[ ] that data can deliver on their potential and that laws, regulations, and policies are effectively enforced."[174]

[171] Id., *at* xi.

[172]

[173] Id., *at* 13, 189-221.

[174]



*Source:* WDR 2021 team.

This overarching framework has helped guide the WB's investments in AI capacity building since 2021. Since 2018, the WB Group has initiated at least 45 AI-linked projects.[175] Some of these projects have focused on FinTech, the prediction and response to poverty using better data analytics, agriculture and the energy sector.[176] Many of these are driven by the WB's in-house **ITS Technology & Innovation Lab** (ITSTI).[177] This lab "appl[ies] a multi-disciplinary approach to explore the potential of different emerging technologies to [ . . . . ] determine whether emerging technologies can help solve development challenges in more effective and efficient ways in specific country contexts."[178] This case-specific research is then "packaged in a [ ] report [to] support the prototype's incubation, scaling, and further action to operationalize the solution for [other World Bank Group]

[175] Rabi Thapa, "Developing AI for development" (Apr. 9, 2024) World Bank Newsletter, https://accountability.worldbank.org/en/news/2024/Developing-AI-for-development.
[176] Sudha Krishnaswami, "How the World Bank Leverages AI to Help Low-Income Families," Blog (Oct. 14, 2023), https://borgenproject.org/how-the-world-bank-leverages-ai/.
[177] WB: ITS Technology & Innovation Lab: Factsheet, (date unknown) *available at* https://thedocs.worldbank.org/en/doc/724241569427635399-0250022019/render/WBGITSInnovationLabDigital.pdf
[178] Id.

operations."[179] GovTech is a separate WB project designed to "bring governments and technology together to transform public services."[180] GovTech generates policy guidance, shares best practices, and partners with the ITSTI to test potential NETs, such as a tool to better identify tax fraud that relies on synthetic AI-generated data.[181]

## The South-South and Triangular Industrial Cooperation Lens

Related but slightly different in emphasis from the sustainable development lens is what we are calling in this paper the South-South and Triangular Industrial Cooperation (SSTIC) lens. This term comes from UNIDO, which is the principal proponent of this frame. The SSTIC lens is prominently reflected in the GDC (see above p.34), in that document's strong focus on closing the digital divide and promoting an inclusive dialogue focusing in particular on the needs of less-developed nations.

The focus of SSTIC development can be described in several ways. One way would be to describe it as a style of development focused on the "decolonialization" of the development enterprise, focusing on strategies of development that no longer rely on a "patron-client" pattern of development cooperation between the industrialized and generally more wealthy nations of the so-called "Global North" and those generally less-industrialized and less wealthy nations of the so-called "Global South." This approach to development focuses on a model of rapid industrialization, often brought about by means of protectionist policies and sometimes at the expense of the human rights agenda (the so-called "Asian values" critique of human rights comes to mind). This model of development is often associated with China, but China is by no means the only nation to embrace such an approach to development.

### United Nations Industrial Development Organization (UNIDO)

Established in 1979,[182] UNIDO became a UN Specialized Agency in 1985.[183] The organization's mandate is to "promote, dynamize and accelerate industrial development" as reflected in SDG-9: to "[b]uild resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation."[184] UNIDO provides support to its 172 Member States. Numerous nations from the industrialized Global North, notably the USA, Australia,

[179] Id.
[180] WB: GovTech: Putting People First: Overview (website, *last accessed* June 8, 2024), https://www.worldbank.org/en/programs/govtech/priority-themes.
[181] *See* Asami Okahashi and Charles Blanco, "How is the World Bank using AI and Machine Learning for Better Governance?" (Mar. 07, 2024), World Bank Blogs, https://blogs.worldbank.org/en/governance/how-world-bank-using-ai-and-machine-learning-better-governance.
[182] UNGA, A/RES/2152 (XXI) of 17 Nov. 1966.
[183] United Nations and United Nations Industrial Development Organization: Relationship Agreement (Dec. 17, 1985) UN Treaty Series 1412, 305.
[184] UNIDO: About Us (website, *last accessed* June 7, 2024), https://www.unido.org/about-us/who-we-are.

Canada, France, the UK, and New Zealand, left UNIDO for ideological reasons in the 1990s. UNIDO survived this crisis, however, and has restructured itself informally as the "development agency for middle-income countries,"[185] seeking to develop South-South and Triangular Industrial Cooperations (SSTIC) as a "complementary pathway to traditional North-South development cooperation."[186] China has emerged as UNIDO's top donor, contributing over 16% of the overall 2020 budget.[187] As part of this effort, UNIDO has taken an active role as a partner in China's Belt and Road Initiative.[188,189]

[185] *See* Dan Runde, "China believes in UNIDO, why don't the rest of us?" (July 11, 2013) Devex (web blog, *last accessed* June 7, 2024), https://www.devex.com/news/china-believes-in-unido-why-don-t-the-rest-of-us-81424.

[186] UNIDO, "UNIDO charts new course in South-South and Triangular Industrial Cooperation" (Sept. 1, 2023), https://www.unido.org/news/unido-charts-new-course-south-south-and-triangular-industrial-cooperation.

[187] Max-Otto Baumann, Sebastian Haug and Silke Weinlich, "China's Expanding Engagement with the United nations Development Pillar: The Selective Long-term Approach of a Programme Country Superpower (Nov. 2022), German Institute of Development and Sustainability and the Friedrich Ebert Stiftung, 10, https://library.fes.de/pdf-files/iez/19692.pdf.

[188] UNIDO, "UNIDO's Director General visits China to bring cooperation to a new level (Oct. 20, 2023), https://www.unido.org/news/unidos-director-general-visits-china-bring-cooperation-new-level.

[189] Gerd Müller, Director General of UNIDO, Speech at the Third Belt and Road Forum for International Cooperation: Thematic Forum on Trade Connectivity (Sept. 18, 2023), https://www.unido.org/news/third-belt-and-road-forum-international-cooperation-thematic-forum-trade-connectivity ("With China and the [Belt and Road Initiative] as partner, UNIDO wants to do bold, innovative projects. Large-scale, flagship projects that set an example to the world and show how development can be accelerated, like the Chinese model. Together we can maximize the positive effects of the BRI in developing countries, reducing poverty and creating jobs.")

**Relevant Standards and Institutional Capacities:**

In July of 2023, UNIDO launched the Global Alliance for AI for Industry and Manufacturing (AIM-Global).[190] The stated objective of AIM-Global is no less than to develop a comprehensive and global AI Governance regime:

> "UNIDO's Global Alliance was conceived as an international platform to foster AI development in industry and manufacturing that would facilitate research, develop ethical guidelines, and make policy recommendations for AI use globally. AIM-Global will prioritize data protection, privacy, and trust in AI systems, advocating for improved mechanisms to safeguard sensitive data. These efforts will be reinforced by the establishment of AI Centers in different regions, as well as investments in education and training to empower individuals and organizations to leverage AI for innovation while upholding ethical standards."[191]

UNIDO's AIM-Global program promises to address the risk that "[w]ithout careful consideration and [a] strategic approach, these advancements [new technologies like AI, machine learning (ML), and intelligent automation] could widen the chasm between rich and poor nations, funneling investment into already advanced economies and leaving emerging ones behind."[192] The program seeks to "redirect the current trajectory of technological advancement to serve a more inclusive purpose."[193]

[190] UNIDO, "UNIDO launches Global Alliance on AI for Industry and Manufacturing (AIM-Global) at World AI Conference 2023," (website *last accessed* Jun. 8, 2024), https://www.unido.org/news/unido-launches-global-alliance-ai-industry-and-manufacturing-aim-global-world-ai-conference-2023.
[191] Mr. Ciyong Zou, Deputy to the UNIDO Director General and Managing Director, Keynote Speech marking the inauguration of AIM-Global, quoted in UNIDO, "UNIDO launches Global Alliance on AI for Industry and Manufacturing (AIM-Global) at World AI Conference 2023," (website, *last accessed* June 7, 2024), https://www.unido.org/news/unido-launches-global-alliance-ai-industry-and-manufacturing-aim-global-world-ai-conference-2023.
[192] UNIDO, AIM-Global: Who We Are (website, *last accessed* June 7, 2024), https://aim.unido.org/who-we-are/.
[193] Id.

UNIDO AIM-Global is a membership organization open to UNIDO Member States, academic institutions, corporations, and other "relevant stakeholders."[194] At the time of publication, membership included a number of firms, non-profit consulting firms, and corporations from the Global North and the Global South.[195] Members commit themselves to a set of five UNIDO AIM-Global Principles,[196] specifically:

| | |
|---|---|
| **(1)** | Transparency and accountability; |
| **(2)** | Inclusiveness; |
| **(3)** | Collaborating for Innovation; |
| **(4)** | Reliability and sustainability; and |
| **(5)** | Privacy & Security |

Noteworthy in these principles is the calculated *absence* of any mention of human rights, despite some seemingly very related concepts being included (for example a focus on "privacy," and "human dignity"). This is consistent with scholarly literature describing the influence of China in multilateral development forums, namely that "UN documents that originate in China [and] country programme documents of UN agencies operating in China, either do not contain words that reflect the UN's human rights framework or use them in ways that can counteract their intended meaning."[197]

[194] UNIDO, AIM-Global: Membership (website, *last accessed* June 7, 2024), https://aim.unido.org/membership/.
[195] Id.
[196] UNIDO, AIM-Global: Alliance Structure (website, *last accessed* June 7, 2024), https://aim.unido.org/alliance-structure/.
[197] Baumann, Haug and Weinlich, *supra* note 187, *at* 25.

## The Technical Lens

> "Before regulation, there needs to be agreement on what the dangers are, and that requires a deep understanding of what A.I. is."[198]

California (USA) Congressperson **Jay Obernolte**
*(the only member of the US Congress with a master's degree in AI)*

Alongside the various policy discussions about AI there are also numerous highly technical considerations that impact how AI and NETs should be regulated. Anyone wishing to engage in these aspects of AI and NETs must be conversant in the technical aspects of these technologies. The GDC makes prominent mention of the importance of developing relevant data standards that will allow transborder data governance, and much of this work will depend on the development of robust, safe, and interoperable technical standards.

Discussions in this domain often revolve around the specifics of what can (and cannot) be expected of these technologies. When a policymaker says, for example, that an algorithm should be "**transparent**," only a technologist can help clarify whether that is even possible in the context of an AI system operating in a "black box."[199] The same is true also for other discussions, for example the discussion about **AI 'sentience'** (when an AI can be said to have reached or surpassed human-level cognition), **alignment** (the challenge of ensuring that AI systems remain tethered to an underlying human-defined set of priorities), how to **'watermark'** AI-generated content, or how to ensure that an AI system not be based on **biased training data**. These ideals all sound straightforward from a policy makers' perspective but might not at all be easy to implement from an engineering or computer science perspective. Thus, a parallel and highly interconnected element of any global governance regime would be the effort to define the actual technical standards that can be realistically used to create common standards and regulatory frameworks globally.

There are currently only two main players at the UN that can play such a technical role: the International Telecommunications Union and the World Intellectual Property Organization. The subject matter of this discussion falls outside the scope of this paper, and so these two organizations and their technical capacities will be discussed only briefly.

---

[198] Congressman Jay Obernolte, quoted in Cecilia Kang and Adam Satariano, "As A.I. Booms, Lawmakers Struggle to Understand the Technology," NYTimes (Mar. 3, 2023) https://www.nytimes.com/2023/03/03/technology/artificial-intelligence-regulation-congress.html.

[199] *See, e.g.,* Steven Levy, "AI is in a black box. Anthropic figured out a way to look inside." Wired (May 21, 2024) https://www.wired.com/story/anthropic-black-box-ai-research-neurons-features/.

## International Telecommunications Union (ITU)

While the ITU's AI for Good initiative focuses on the use of AI in furtherance of the SDGs, the ITU standardization sector (ITU-T) has convened focus groups, each tasked with developing concrete technical specifications, to focus on (1) machine learning for future networks including 5G, (2) AI for Health (run in collaboration with the World Health Organization), (3)

Environmental Efficiency for AI and other Emerging Technologies, (4) AI for Autonomous and Assisted Driving, and (5) AI and Data Commons.[200]

## World Intellectual Property Organization (WIPO)

First established in 1893 and then eventually incorporated into the UN as a Specialized Agency in 1974,[201] WIPO's mandate is "to lead the development of a balanced and effective international IP system that enables innovation and creativity for the benefit of all."[202] WIPO is primarily self-funded and derives much of its revenue from its fee-based arbitration and mediation services. AI and other NETs stand to significantly affect WIPO's mandate of protecting intellectual property.[203]

WIPO began to focus on AI and NETs in 2019 with the publication of a report on AI and its impacts on the IP system.[204] Currently, WIPO's primary organizational response to the advent of AI is to convene a semi-annual multistakeholder dialogue called the **WIPO Conversation on IP and Frontier Technologies** in which these issues are discussed. In coming years one could imagine some of these discussions factoring into the standard-setting work of the WIPO Committee on WIPO Standards (CWS), as it has in response to the advent of a different NEDT (blockchain) following a similar series of discussions that began in 2018.

---

[200] ITU and World Bank, Digital Regulation Platform (website, *last accessed* June 7, 2024), https://digitalregulation.org/the-work-the-itu-standardization-sector-itu-t-on-artificial-intelligence/.

[201] Agreement concerning the relations between the two organizations. Approved by the General Assembly of the World Intellectual Property Organization on 27 September 1974, and by the General Assembly of the United Nations on 17 December 1974 (Dec. 17, 1974) UN Treaty Series 956, 405.

[202] WIPO: inside WIPO (website, *last accessed* June 8, 2024), https://www.wipo.int/about-wipo/en/.

[203] WIPO IP and Frontier Technologies Factsheet, https://www.wipo.int/about-ip/en/frontier_technologies/. (highlighting a number of questions about IP and NETs, including whether AI can invent and what AI might mean for human inventions; how best to foster access to data for machine learning and AI?
What role IP plays in balancing access to data, fostering generation of data, and protecting legitimate interests; whether the current law of trade secrets strikes the right balance between protecting innovations in the AI field and the legitimate interests of third parties in having access to certain data and algorithms; whether frontier technologies assist innovators and creators everywhere to benefit from IP; and what policy measures could contribute to the reduction in the technology gap and help frontier technologies solve pressing global issues?)

[204] WIPO (2019). WIPO Technology Trends 2019: Artificial Intelligence. Geneva: World Intellectual Property Organization, https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf.

## The National Security Lens

The final disciplinary lens through which AI and other NETs can be evaluated is that of national security. Indeed, this is one of the most prominent ways many commentators speculate about AI's potential future impacts. Dystopian scenarios of autonomous weapons, the intensification of conflict through AI-driven misinformation campaigns and malicious cyber-operations, doomsday scenarios of human extinction caused by AI, and other threats to peace, security, and global stability dominate headlines (and science fiction movies) about AI. The national security lens is notably not frequently addressed in any of the standards described above. National security considerations tend to be reserved for national governments to consider, often to the consternation of some human rights activists who consider the deployment of AI and other NETs for national security purposes to represent one of the most prominent potential human rights risks of these technologies.

### United Nations Security Council (UNSC)

The UN Security Council is a permanent body within the United Nations tasked with managing international peace and security issues. In July of 2023, it convened a **High-Level Briefing on AI**, in which Council Members discussed AI and its implications for global security.[205] Briefings stressed both AI's positive and negative potential to impact global peace and security. On the tech-optimist side, presenters focused, for example, on AI's ability to support peace mediators and predict outbreaks of violence, whereas others focused on the potential use of AI technologies to produce misinformation, facilitate cyberattacks, and fuel new forms of terrorist attacks. UN Secretary António Guterres also stressed the need for a comprehensive ban on the use of AI in automated weapons of war.

Article 12 of the UN Charter grants the UN Security Council the power to quash any ongoing discussion about an issue at the General Assembly. The Security Council also has the power to take measures, including those requiring the use of armed force, to maintain or restore international peace and security. The implications of these provisions are clear—in the event that AI or NETs would ever emerge as a tangible threat to international peace and security, and in the event the Security Council membership should find consensus on an adequate response ot that threat, there should be no doubt as to the Security Council's capacity to assume control over any other efforts to govern AI that may have emerged from any of the Agencies or UN Institutions described above. Should this

[205] UN Meetings Coverage and Press Releases, Security Council, 9381st Meeting (SC/15359), July 18, 2023, "International Community Must Urgently Confront New Reality of Generative, Artificial Intelligence, Speakers Stress as Security Council Debates Risks, Rewards," https://press.un.org/en/2023/sc15359.doc.htm.

occur, a national security disciplinary lens would most likely re-emerge as the ultimate 'lodestar' guiding such deliberations.

However, there is a caveat to the point above. Many analysts have noted the inability of the UN Security Council to decide on contemporary security issues, especially if they involve the key interests of the Council's Five Permanent Members (P5). When it comes to AI, there are well-known policy divergences between the regulatory approaches taken by China, the United States, Europe, and the United Kingdom. These differences pertain primarily to *non-military* use cases for AI and other NETSs. Regarding the *military* use of NETs, it seems even more clear that most nations, including the P5, are still absolutely unwilling to subject their use of AI to any international governance mechanisms. The chances of clear substantive guidance from the Security Council relating to the global governance of AI are, therefore, exceedingly slim, except in the hypothetical scenario of an existential security threat that *transcends* the hardened security interests of the Council's P5 Membership.

Given these realities, the various multistakeholder efforts described above, under the auspices of the UN General Assembly, are likely to continue. **If progress is to be made at the UN level on the global governance of AI, it will most likely occur under the auspices of the UN General Assembly and its subordinate bodies.**

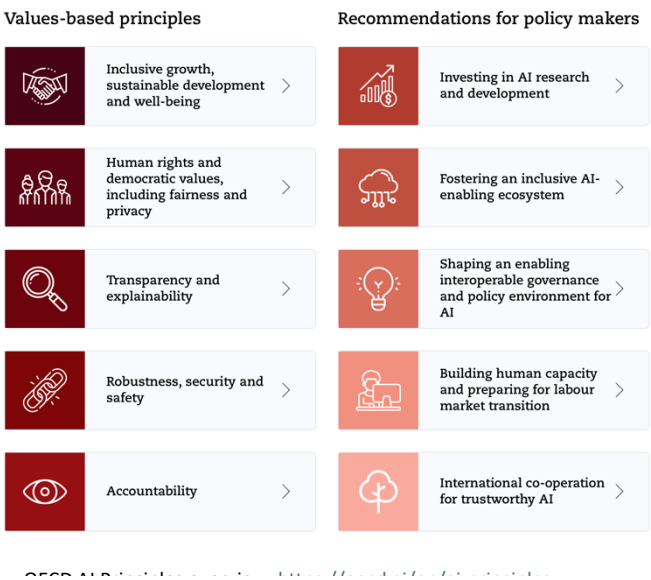
## Standards and Institutional Capacity Building Efforts taking place outside of the United Nations System

The UN is not the only entity concerned about the international governance of AI and other NETs. Several other organizations have also taken a leadership role in this discussion. While none of them have the same institutional reach as the UN in terms of mandate or membership, the UN would likely need to be in constant dialogue with these organizations, adopting an approach to governance that builds on (and does not compete with) these other approaches.

The specifics of these various regional efforts go beyond the scope of this paper and will be only briefly discussed here.

### OECD & G20

The OECD brings together 38 member states, mostly from the industrialized 'Global North.' The G20 brings together the 20 largest economies, the European Union and the African Union. In May 2019, the OECD developed the **OECD AI Principles** to promote innovative and trustworthy uses for AI, built around respect for human rights and democratic values. The OECD AI Principles consist of 5 Values-based principles and 5 recommendations for policymakers seeking to regulate AI.[206]




Updated recently in May 2024, the OECD AI Principles represent the first intergovernmental standard on AI. They have since been used by the G-20 as the basis for its own AI principles, which it endorsed in June 2019. The OECD maintains an OECD AI Policy Observatory which hosts a range of resources related to the gradual implementation of the OECD AI Principles, including data from OECD (and non-OECD) member nations comparing their AI-relevant policy environments.

### Regional Bodies

Numerous regional organizations have also taken a proactive approach to the supranational governance of AI. The European Union's efforts to regulate AI are frequently discussed in this context, but by no means alone.

#### African Union (AU)

In February 2024, the African Union Development Agency (AUDA-NEPAD) released an **AI Continental Roadmap for Africa**.[207] Work on this roadmap began in 2016. The document

[206] OECD.AI Policy Observatory, "OECD AI Principles overview," (website, *last accessed* June 8, 2024), https://oecd.ai/en/ai-principles.
[207] AUDA-NEPAD, Artificial Intelligence Continental Roadmap for Africa, Feb. 2024, Johannesburg, South Africa, www.nepad.org.

proposes the "necessary measures that African countries should adopt to potentially facilitate inclusive and sustainable AI-enabled socioeconomic transformation,"[208] breaking those efforts down into six strategic priorities: (1) the development of human capital for AI; (2) the establishment and use of infrastructure and data to fuel AI-driven economic growth; (3) the creation of an enabling environment for AI-driven innovation; (4) establishing a conducive economic climate for AI; (5) building sustainable partnerships, and (6) monitoring and evaluation.

**Association of Southeast Asian Nations (ASEAN)**

In 2024, ASEAN published its **Guide on AI Governance and Ethics**.[209] This publication proposes seven guiding principles for the use of AI.[210] It also proposes a voluntary AI governance framework that organizations can use as they implement AI systems in commercial, non-military, or dual-use scenarios. It also contains case studies and best-practice recommendations for national governments and regional organizations.

**Council of Europe (CoE)**

On May 17, 2024, the CoE finalized the **Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law**, the first-ever binding treaty on the use of AI. This treaty opened for signature during the CoE Conference of Ministers of Justice in September 2024 and will enter into force on January 1, 2025.[211] (Article 30.3). Non-CoE member states such as the United States and Israel are also eligible to sign the treaty, and have already done so.[212] The object of this treaty is "to ensure that activities within the lifecycle of artificial intelligence systems are fully consistent with human rights, democracy and the rule of law."[213]

[208] Id., *at* 13.
[209] ASEAN, ASEAN Guide on AI Governance and Ethics (2024), https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf.
[210] (1) transparency and explainability; (2) fairness and equity; (3) security and safety; (4) human-centricity; (5) privacy and data-governance; (6) accountability and integrity; and (7) robustness and reliability.
[211] At least three of those ratifying states must be members of the Council of Europe.
[212] Council of Europe: Committee on Artificial Intelligence (CAI) (website, *last accessed* June 8, 2024), https://www.coe.int/en/web/artificial-intelligence/cai.
[213] Article 1.1, Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, [Vilnius, 5.IX.2024].

Article 3.3 of the Convention makes it clear that this convention does not apply to research insofar as those activities do not "have the potential to interfere with human rights, democracy and the rule of law," and Articles 3.2 and 3.4 make it clear that it also does not apply to AI applications related to national defense and national security.

The Convention begins with the general principles that AI systems are not to be used to interfere with or undermine human rights, democratic processes, or the rule of law (Articles 4 & 5). The treaty commits member states to implement seven common principles into their domestic AI regulatory framework, consistent with the form and methods of their domestic legal system:

| | |
|---|---|
| **(1)** | Human dignity and individual autonomy; |
| **(2)** | Transparency and oversight; |
| **(3)** | Accountability and responsibility; |
| **(4)** | Equality and non-discrimination; |
| **(5)** | Privacy and personal data protection; |
| **(6)** | Reliability; and |
| **(7)** | Safe innovation. |

States under the Convention are obligated to ensure the availability of remedies to contest decisions made by AI systems and that users and judicial authorities have access to the requisite data to know whether they have been treated fairly by an AI system (Chapter IV). Parties are also required to conduct context-specific due diligence to mitigate potential adverse impacts of AI systems on human rights, democracy, and the rule of law (Article 16.3). Mitigation measures must include, if necessary, the option of banning certain AI systems if their use is deemed incompatible with human rights, democratic governance, and the rule of law (Article 16.4).

The Convention also calls for the creation of a Conference of Parties, which will serve as the guardian of the Convention, overseeing and facilitating its implementation and—when necessary—revisiting certain aspects of the Convention based on unforeseen developments. The eventual entry into force of this Convention, should it happen in late

2024, will represent the first *binding* and global treaty to articulate State obligations with regard to AI governance.

**European Union (EU)**

The EU has taken a proactive role vis-à-vis artificial intelligence. In addition to various measures to promote investment in AI and create a favorable enabling environment within the EU for AI enterprises to flourish, the EU has also developed the first binding regional standard for AI. The Council of the EU gave final approval to the **EU AI Act** in May 2024, and analysts expect it to enter into force across the entirety of the EU by mid-June of 2024.

The structure and methodology of the EU AI act remains a hotly debated topic, but analysts also expect it to set a benchmark against which other regulatory approaches, should they choose an alternative model, will have to define themselves. The EU's model is based on a **risk-based model of gradated regulation**. AI technologies are to be sorted according to their level of risk. AI models posing an *unacceptable risk* are to be categorically banned. This includes AI-fueled social scoring systems or systems designed intentionally to distort a person's behavior in a way that would cause the person harm. *High-risk* AI systems include those relating to health and safety or those concerning education, employment, access to essential public and private services, law enforcement, migration, and the administration of justice. High-risk systems are to be registered in a centralized database and are subject to the highest level of regulatory scrutiny before they can be deployed in the EU market. *Limited-risk* AI applications, which would include most of the Chatbots, image generators, or other public-facing AI applications we have grown used to since the launch of ChatGPT in 2022, would be subject primarily to transparency requirements. Any other uses of AI would be categorized as *low or minimal risk* AI use cases and therefore would not be subject to any regulatory burdens.

The provisions of the EU AI Act are to be enforced at the national level, with high-level coordination across the EU by an AI Office within the EU Commission. The Act also calls for significant penalties for willful noncompliance with the provisions of the Act or the act of willfully supplying misleading information relating to the risk assessment for a particular AI use case. Detractors of the EU's approach worry that these requirements will stifle innovation and competition in the European Union, whereas the proponents of the Act hope that it will create the world's first economic zone where consumers can be confident that they will be exposed to only safe and trustworthy AI applications.

**Organization of American States (OAS)**

So far, the OAS is the only regional organization that has remained largely unconcerned with a regional approach to AI governance.[214]

In addition to regulations at the regional level, there is also a boom of regulations at the national[215] and corporate levels. In addition, numerous industry or civil-society-led initiatives[216] are emerging intending to regulate AI applications. A survey of these mechanisms goes beyond the scope of this paper.

# Analysis of Existing Efforts to Work Towards a Global AI Governance Regime

The discussion above shows that **there is no universally accepted institutional "lead organization" in the discussion over AI Governance**. There are numerous institutional actors, all of them with a unique mandate to engage in this multifaceted discussion. This naturally results in a plurality of voices speaking in parallel to one another. Of course, the perennial scramble for funding inherent to any UN Agency or body contribute to this plurality of voices. Artificial Intelligence is currently considered to be an extremely "hot" topic, likely to attract funding and interest from a range of sources, including private sector sources keen to claim a seat at the table for these discussions.

Existing efforts to coordinate AI Governance at the UN level have been piecemeal, to say the least. UNESCO, the ITU, UNIDO, UNDP, the World Bank Group, UNDP, and the OHCHR have all, in one way or another, claimed for themselves some leadership role regarding global AI governance. None of these actors are acting in bad faith, and most are articulating models of

[214] OAS: Speeches and other Documents by the Secretary General. Speeches and other documents by the Secretary General, High-Level Roundtable Policy Dialogue "Artificial Intelligence: Public Policy Imperatives for the Americas" Draft Opening Remarks, May 4, 2023, https://www.oas.org/en/about/speech_secretary_general.asp?sCodigo=23-0015; *and* Busola.tech and OAS and Bússola Tech (2024) Tecnologías Generativas y el Ciberespacio en la Agenda Legislativa 2024, Madrid, España, https://news.bussola-tech.co/generative-tech-and-legislative.

[215] For good reviews of national AI policies, see the Repository of over 1000 AI Policy Initiatives from 69 countries found in the OECD's AI Policy Observatory (https://oecd.ai/en/dashboards/overview); The UN Institute for Disarmament Research Artificial Intelligence Policy Portal (https://www.caidp.org/reports/aidv-2023/), and the non-profit Center for AI and Digital Policy's annual Artificial Intelligence and Democratic Values report (https://www.caidp.org/reports/aidv-2023/).

[216] *See e.g.,* < encode justice >, AI2030, https://ai2030.encodejustice.org.

AI governance that overlap substantially with each other. Almost all (with the possible exception of UNIDO's efforts) claim that human rights must be the central disciplinary "lens" (lodestar) for any governance regime, and all also justify the need for an urgent embrace of AI's potential in terms of the fulfillment of the SDGs. A great deal of consensus exists about the different approaches to AI governance, even if each model emerged from largely isolated consultation processes and from a range of different institutional mandates. In short, the plurality of self-purported 'leaders' in this discussion has little to do with substantive disagreements, and more to do with institutional momentum lacking centralized coordination.

Nonetheless, many of these discussions still tend to be dominated by voices from the "Global North," specifically, from voices in North America and Europe. This is especially true in the corporate space, where the loudest voices often belong to those corporations with the resources to consistently engage in the numerous and rapidly proliferating discussion forums – in Geneva, New York, Vienna, Washington, London, and Seoul (among many other locations) – that are so often meeting in parallel to one another. However, it is also true for overstretched diplomatic representatives from countries in the Global South and UN Agencies that depend heavily on consultations in their corporate headquarters.

Most of the frameworks that have emerged, save for the ones emerging from the European Union and the Council of Europe, are still largely premised on voluntary and unenforceable commitments, and many also still rely on abstract normative principles, with scarce guidance on how to translate those into practice. Finally, few regulatory frameworks offer anything but a *rhetorical* commitment to the *promise* of AI. The first obvious exception to that observation is corporate ethics frameworks, all of which are designed to not constrain that corporation's ability to innovate. The second is efforts, like the AI for Good initiative driven by the ITU, to highlight the positive potential of AI without much thought on how to regulate those same technologies. Merging those two prongs will be one of the great challenges of an integrated global governance approach towards AI.

**Now that the UN has firmly established its intention to develop a global governance regime for AI and other NETs, the question is not who would like to take ownership over that governance regime or even who has already demonstrated leadership in this field, but rather how, from an institutional design perspective, a coordinated and coherent governance function would be best structured.**

The UN Secretary General and the General Assembly have recently re-assumed for themselves a leadership role in this discussion. In March of 2024, the UNGA adopted a resolution on AI that sketched out an UN-wide institutional approach to AI and other NETs. Even before that resolution, the UN Secretary General in 2023 had already convened a High-Level Advisory Group of globally renowned experts to advise him on a framework approach to AI, which is scheduled to be made public in September of 2024. This will be the moment when the UN can finally cut through the administrative and bureaucratic cacophony to articulate a clear vision of how it will approach the global governance of AI and other NETs.

# The Case for Investing in the Capabilities of the Human Rights Council

This paper makes a case for strengthening the capabilities of the Human Rights Council as a central player in the ongoing discussions about the global governance of AI and other NETs. The argument centers on five key arguments, each drawn from the analysis above. **The HRC can play an essential coordination role without the need to first assemble a massive new institutional machinery, ensuring that human rights remain the 'lodestar' of discussions over how best to govern NETs, while also empowering the many other actors of the UN system to do what they do best in support of that overall effort.**

## Substantive Anchoring to Human Rights

The above analysis has demonstrated a near-unanimous consensus that human rights must lie at the heart of any global efforts to govern the use of AI and other NETs. No other UN institution is as uniquely situated to guarantee the centrality of the human rights discourse as the HRC. This is by design: since its creation in 2006, the HRC has been mandated to serve as the central coordinating body tasked with overseeing and orchestrating all efforts across the United Nations to advance the global human rights agenda, answerable only to the UN General Assembly in that effort. This includes, by design, the mandate to advance the right to development.[217]

Crucially, the HRC has been designed to serve this role *in collaboration with* (not "in the place of…") all other relevant actors at the UN with a human rights mandate. As a result of past efforts to mainstream human rights throughout the UN, almost all UN Agencies or Programmes have human rights embedded centrally into their organizational mandates. However, above all of those mandates hovers the continued need for ongoing dialogue about emerging human rights issues, cross-disciplinary coordination, civil society engagement, compromise, joint strategizing, and diplomatic reality testing. That role can only be played by

[217] UNGA Res 60-251, *supra*, note 90, Article 4.

the HRC, which has been specifically designed with these purposes in mind. That role is also crucial in this current effort to develop a global and human rights-based approach to the governance of AI and other NETs.

## The Gradual Elaboration of Human Rights in the Digital Space:

Our social, political, economic, educational, and cultural realities are rapidly shifting from the real world into digital spaces. This shift from a "1.0 reality" towards a "2.0 reality"—or at least a hybridized reality somewhere in between—has not yet been accompanied by a real interrogation of how existing human rights, all of which were developed with a firm "1.0 reality" as the backdrop, apply in a new digital "2.0 reality."

Take, for example, the quintessential "1.0 reality" right not to be "subjected to torture or to cruel, inhuman or degrading treatment or punishment" (Article 5 of the Universal Declaration of Human Rights) or the right to "recognition everywhere as a person before the law." (Article 6). What do those rights mean in the context of a 2.0 reality? If torture, cruel, inhuman, or degrading treatment is equated only with *physical* abuse that can take place in the "1.0 world," then what protects us from similar forms of abuse that might take place online? And if the right to recognition before the law is equated with notions of "jurisdiction" as defined by judges, lawyers, courts and borders in the "1.0 world," then what are we left with when we have a grievance to air in the "2.0 world?" The ongoing conversation about human rights needs to pay more attention to the shape and texture of existing human rights and freedoms in the digital space.

This "translation" also needs to occur in the context of discussions over AI and other NETs in our digital lives. What does it mean, for example, to have a "right to due process" when our grievances, job applications, or requests for social security services are mediated by automated systems? What would a right to dignified work resemble in a world where traditional jobs can increasingly be replaced by automated systems? And does the right to freedom of expression and opinion in today's world necessitate a consequential "right to the internet?" The provisions in the Council of Europe's Framework Convention (see above, p. 70) introduces the right to contest algorithmic decisions, offering an excellent example of such a "translation" in practice.

The HRC is ideally structured to gradually explore such questions. Special Procedures at the HRC are well known for their ability to quietly, thoughtfully, and methodically explore such questions and then bring their emerging ideas before the HRC for discussion and consensus building. Numerous special mandate holders have been tasked with similar tasks before. A Rapporteur or Working Group on NETs could do the same for human rights in the digital

space. Such a mandate would not need to come at the expense of other existing mandates, but it could absolutely coordinate and interact with them constructively and collegially to explore this philosophically challenging question, building on some of the groundbreaking work already done by existing mandate holders (see above, p.51).

We mention three examples of such translations in the following: the right to the internet, the right to data privacy, and the right to digital literacy. All of these 'rights' can be derived directly from existing human rights.

The **"right to the internet"** is arguably a derivative of the right to freedom of expression and opinion. Given the infrastructure required to give meaning to this right, however, one can imagine discussions of a digital divide impeding the immediate realization of this right for many parts of the world. One might, therefore, imagine this "right" (if it existed) to be subject to some form of a progressive realization clause, in stark contrast to the more immediate state obligations inherent in most civil and political human rights. Other questions are also inherent in the idea of a human "right" to the internet. Can the Internet be a platform for citizens to carry out their civic duties, and would a "right" to the Internet also necessitate access to the *entire* Internet? Would a right to the internet also entail a right to access and use NETs? And how would questions of jurisdiction be resolved in the context of such a "right to the internet?" These are exceedingly complex questions, and a dedicated special mechanism would be the ideal vehicle for exploring them further.

The **right to privacy** is a fundamental human right, but what does that right look like in a digital space, where data that in the "real world" might be considered confidential is being "read" and interpreted by AI-enabled algorithms all the time? Does privacy only apply to human "readers," or are machine readers of confidential or private information also barred from scrutinizing such information? To answer this question competently and in line with human rights norms, one must engage in thorough philosophical, technical, and legal discussions – a task that would lend itself perfectly to a special procedure.

Finally, the **"right to digital literacy"** might be understood as a subordinate right to the existing right to education, which is traditionally understood to be an Economic, Social, and Cultural right, and thus subject only to the requirement that States work towards its gradual and progressive realization in keeping with their capabilities to do so. But digital literacy is increasingly a prerequisite for exercising of many other rights in the digital space, for example, the right to take part in the conduct of public affairs, especially if those move into the digital space. The right to digital literacy opens the doors towards becoming a "global citizen" in both

the physical and digital worlds. Do human rights in the digital era, therefore, require that countries investing in AI-enhanced public service delivery methods simultaneously waive their right to only "progressively realize" the right to digital literacy?

To keep up with the proliferating technological reality, the HRC—with the help of a specialized procedure tasked with exploring these knotty issues—must continuously interrogate itself as to how existing rights translate into novel digital spaces of human activity.

## Procedural Legitimacy

The HRC is one of only a few UN deliberative bodies. By design, it brings together representatives from various regions of the world to periodically debate human rights issues. It also has the power to involve civil society, other UN Agencies, and the private industry in these discussions. Much more so than other forums, the HRC can claim to be a genuinely consultative body, mandated to serve as a place of dialogue, consensus building, and critical interrogation. This is precisely the kind of forum that is most needed to host ongoing discussions about the interrelationship between AI and human rights protections and produce the kind of "**thick stakeholder consensus**" that John Ruggie also described as essential for the unanimous endorsement of the UNGPs in 2011.

The HRC is also not a bureaucracy. It is supported and facilitated in its work by the OHCHR, which serves—in some ways—as the administrative apparatus giving life to the HRC's decisions. The HRC cannot fall victim to institutional capture, whereby an established bureaucracy becomes institutionally entrenched in a certain way of thinking or doing things. This capability to be free from unnecessary talk of "how we've always done things…" is essential for any organ engaging in the highly dynamic and constantly evolving world of AI and NETs.

## Existing Capacity

The HRC already has the institutional features to play the coordination and norm elaboration roles described above. The HRC is—at heart—a plenary forum. This forum is ideally suited as a high-level discussion forum. The HRC also has the power to create special mechanisms that – as discussed above – are ideally suited to elaborate on how existing human rights apply in novel contexts. The HRC has the mandate to gradually elaborate upon emerging issues of human rights law and—if and when relevant—to bring those to the attention of the General Assembly for further deliberation and possible formalization. The HRC's universal periodic review process also provides a universal process of constructive engagement in which national authorities can be directly engaged about their policies on AI and other NETs. The HRC even has a built-in complaints mechanism, and individual special mandate holders could

easily be tasked with a constructive problem-solving role in response to individual complaints they receive relating to their mandate.

In short, no new institutions or bureaucratic structures would need to be created for the HRC to simultaneously fill a number of gaps in the UN's current approach to the global governance of AI.

## Catalytic Synergies

Finally, and perhaps most importantly, the HRC as an institution is structured as a "forum for dialogue on thematic issues on all human rights."[218] The HRC is a specialized plenary assembly, subordinate only to the General Assembly. By design, it is a multi-stakeholder forum with the ability and mandate to orchestrate the activities of other UN Offices, Programmes, Agencies, and Specialized Agencies on matters relating to human rights. The HRC can work seamlessly and without contradiction with the OHCHR, UNESCO, the ILO, the World Bank, UNDP, and the ITU to weave a common strand of human rights-based thinking throughout each of these actors' respective actions regarding AI. This role, as discussed in the next and final section of this paper, is crucial in a polycentric world of global governance.

[218] Id., Article 5.b.

## Discussion paper 3-2

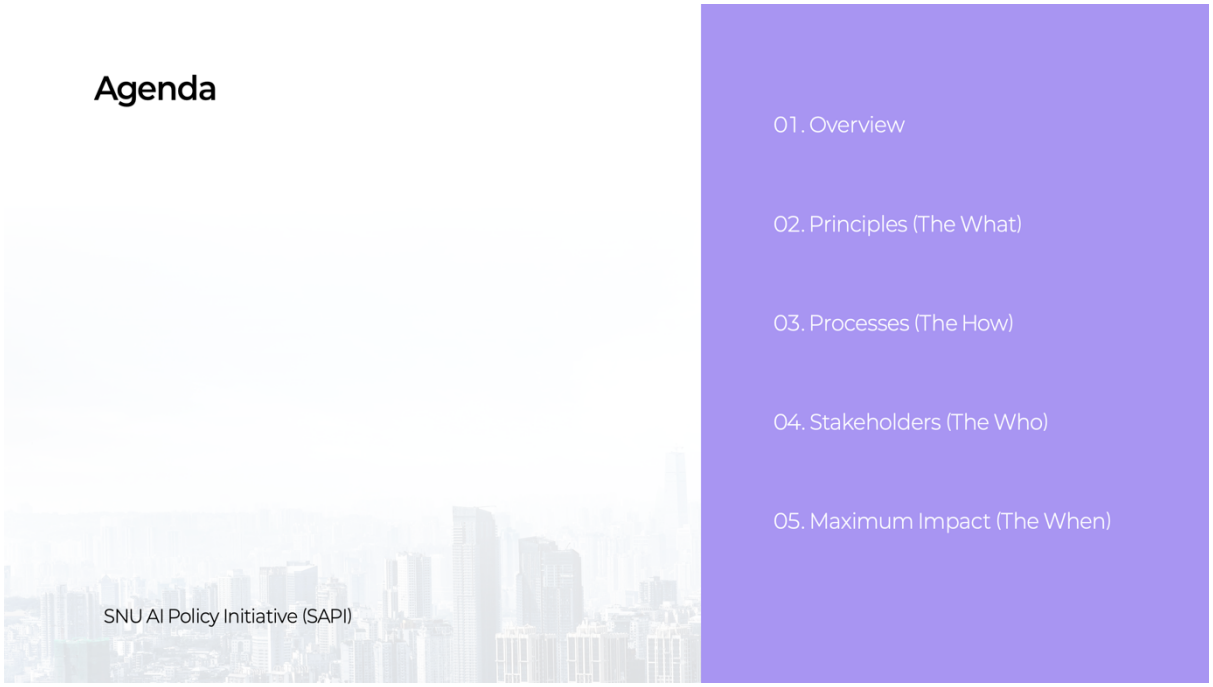
# The HRBA@Tech diagnostic tool and management consultant's toolbox: an introduction to the approach

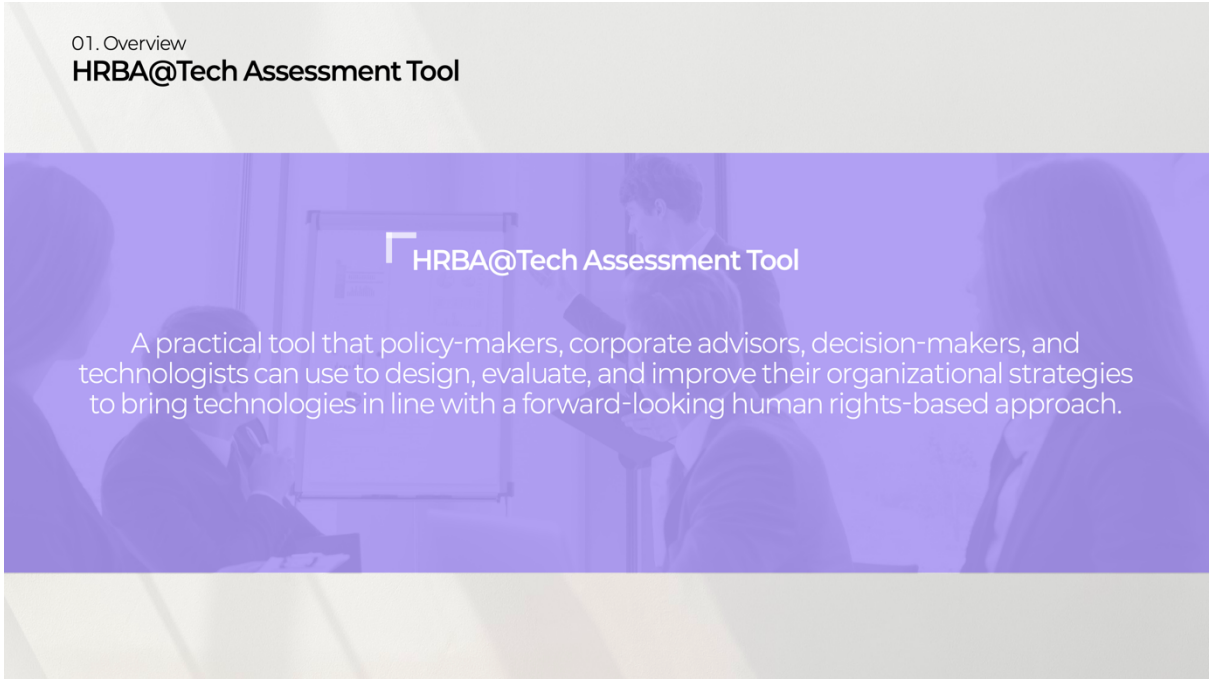
This paper was authored independently by the Seoul National University Artificial Intelligence Policy Initiative (SAPI) as part of an ongoing collaboration between SAPI, the Universal Rights Group (URG), and the Permanent Mission of the Republic of Korea to Geneva. It was presented for discussion in Geneva on December 5, 2024.


# Table of Contents

# Introduction

The **HRBA@Tech Consulting Manual** is a comprehensive and collaborative engagement methodology developed by the Seoul National University Artificial Intelligence Policy Initiative (SAPI) to help organizations and consultants work together to ensure that new and emerging technologies (NETs) are designed and deployed in line with human rights principles. This tool is grounded in a Human Rights-Based Approach (HRBA) and is valuable for consulting in contexts where technology intersects with public accountability, user safety, and ethical impact.






The tool was designed to reflect the Human Rights Based Approach to New and Emerging Technologies Model[219] (HRBA@Tech model) that SAPI produced in collaboration with the Universal Rights Group (URG), a Geneva-based human rights think tank, in 2022. The HRBA@Tech model was informed and subsequently enriched by extensive consultations with technology firms, academics, ethicists, and regulatory agencies at the national and international level.[220] The intent was to create an actionable framework for a host of different stakeholders, *including* but not limited[221] to tech corporations, to 'nudge' NETs in ways that would make their impact more consistent with human rights objectives. It is hoped that the HRBA@Tech model will contribute substantively towards the future work at the Human Rights Council through appropriate mechanisms.

Five key realizations differentiate the HRBA@Tech model from other existing human rights frameworks, including existing human rights tools as well as the UN Guiding Principles for Business and Human Rights.

1. **Capturing the *Upsides* of NETs**

   The HRBA@Tech model assumes that the development and deployment of NETs can undermine human rights protections, but that those same NETs can also serve to advance the cause of human rights. This dualistic nature of technology, coupled with

[219] Perm. Mission of the Republic of Korea to Geneva, SNU AI Policy Initiative, and Universal Rights Group. (2022) Towards a Human Rights-Based Approach to New and Emerging Technologies.
[220] Perm. Mission of the Republic of Korea to Geneva, SNU AI Policy Initiative, and Universal Rights Group. (2023) HRBA@Tech: AI Tech Startups, Climate Change, and Global Normative Governance.
[221] The HRBA@Tech model identifies six relevant categories of stakeholders for whom this tools can provide actionable guidance, namely (1) national governments, (2) private enterprises, (3) individual actors, (4) non-profit educational institutions, (5) civil society organizations, and (6) international organizations.

the impossibility of designing our way out of this paradox, requires human rights actors to adopt a nuanced and at-times contradictory attitude towards the emergence of NETs, guarding against their potential human rights *downsides* while at the same time also partnering with technologists and policy makers to maximize the potentially transformative *upsides* of those same technologies.

Most existing human rights frameworks tend to focus only on guarding against the potential downsides of NETs (a risk management, or "do-no-harm" approach). Some rhetorically acknowledge the transformative potential of NETs to promote human rights but fail to offer any guidance on how to maximize that impact. The HRBA@Tech model seeks to fill this void by offering normatively-grounded guidance to human rights actors seeking to capture the potential human rights upsides of NETs ("making the world a better place").

**2. Emphasis on *Processes* in addition to Principles and/or Standards**

Most frameworks to promote "ethical", "reliable", "trustworthy", "safe", or "human centric", or "human rights based" technologies rely either on abstract overarching principles or highly detailed technical standards in pursuit of those outcomes. The HRBA@Tech model also recognizes the importance of both principles[222] and standards[223] as part of a comprehensive human rights-based approach to NETs.

The key recognition of the HRBA@Tech model, however, is that principles and standards *alone* provide very little actionable guidance to the decisionmakers tasked with ensuring the rights-compatibility of a particular technological product. Principles, by necessity of having to be universal in scope, are often worded using vague and aspirational language. The interpretation and application of these principles is left to individual decisionmakers, including bad-faith actors intent on pursuing their individual advantage or profits over a commitment to principle. Standards, which can provide more clarity through specific benchmarks and technical requirements, are more suited to a risk management-based approach and thus can still be inadequate in capturing the much needed upsides of NETs in further human rights. Also, in practice this can discourage innovative solutions on and replace a good faith commitment to aspirational human rights principles with the box-ticking drudgery of implementing specific standards, some of which may not even be tied quantifiably to the advancement of those overarching human rights principles.

[222] The HRBA@Tech model proposes its own set of 7 overarching principles that should hold true for any human rights-based approach to the design and deployment of NETs. Subordinate to the model's "do no harm" objective, it posits that any efforts to develop NETs should be guided by a commitment to (1) legality, (2) non-discrimination and equality, (3) safety, and (4) accountability and access to remedies. In line with its objective to also harness technologies' potential to "make the world a better place", the model also dictates that actors should be committed to the principles of (5) empowerment, (6) transparency, and (7) participation.
[223] Under the principle of "safety," the HRBA@Tech model highlights the engagement with standardization processes as one essential element of a comprehensive human rights-based approach to NETs.

The HRBA@Tech model is built around an embrace of concrete processes that various stakeholders can harness in pursuit of more trustworthy and human rights-based technologies. This focus marries the visionary power of aspirational principles with the concreteness of standardization.

### 3. Expanding the discourse to *Other Stakeholders* - moving beyond the relationship between a State and the Individual Rights Holder

Traditional human rights tools focus on the relationship between a State and an Individual as the primary forum where human rights are asserted and vindicated. Other stakeholders, for example private corporations or the interests of "mother nature", are typically molded into this overarching State-Individual dyad, for example by the creation of legal fictions (*e.g.*, "legal personhood") and by an almost exclusive reliance on the State to regulate entities within its jurisdiction.

The HRBA@Tech model does not propose a complete rupture with the centrality of the State-Individual dyad. It does, however, propose that specifically four other categories of stakeholders deserve to have their unique roles in the design and deployment of NETs highlighted:

**International Organizations:**

Traditional human rights theory sees international organizations such as the UN as mere conglomerations of States working in concert with one another. But it is also undeniable that they have their own powerful role to play when promoting and spreading universal norms, such as human rights frameworks that apply to NETs. This idea is core to the HRBA@Tech model, which is itself being promoted as part of a push to develop *global* governance standards for AI and other NETs.

**Civil Society Actors:**

Traditional human rights theory sees civil society as conglomerations of individuals working in concert with one another to assert their human rights (a mirror image of international organizations just with regard to individuals instead of States). But here, too, one must acknowledge that civil society activism is both cause and effect of individuals understanding and asserting their rights, not to mention a powerful vehicle for individuals to find a voice that they might not otherwise have against more powerful state or corporate interests.

**Private Corporations:**

Like civil society, traditional human rights doctrine considers private corporations to be mere amalgamations of individuals acting in concert with one another. This construct ignores the various legal, governance, financial, and market-based structural constraints that most corporations must operate within. Furthermore, by equating corporations with individuals in the State-Individual dyad, traditional human rights doctrine obfuscates the empirical reality that corporations are often a primary and direct threat to the human rights of individuals and communities in which they operate, especially in contexts where the State is either unable or unwilling to regulate corporate entities within its jurisdiction. Moreover, in more modern times, traditional human rights doctrine has only inadequate answers on how to hold transnational corporations accountable. The HRBA@Tech model, like the UN Guiding Principles on Business and Human Rights upon which the model builds in this regard, speaks directly to the powerful role that the private sector undoubtedly plays when it comes to the development and deployment of NETs.

**Universities and other Educational Institutions:**

Universities and other educational institutions also deserve to be highlighted as a standalone category of stakeholders when it comes to NETs. This holds *especially* true when those institutions are in their essence non-for-profit institutions, and when they are guided by an overarching commitment to academic freedom and evidence-based research. Such institutions have played, and will continue to play, a major role in the development of NETs and the governance structure that surrounds them, and yet they do so without the various structural factors influencing corporate behavior mentioned above. In that sense, universities and other educational institutions can much more directly tether their educational and research investments to their ethical responsibility to make the world a better place.

### 4. Acknowledgment that "Rights Talk" must be supplemented by "*Responsibilities Talk*"

Although individuals are obviously part of the original State-Individual dyad, the HRBA@Tech model sees them not merely as rights holders, but also as agents who can (and should) act upon their individual responsibilities vis-à-vis others in society. Traditional human rights doctrine focuses on "rights talk"[224] as the primary vehicle for institutional and social progress. Individual rights-holders assert their rights, which they have by virtue of a universally applicable human rights doctrine, against the power of a State. The State, by contrast, is obligated by that same universal human

[224] Intentional reference is made here to terminology introduced by Mary Ann Glendon (1993) Rights Talk (Free Press).

rights doctrine to respect those individual rights. The State, in this traditional model, serves as the sole guarantor and intermediator of human rights entitlements.

The traditional model does not necessitate the introduction of responsibilities-talk, whereby individual actors might also have responsibilities towards one another. Thus, terms like "corporate ethics" or "ethical AI" are eschewed by traditional human rights thinkers.

The HRBA@Tech model, in contrast, embraces such talk as part of the overall approach to NETs. For the HRBA@Tech model to function, individuals, private corporations, civil society actors, educational institutions, international organizations and governments must all be willing to engage with both the traditional rights-based dialectic of social progress *as well as* a more ethically grounded logic of responsibilities and duties of care towards others in society.

5. **Adding a *Temporal Dimension* to a human rights-based approach for NETs - understanding that different interventions make sense at different points during a technology's lifecycle**

   Finally, the HRBA@Tech model recognizes that not all process interventions to 'nudge' technologies towards human rights make sense at all points along a technology's lifecycle. Speaking in generalities, almost all technologies start with a creative idea or an innovation, progress through a range of refinements to eventually become a mature (and perhaps profitable) technology or technologically-enabled product, and eventually start to fade as even newer technologies emerge. Market pressures, as well as the relentless pace innovation, render this an inevitable reality for the vast majority of technologies, especially for companies that do not constantly re-invent, refine, and improve their technologies. The HRBA@Tech model embraces this evolutionary aspect of NETs, thus allowing the model to be adaptable and tailored to the specific context for the NET as it evolves over its lifecycle.

   Certain processes highlighted by the HRBA@Tech model make more sense at different points along this technology lifecycle. Some human rights safeguards are the most effective when they are conceptualized as part of the initial design process, whereas other interventions (for example supply chain monitoring) make more sense at a later stage in the technology lifecycle when a product (assuming there is one) is already being manufactured. Having an awareness of the specifics of a product's lifecycle, as well as a sense of where – roughly – along that lifecycle a particular technology happens to be, allows the analyst to prioritize the most impactful processes for an organization to become engaged in.

## Why Use the HRBA@Tech Diagnostic Tool?

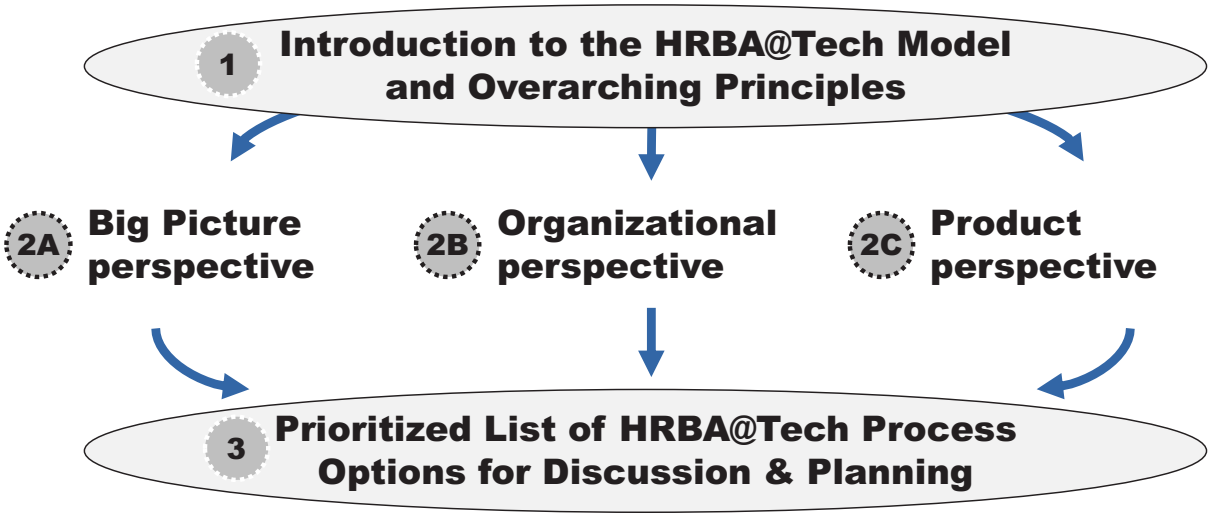
In today's rapidly evolving technology landscape, a variety of private sector companies, non-for-profit organizations, and governmental agencies face the challenge of balancing

innovation with a steadfast commitment to respect human rights. The HRBA@Tech Diagnostic Tool provides organizations and their advisors with a structured method to assess and align technological strategies with human rights norms and priorities. It enables consultants to help their clients mitigate risks, identify potential areas of harm as well as opportunities for pro-human rights impact, and create frameworks that prioritize fairness, non-discrimination, safety, and accountability.

## Structure and Use of this Training Manual

The HRBA@Tech Consulting Approach is divided into five sections. Participants and facilitators are not encouraged to go through the manual in a linear fashion; starting from slide 1 all the way to the end. Rather, the facilitator should be well-versed in the material and ready to skip forward (or backwards) to the appropriate section, discussion exercise, or brainstorming topic – depending on the flow of the conversation. This structure necessitates a facilitator with the experience and self-confidence to treat this manual not as a linear lecture, but rather as a forum in which to host an outcome-oriented discussion.

All consulting efforts should begin with Section 1 of this training manual, and all should end, ideally, by exploring in greater depth one or more of the 24 HRBA@Tech Processes which are found in Section 3 at the end of this manual. Those discussions should lead, ideally, to a customized and targeted action plan specific to the needs of the organization commissioning the effort. How an organization decides upon its own prioritized list of HRBA@Tech processes to focus on depends on the purposes the client might have for conducting this analysis in the first place. The methodology envisages three such possible use cases, which are described in Section 2 of the manual.



## Part 1: Overview of the HRBA@Tech Model, and its foundational Principles

The first section of the manual provides an overview of the HRBA@Tech Model. It serves as the intellectual foundation for this consulting methodology, and therefore all engagements should begin with this unit. It outlines seven overarching human rights principles that serve as the foundation for the HRBA@Tech Model, categorized by whether they are primarily focused on "doing no harm" or "making the world a better place."

The "Do No Harm" pillar and its associated principles, we argue, are mandatory for all stakeholders under *existing* human rights norms. The obligation not to discriminate, or the need to create a viable judicial system, for example, *already* applies to all stakeholders in the technology sector. These principles are thus merely restatements of existing human rights norms, encompassing also the provisions found in the UN Guiding Principles on Business & Human Rights.

The second "Make the World a Better Place" pillar, on the other hand, is more novel, generally voluntary, and focused on *actively* working towards the progressive achievement of human rights-based outcomes in society. The difference between the two pillars can be analogized to the difference between basic mandatory food safety standards *(put in place by a government and applying to all food producers wishing to sell produce in a given market)* and organic food qualification standards *(applying only to those food producers wishing to meet a higher food safety standard, perhaps in exchange for higher prices at the supermarket)*.

### Do No Harm

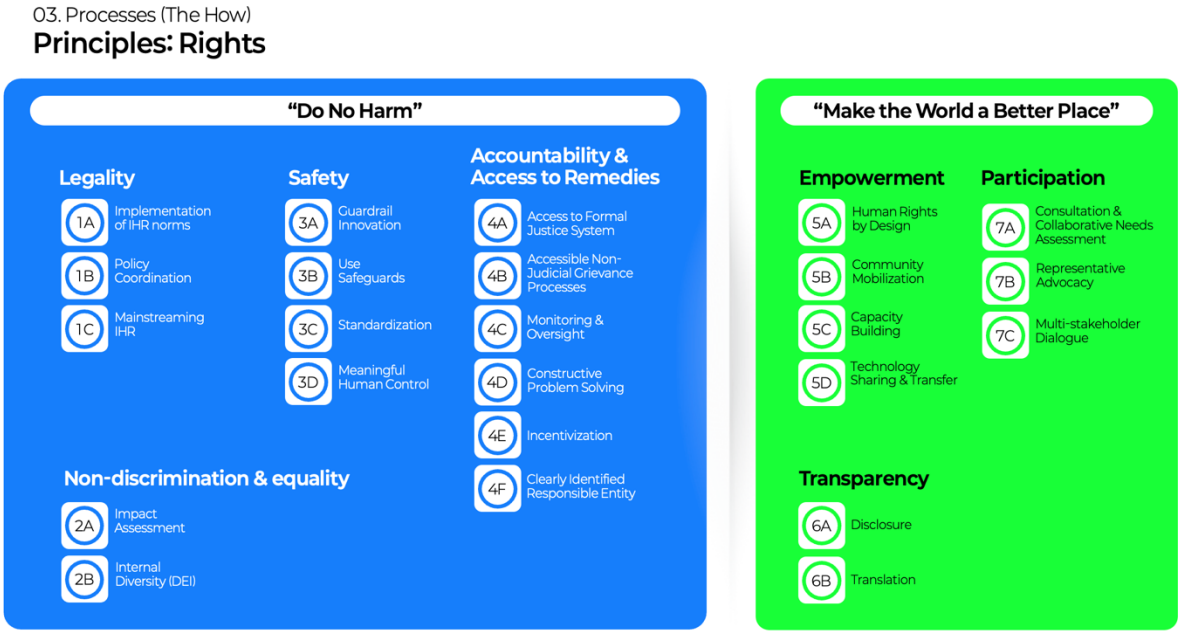
1. **legality**
2. **non-discrimination and equality**
3. **safety**
4. **accountability and access to remedies**

### Make the World a Better Place

5. **empowerment**
6. **transparency**
7. **participation**

Participants in this introductory training are asked to consider these principles and how they might match with the priorities and objectives of the organization conducting the assessment. This can be a prolonged and deeply philosophical discussion. It can also involve very realistic cost-benefit discussions. Facilitators should be ready to help build conceptual bridges between an organization's existing vision and mission statements and the foundational principles of the HRBA@Tech Model, and manage these discussions without judgment or evaluation.

Assuming buy-in to the full set of principles, participants are then introduced to a set of 24 processes associated with the above 7 principles. The HRBA@Tech model proposes that these processes collectively serve to make real the foundational principles listed above by "nudging" a new or emerging technology (NET) in the direction of human rights outcomes and priorities.




Facilitators may wish to spend some time describing the processes in some detail, without exhausting the audience. A key point, however, is that few of these processes can be derived specifically from a "classic" human rights doctrinal source (a treaty or other legal instrument). Here it might be helpful for the facilitator to illustrate one of the principles and its subordinate processes, choosing the principle based on the likely interests or pre-existing needs of the workshop participants. A technology company, for example, might be interested in the principle of 'Safety' and its subordinate processes, whereas a civil society group might be more primed to hear an elaboration of 'Accountability and Access to Remedies.'

At the conclusion of this introductory session participants are asked to decide whether they wish to conduct a systems-wide "climate assessment" (see below, Section 2A), an organizational assessment (see below, Section 2B), or a product-specific assessment (see below, Section 2C). Their answer is usually obvious from the outset, and might easily have

already been pre-determined before the start of the workshop. Depending on their objectives, the facilitator should then proceed to Section 2 of this training manual.

Some groups will want to take the time internally to discuss the contents of Section 1 of the manual, as well as the question posed at the end of this training, with key stakeholders within the organization who are not part of the training. In such situations, the facilitator should call for a break in the training to allow those discussions to take place.

## Using the HRBA@Tech Chatbot

Participants are encouraged to use the custom-made **HRBA@Tech Chatbot Tool** to help with these brainstorming discussions. The role of the tool is to emulate the role of a facilitator if one were actually present in the room during that discussion while providing substantive feedback and direction aligned with the HRBA@Tech model.

This HRBA@Tech consulting manual assumes that facilitators will want to minimize the invasiveness (and costliness) of their consultancies. Doing so will make the assessment and action plan formulation process more sustainable, not to mention less potentially threatening for an organization that is still exploring whether it wishes to proceed with a given course of action.

The HRBA@Tech Chatbot can be particularly helpful when workshop participants are sent away to liaise with their colleagues and come back with a certain decision in hand. This might happen in the context of a short small-group discussion session (for example when workshop participants are split into smaller groups to discuss a certain issue, and one of those participants forgets how the workshop facilitator defined a certain term) as well as when the workshop facilitator calls for a break in the flow of the workshop to give participants the opportunity to discuss an issue in private, consult with colleagues, and return with a decision on how to proceed. In both scenarios, the HRBA@Tech Chatbot can serve as a **sounding board** for ideas, can help explore in a non-binding way the **repercussions of a decision**, and help an individual or group **generate new and creative ideas** in line with the overarching HRBA@Tech model.

As an integral part of the workshop structure, the HRBA@Tech Chatbot helps to make the HRBA@Tech model more streamlined and accessible. It also operates as an accessible tool to be used by any interested party. One might think of the Chatbot as an avenue to evaluate, on a preliminary basis, existing policies, structures, and protocols to examine and identify processes from the HRBA@Tech model that individuals, organizations, and products could incorporate into their routine operations. It can also serve as a virtual brainstorming counterpart, ready to help workshop participants formulate and explore budding ideas of how to take action in line with the HRBA@Tech model.

While the HRBA@Tech Chatbot is designed to be easily accessible and grounded in the HRBA@Tech model, it is also important to remember that it is not a human or a professionally trained facilitator. At times, it too – like many AI systems – can be prone to hallucination. Its

role is merely that of a brainstorming tool that groups can use to help mature their own thinking on a certain topic. All actual decisions, especially those of any real consequence, should be made together with the real (human) facilitator.

## Part 2: Finding the Right Scope for the Assessment

As discussed above, participants will ended the discussion in Section 1 of this manual with the question of whether they wish to conduct (1) a comprehensive "climate assessment" of how technology is 'nudged' in the direction of human rights *in general* in a given society, or they may wish to focus (2) on one particular organization or institution with a view towards maximizing that organization's efforts to promote human rights outcomes, or finally they may wish to focus (3) much more narrowly and specifically on one particular piece of technology at a certain moment in time.

### Section 2A: Conducting a Comprehensive "Climate Assessment"

Some analysts may wish to conduct a comprehensive "climate assessment" (a systems-level perspective on the existing strengths and weaknesses of the environment in which an organization operates). These kinds of snapshots can help policy makers and other analysts assess the need for improved policies or incentives for human rights-based action. This could be useful, for example, in the case of a program manager at an international development organization or foundation trying to determine what kind of capacity building might be the most impactful in a given situation, or a government agency seeking to solicit targeted bilateral support to support a certain priority. Given their sweeping scope, these kinds of analyses focus primarily on a macro-systems analysis, examining how "healthy" the interactions between various stakeholders are, and how well that system serves to produce, on balance, human rights consistent outcomes.

Such an analysis would typically proceed in three steps:

1. **Comprehensive heatmaps**

   As a first step, the analyst conducting a "climate assessment" would have to go through each of the 24 processes, paying particular attention to the kinds of stakeholders who should be most active driving each of those processes. Comparing that theoretical understanding with the real-world context under review, the analyst would then have to conduct focused research to "score" the system regarding the "strength" or effectiveness of each of those processes is functioning.

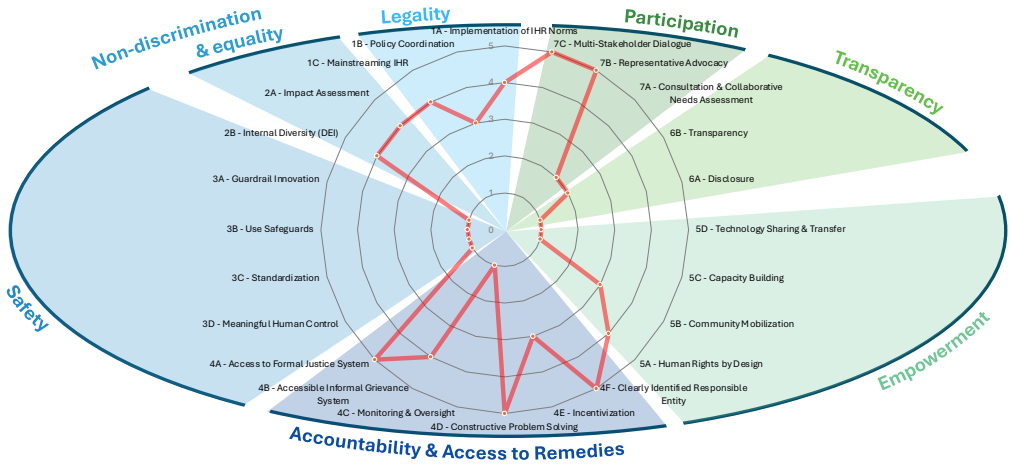



*Figure 1: Hypothetical heat map, "graded" on a 1-5 scale, indicating the relative strengths and weaknesses of a system.*

Although the output of this analysis takes the form of an "objective" quantitative data visualization, it is still (obviously) based only on a group's subjective assessments. Nonetheless, it gives the analyst a sense of where, within the overall HRBA@Tech model, a given system demonstrates its greatest strengths and weaknesses.

2. **Brainstorming**

   What else could be done in this system? What existing processes could be improved, and how? This is a classic brainstorming phase. Groups can take advantage of the HRBA@Tech Chatbot to help animate this discussion.

3. **Action Planning for needs-based capacity building**

   Based on this assessment, the analyst can then begin to think about building a systems-wide action plan for strategic capacity building, based on the perceived weaknesses of the above analysis.

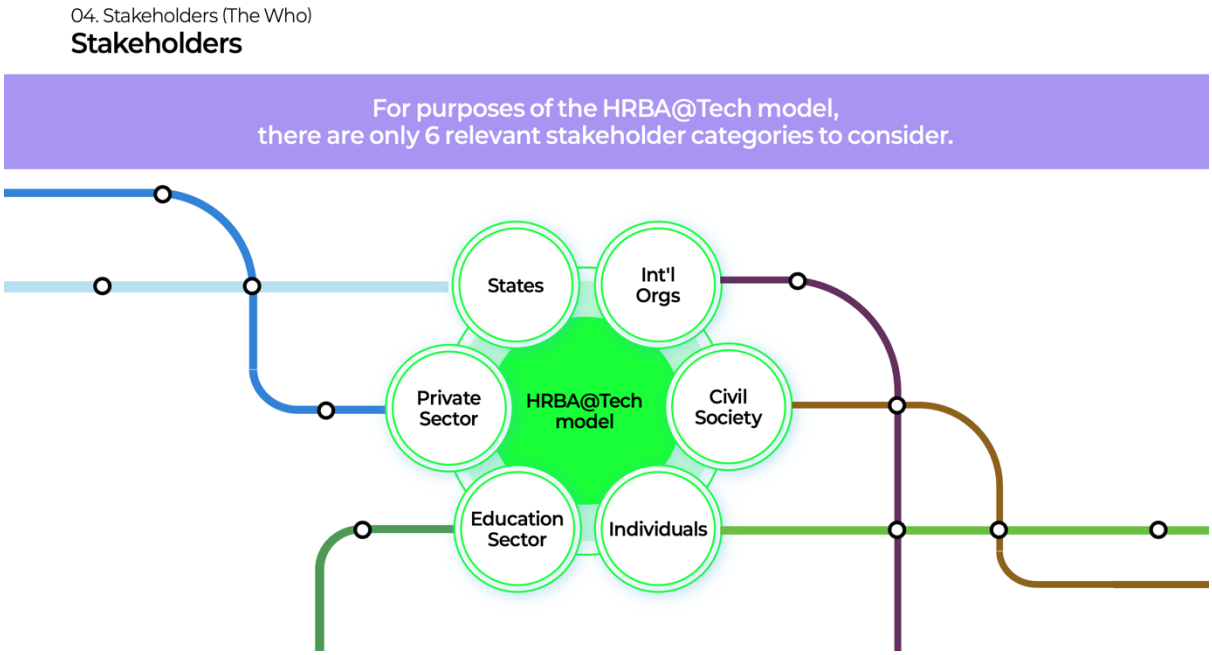
   Doing so will require identifying the "weakest" processes and proceeding to develop a series of process-specific action plans using the corresponding units in Section 3 of this manual.

## Section 2B: Conducting an Organizational Assessment

Other analysts may wish to analyze an organization's strengths & weaknesses according to the HRBA@Tech model, and – building on that analysis – prioritize areas where that organization alone may wish to reinforce its institutional capacity to engage constructively with NETs.

Several types of organizations might be interested in conducting such an assessment. A company, for example, might decide it wishes to review and improve its existing strategy towards developing trustworthy technologies – perhaps in response to a public relations

crisis, or to avoid one from happening in the future. Likewise, a civil society organization may wish to engage more deeply with the human rights implications of technology and consequently begin an assessment of how it might use its institutional power and advocacy tools to greatest effect. Government actors, universities, and even individual groups of researchers might also decide to engage in such a discussion. Any such assessments would likely focus on the formulation of organizational policies that can then be publicized and applied evenly throughout that organization. Such policies must be abstract enough to encompass numerous potential workflows or product streams within an organization, but also tied the organization's unique institutional culture, mandate, and other pre-existing policies.




Such an analysis would proceed in six steps:

1. **Self-Identification**

   The first step in this process would be to clearly identify which of the six categories of stakeholders best describes "us" (the organization conducting this analysis). This should be a relatively straightforward exercise.

   The sole source of confusion may arise between "individuals" and any of the other stakeholders. All of the workshop participants will clearly be both participants *and* potentially representatives of one of the five other categories of organizations (civil society, international organization, private sector, educational institution, or government). In that case, participants should choose for themselves which organizational perspective they wish to adopt, and always keep in mind that all of us – in whatever organizational capacity we are operating, are *always* also individuals who are both agents and subjects of human rights.

2. **Theoretical stakeholder capacity analysis**

   Next, the facilitator should list the 24 HRBA@Tech processes and cross off (or delete) those processes where a particular stakeholder has no role to play.

   This exercise has the potential of being more frustrating than one might initially think, since stakeholders often have peripheral or indirect roles to play in *almost all* of the 24 processes. Consider the process of Implementing IHR Norms, for example (the process of giving international human rights norms binding effect domestically). At first glance, it would seem that it falls exclusively to the State to carry out this process. Upon closer inspection, however, it also becomes clear that civil society groups,

international organizations, individual actors (for example individual lawmakers), educational institutions and even corporations all play significant secondary roles in any effort to give domestic effect to international human rights norms. Thus, it would be inappropriate to suggest that private individuals, for example, have no role to play in this process.

The slides illustrate the example of a private sector corporate actor presumably not being very active in the process of representative advocacy (the process of advocating on behalf of a vulnerable community that cannot or is not speaking up for itself). Here too, though, some eager participants may volunteer that the corporation can engage in lobbying activities, and also that technically speaking it is acting on the interests of its owners or shareholders. This example serves to illustrate the challenge for the facilitator, which will be to manage this conversation but NOT push back if the group wishes to make a point. Groups will sometimes do this to throw a facilitator off-balance, and the facilitator at this point should not take the bait.

The result of this discussion will ideally be a list of less than 24 remaining HRBA@Tech processes. From that point on, the analysis would focus *only* on that smaller subset of processes.

3. **Organizational heatmap (what are we currently doing?)**

Similarly to the comprehensive heat map described above, the analyst would now work with the workshop participants to "score" the organization regarding each of the highlighted HRBA@Tech processes.

This discussion would also have to focus on how successful participants feel that current efforts are. For example, if a corporation feels that it is doing impact assessments but isn't seeing results it had hoped this process would achieve, the group can decide to re-examine and perhaps re-design that process.

4. **Brainstorming**

What else could we be doing? What existing processes could we be doing better? This is a classic brainstorming phase. Groups can take advantage of the HRBA@Tech Chatbot to help animate this discussion.

5. **Capacity | Mandate | Organizational Appetite Assessment**

Not all organizations have the capacity, mandate, or organizational "appetite" to engage in each of the 24 HRBA@Tech processes, even if *hypothetically* they could. In fact, doing so would likely lead to failure for lack of focus. Profit-driven businesses have obvious constraints on their ability to dedicate large sums of time and money towards the achievement of human rights. Similarly, some civil society organizations have a significant incentive to remain in close alignment with their organizational reputation. A non-profit known for its adversarial strategic litigation strategy, for example, might have a very hard engaging in "constructive problem solving.

Facilitators will need to work with groups to compare an assessment of where the organization's current activities could hypothetically stand to be strengthened with a cleareyed assessment of where an organization might be willing to expand its investment. These conversations will often also need to take place behind closed doors – without the facilitator's presence.

Here too, the HRBA@Tech Chatbot can again play a very useful role in helping groups carry on frank and substantively informed discussions in a way that feels conducive to real brainstorming and decision-making.

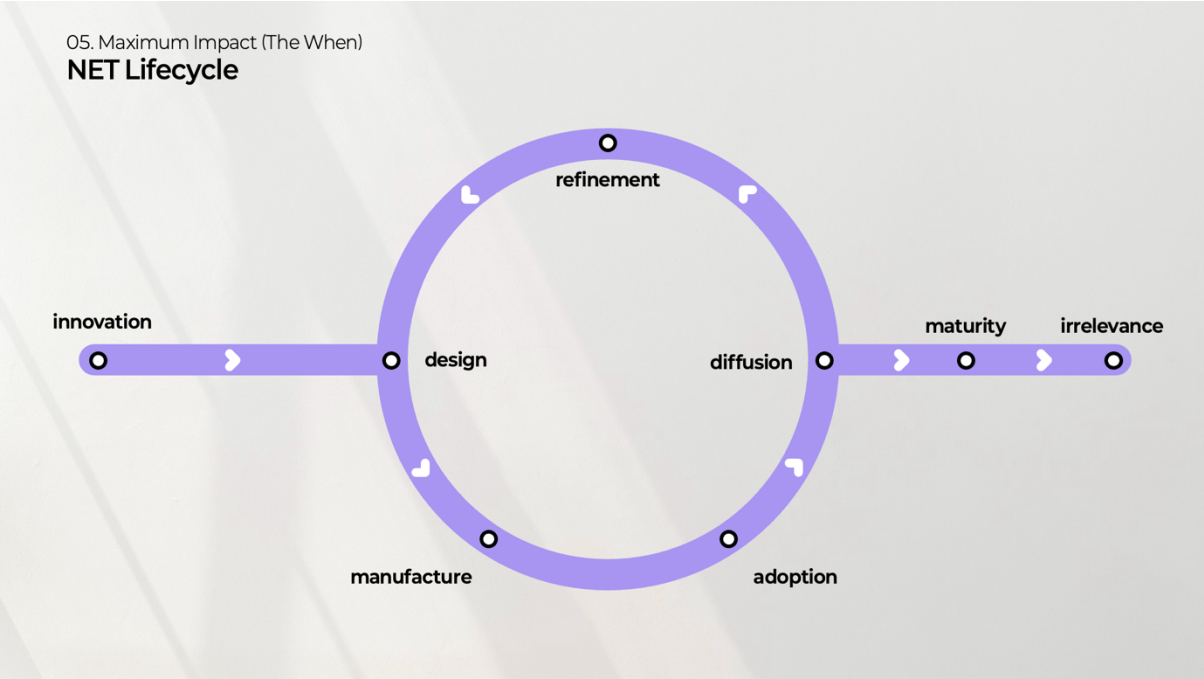
6. **Action planning**

Once a group has completed the above steps, they can then brainstorm concrete actions for their organization can take, based on the perceived weaknesses of the above analysis, to improve its organizational approach to NETs. This will involve developing a series of process-specific action plans using the corresponding units in Section 3 of this manual.

## Section 2C: Conducting a Product-Specific Assessment

A final use case for this HRBA@Tech manual would be to apply the HRBA@Tech model to one specific technology (or tech-product) at a specific moment in time (presumably the present). This type of assessment makes sense to project managers focused on a particular technology.

The analyst in such situations would first need to understand where along the technology lifecycle a specific technology finds itself. This process requires first fine-tuning the *generic* technology lifecycle to reflect the terminology and specific characteristics of the technology in question.



The analyst will want to do this "fine-tuning" process collaborative, drawing on the wisdom of the workshop participants to share their preferred terminology.

Such an analysis would proceed in seven steps:

1. **"Translating" the HRBA@Tech model's generic Technology Lifecycle into one that has more meaning to the group and product under discussion**

   The first step in such a product- or technology-specific analysis is to first understand where along the technology lifecycle (TLC) a specific technology finds itself. This process often requires first fine-tuning the generic TLC to reflect the terminology and specific characteristics of the technology in question.

   This step requires the modification and customization of the terms and nodes one would use to describe a TLC. The HRBA@Tech model proposes a generic TLC, and each node in that lifecycle is described in abstract but functional terms. When an analyst begins to work with an individual technologist or entrepreneur, they may often be fixated on very different defining milestones of their product's or technology's lifecycle. Some of the nodes in this generic product lifecycle do not apply 1:1 to every NET. In other situations, a certain phase in the generic product lifecycle may require clarification (an possible sub-division) in order for it to be meaningful for a given technology. In such a situation, it makes little sense for the facilitator to 'force' the workshop participants into an imposed description of a generic TLC! Doing so would likely alienate the workshop participants and lead them to suspect that the facilitator doesn't know much about (and worse—also isn't interested in) the specifics of the industry or technology under discussion.

   Much more advantageous is the decision to first 'customize' the technology lifecycle.

   An example serves to illustrate this process. When speaking with AI technology startups, for example, the founders' and investors' attention may be focused less on the generic terms of the technology lifecycle and much more on the particular challenges of getting a startup to "exit" – either being purchased by larger technology company or going public on the stock-market. Those stages tend to be much more prominent in the minds of startup entrepreneurs as they consider which HRBA@Tech processes they feel comfortable engaging in, not to mention how they would justify those investments at any given point in time.

   Moreover, the specific phases of bringing an AI technology to market differ. Assuming an AI product housed in the cloud, there is virtually no "manufacture" phase to speak of. That said, there is a functional equivalent of the generic "diffusion" stage of an AI product, where the product is released and gains widespread acceptance among the potential userbase. This phase, in the parlance of many AI entrepreneurs we spoke to, was more appropriately described as "maturity."

Sticking with this example, the analysts might have determined (as illustrated in the chart above) that one might leave the generic "Innovation" phase unchanged, but 'translate' the generic "diffusion" stage to "maturity." Filling in the intervening nodes, a discussion might suggest that after an AI product is first innovated, it then needs to go through a period of intense research, followed by a provisional release and another period of intensive refinement before it can finally be described as having reached maturity.

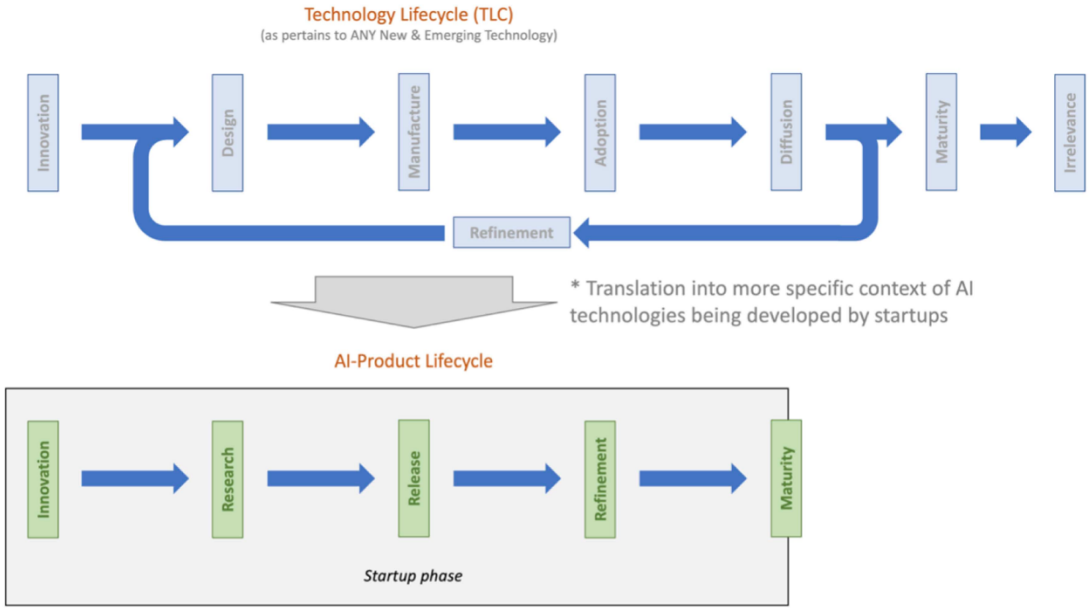



*Figure 2: excerpt from the 2023 report "A Human Rights-Based Approach to AI for Tech Startups and Global Normative Governance" (p. 45)*

From a facilitators' perspective, the specifics of such a description are absolutely secondary to the ideal that a group would feel that the translation they generated – whatever it may be – feels right to them. This is a crucial moment for the facilitator to cultivate the intellectual 'ownership' of a group.

## 2. Mapping processes onto the newly customized Technology Lifecycle

As a next step, the group would want to discuss which of the 24 HRBA@Tech Model processes are applicable to each of the new nodes of the group's customized TLC. This discussion can take inspiration from the generic TLC, which is described in the slide-deck, but will likely require additional and sometimes in-depth discussion about the relative value of certain HRBA@Tech model processes.

Using the slide-deck, the facilitator should refer to the relevant slides pertaining to each "phase" of the TLC to kick-start the process-mapping exercise. Each section begins with an overview definition of that stage of the lifecycle, provides a through general-level considerations, and then goes on to illustrate a few common processes that companies engage in during that stage of the technology lifecycle.

**Note**: The names of these processes are *different from* the list of 24 processes featured throughout the workshop training. Some of these processes are context or hybrid processes, and must therefore be mapped back on to the list of 24 processes. **Grievance Process Design**, for example requires that the designers not block claimants' access to the formal justice system (Process 4a), focuses on the creation of an accessible non-judicial grievance process (Process 4b), forces the designers to engage in constructive problem solving, ideally with concerned stakeholders (Process 4d), and finally requires for there to be a clearly identified responsible entity (Process 4f).

**Note:** Facilitators will want to read the relevant section in the report on the innovation phase. That section is unique from the others in that it requires a different response based on what kind of organization is being discussed, and also what kind of a technology that organization wishes to create. A private business, for example, should be held to completely different standards during the innovation phase than a non-for-profit public university or government, for example. Whereas a private enterprise might well embrace profit as a primary motive for its technology innovation, a university or government could be expected to embrace a more altruistic, "make the world a better place" approach towards its innovation.

Likewise, participants will be asked whether the purpose of a NET is (a) to make the world a better place, (b) make profits, or (c) actually make people worse off (for example a weapons manufacturer incorporating AI into those weapons). The facilitator should be agnostic about the groups' choice, but depending on their answer (again) the degree of care demanded by the HRBA@Tech approach would be quite different, as described in the slides and 2022 paper.

Referring back to our earlier example of the HRBA@Tech model as applied to AI technology startups, a discussion might yield an assessment that the best time to engage in efforts to 'make the world a better place' might be during the research phase of an AI product's TLC, and the best time to invest in 'do no harm' initiatives might equally be during the research and release phases of the TLC.

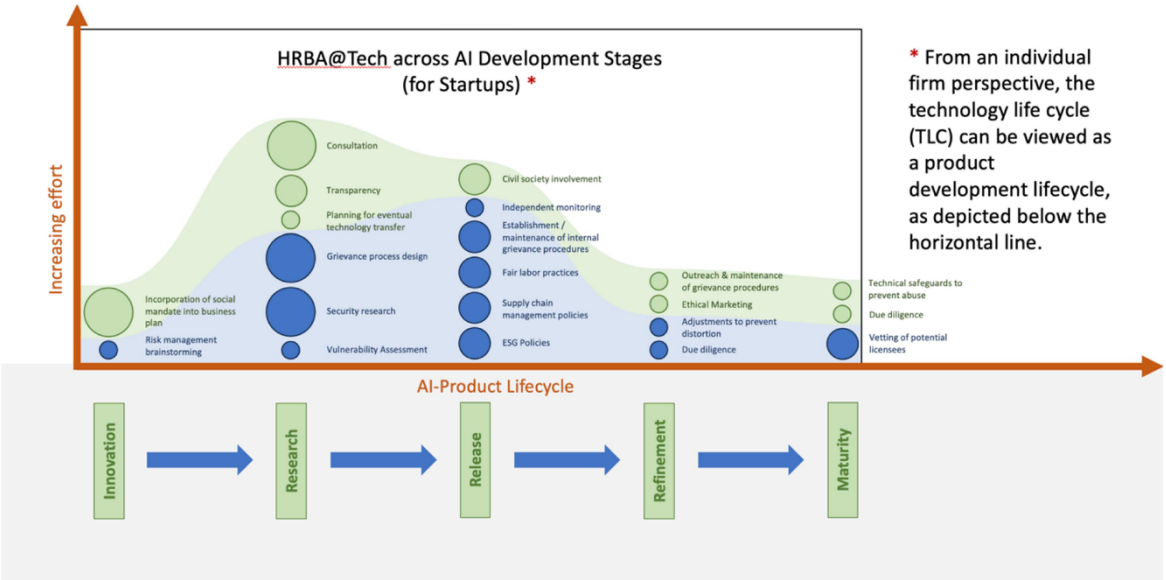



*Figure 3: excerpt from the 2023 report "A Human Rights-Based Approach to AI for Tech Startups and Global Normative Governance" (p. 60)*

This kind of analysis lays the groundwork for a narrowing down of processes that the group may wish to focus on as part of its concrete action planning.

3. **Pinpointing a Technology's Progress along the newly customized Technology Lifecycle**

   Once the group has 'customized' its TLC model, it can then discuss where along that newly translated TLC the particular product or technology under discussion happens to find itself, and how close that technology might be to progress to a subsequent phase or phases.

4. **Checklists & heatmaps (what are we currently doing?)**

   Once the group has 'customized' its TLC, mapped potentially impactful processes onto that new TLC, and also positioned the technology or product under consideration *roughly* along that customized TLC, the discussion can then move to an analysis of what the organization is already doing to 'nudge' the technology or product in the direction of human rights. Groups may also wish to discuss their assessment of how effective those existing approaches may be.

5. **Brainstorming**

   What else could we be doing? What existing processes could we be doing better? This is a classic brainstorming phase. Groups can take advantage of the HRBA@Tech Chatbot to help animate this discussion.

6. **Capacity | Mandate | Organizational Appetite Assessment**

Just as with the organizational assessment, groups would next need to discuss their particular team or organization's "appetite" to engage in highlighted HRBA@Tech processes.

Facilitators will need to work with groups to compare an assessment of where the organization's current activities could hypothetically stand to be strengthened with a cleareyed assessment of where an organization might be willing to expand its investment. These conversations will often also need to take place behind closed doors – without the facilitator's presence.

Here too, the HRBA@Tech Chatbot can again play a very useful role in helping groups carry on frank and substantively informed discussions in a way that feels conducive to real brainstorming and decision-making.

7. **Action planning**

   Finally, as with each of the other assessments, the facilitator can then move on to concrete action planning with the group using the corresponding slides in Section 3 of this manual.

## Part 3: HRBA@Tech Processes

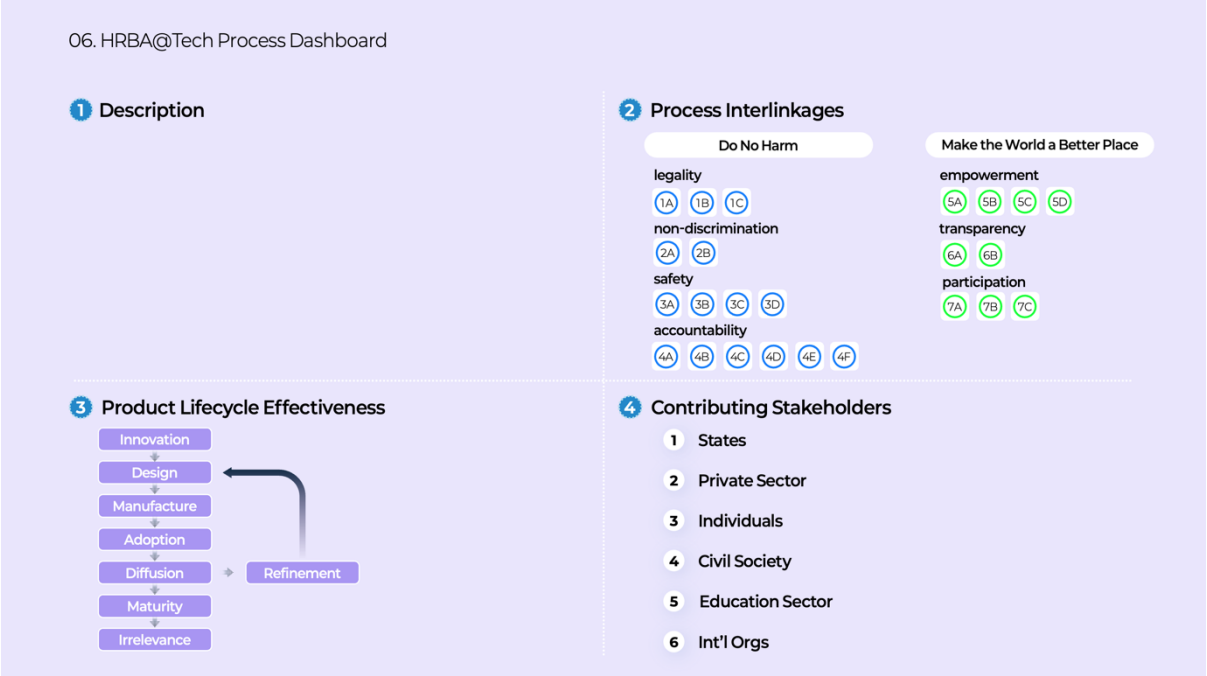
By the time participants have completed parts 1 & 2 of the HRBA@Tech Diagnostic Workshop they should have come up with a significantly shortened subset of the 24 processes that collectively constitute the HRBA@Tech model. Ideally, the group would have chosen to focus on no more than 1 or 2 processes. That shorter list of processes should be the focus of the last unit of this workshop. In this stage, participants discuss in greater detail each of the specific action plans they selected, contextualizing it to the specifics of their organizations' or product-line's unique context. Building on that understanding, participants then craft an action strategy on how they envisage sustainably engaging in that particular process. This action plan, will then constitute a contribution towards the organizations' overall "human rights based approach" to NETs.

Each of the individual Process "Units" in Part 3 of this workshop are structured identically. The facilitator will want to use four slides to guide this conversation. These are best reproduced on a whiteboard or on flipcharts for participants to fill out themselves as part of their discussions.

1. **The Process "dashboard"**

   The "dashboard" is the summary of all aspects of the process. These dashboards are not pre-filled. Instead, they are a vehicle for the group to develop a contextualized understanding of that particular process *collaboratively*. "Forcing" the group to develop that understanding fosters more authentic ownership of the concept, which

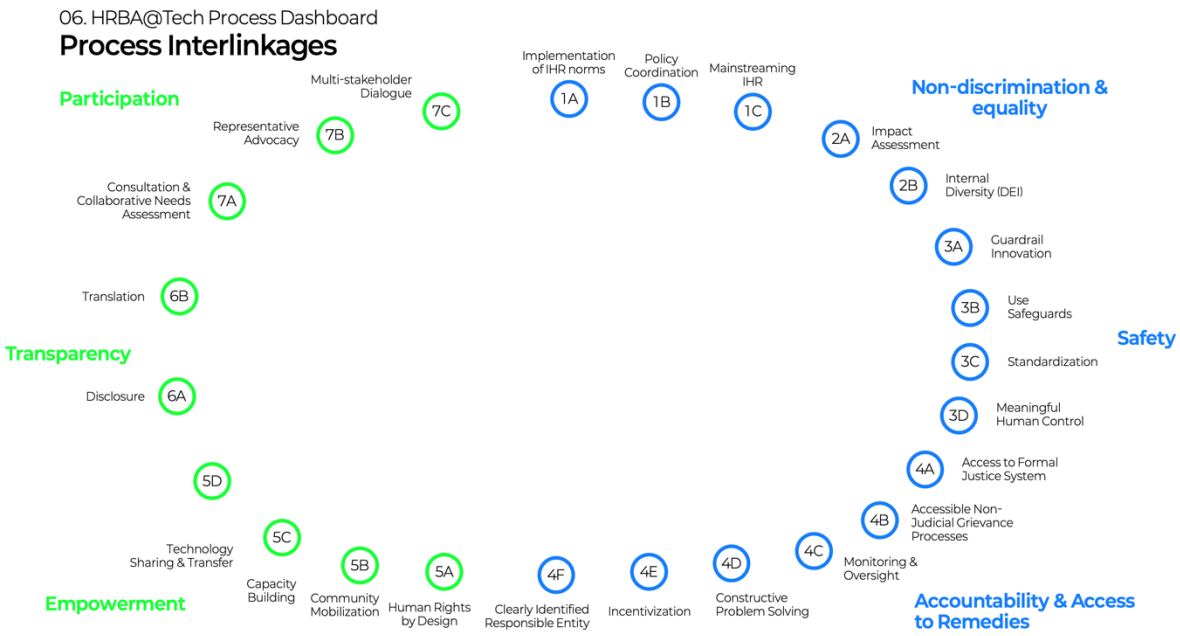
serves as an important prerequisite for the constructive brainstorming effort which follows.




The facilitator should introduce the idea of a Process "dashboard" by means of a simple example. Each of the three subsequent worksheet "slides" will contribute one essential element to that dashboard.

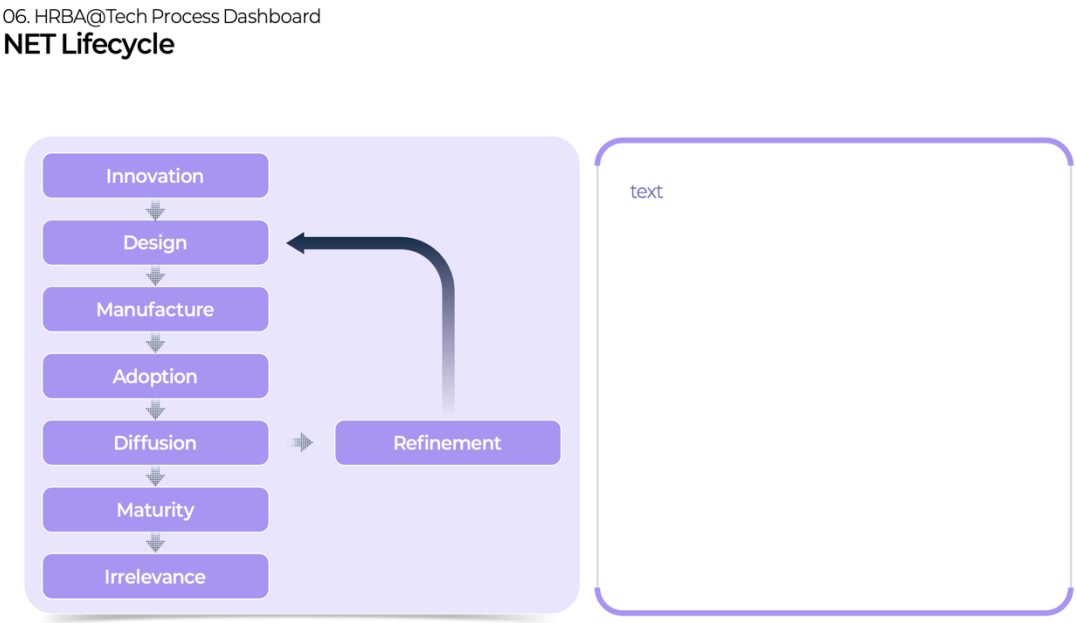
2. The **process interlinkages map**

   Participants should fill this 'map' out twice. They can use the same physical "map" for each cycle of this analysis but should draw arrows using different colors if working on a flipchart or whiteboard. The first cycle should highlight process that *contribute positively to* the process under consideration (the "process influencers"). The second should identify processes that are themselves impacted by the process under consideration (the "process multipliers"). Filling out this map will help illustrate how organizations can contribute constructively but *indirectly* to a process under consideration and build a stronger justification for the instrumental value of investing in a certain type of HRBA@Tech process. The results of this analysis can be used to fill in the northwest quadrant of the process dashboard.




3. The **product lifecycle map**

   Participants should use this worksheet to pinpoint where along a product lifecycle a particular process makes sense. This is important to highlight instances where investing in a certain process may have very low impact in terms of the actual products an organization seeks to influence. The results of this discussion can be used to highlight the corresponding product lifecycle nodes in the southern quadrant of the dashboard.



4. The **stakeholder role** map

   Finally, workshop participants will want to broaden their understanding of how other types of stakeholders might also be essential for a process to function effectively. This can be essential later in the workshop, when participants are asked to conduct a stakeholder mapping and influence strategy. The results of this analysis can be used to fill in the northeast quadrant of the process dashboard, primarily by adjusting the 'relevance' of certain stakeholders vis-à-vis the central process definition.


5. **Return to the Process "dashboard"**

   After filling out the three peripheral quadrants of the dashboard (process influencers, process multipliers, applicable product lifecycle stages, and influential stakeholders), the group should then be asked to revisit (update and customize) the generic process description first presented. This will be the culminating step in building a group's shared expertise on what a process truly means *to them* given their organization's unique context, mandate, and technology exposure.

## Action Planning

Once the group has finalized its process dashboard, it can then move to action planning. At this point the facilitator enters into management consulting or change management mode, drawing on classical consulting tools and methods to push the conversation forward.

At this point the workshop participants will ideally transform from a group of passive workshop participants into a more agentic planning committee (and will hereinafter be referred to as the "workshop committee"). They should ideally begin to see themselves as the champions of a particular HRBA@Tech innovation within their organization.

Facilitators, at this point, will also transform themselves from content deliverers to action planning coaches, using simple exercises to help groups craft their individual action strategies. They should take care to NOT, suggest, however, that there exists a universally applicable 'best practice' out there that would be appropriate for all organizations and in all contexts, nor should they reinforce their own role as an "expert" who somehow has more to contribute to the action planning process than the workshop committee themselves.

**Step 1: Does the group wish to champion this organization's efforts to engage in a particular process under consideration?**

Before embarking on any genuine action plan, the workshop committee will have to first determine whether they intend to more forward with a particular reform idea. Some groups will have already been delegated by senior management to act as the point-persons for an organization's investment in a certain process. In such cases, the question may have already been *partially* solved. For such workshop committees, this might be a good occasion to return to those senior managers and communicate with them the implications of what it would mean for the organization to move forward: verifying – in a sense – that they still enjoy their managers' endorsement for their efforts.

Other workshop committees will have been asked to participate in this workshop with a more exploratory mandate. Such workshop committees should take time at this point to consult with colleagues within the organization, using what they have learned in the workshop so far to facilitate transparent discussions with colleagues about the institutional desire to move forward with a particular reform idea. The facilitator should ideally NOT serve as a resource person for these internal deliberations, even if doing so might appear to be more efficient. The functional value of (essentially) "forcing" the workshop committee to plan and conduct these discussions is that they will then also be much more likely to grow into the role as institutional champions and "owners" of a reform project.

A secondary impact of forcing the workshop committee to take ownership over these internal consultation processes is that they will also (inevitably) localize, customize, and adapt the language they use to describe a particular process to be much more closely aligned to existing institutional norms and existing process. This 'endogenization' process is also crucial to its ultimate success of the project.

What would a "green light" from decision makers and colleagues look like?

Ideally, the group could achieve an agreement to conceptualize, within a given period of time, a particular HRBA@Tech process proposal to a sufficient level of granularity such that management and other relevant stakeholders can decide if they wish to engage in that process.

**Step 2: Decision Maker Systems Analysis**

Not all decisions to engage in a particular process have one single decision maker, and often even if there is one such centralized decision-making figure, the pathways to secure the attention of that decision maker may not be immediately apparent. A facilitator can work with a workshop committee to identify the various stakeholders who may have an interest in a reform proposal, whether they might tend to be generally positively or negatively inclined towards that proposal (or neutral), the degree of power they have to influence a decision about that reform proposal, and finally what influence relationships exist between those various stakeholders. This kind of stakeholder mapping exercise can be done together as a group, and serves to highlight key initial stakeholders to approach.

**Step 3: Action Mapping & Consensus Building**

Having mapped an institutional system, the workshop committee would then have to go out and actually engage with colleagues, stakeholders outside of the organization, etc., essentially doing their best to build a coalition in support of their reform proposal. Over the course of these consultations and engagements, the workshop committee will likely be confronted with new insights and perspectives they had not previously considered, and may be tempted to substantively modify a reform proposal. As "owners" of that proposal, they should be fully empowered to do so.

**Step 4: Policy Formulation & Capacity Building**

Pending sufficient success in this consensus building process, the final task for the workshop committee would be to craft a series of policies and capacity building efforts internally to ensure that the organization's reforms are embraced by all relevant parts of the organization, and that these process participants ultimately also understand the overall purpose and vision motivating a particular HRBA@Tech reform.

### Monitoring & Evaluation

Planning committees should also discuss what metrics or indicators they will look to when evaluating the "success" of any process innovations they decide to champion. The best time to have these discussions is during the brainstorming and action planning phases, ensuring that the performance of a particular reform effort can be tracked longitudinally from the outset of a reform effort.

Discussions about monitoring and evaluation should focus on finding meaningful indicators that would represent success. These should be quantitative indicators that meaningfully track the planning committee's understanding of what 'success' looks like. Efforts to regularly gather and report on these indicators should be incorporated from the outset into the workflow process.

Planning committees should also consider scheduling a 'one-year multi-stakeholder check-in and feedback session', where the committee convenes a broad range of relevant stakeholders to solicit early feedback and reactions to the newly implemented systems, allowing groups to supplement any quantitative data about the performance of a new system or process also with qualitative inputs. Various models of soliciting feedback can be appropriate, including focus groups.

This consulting model is unusual, and particularly dynamic, in that it represents a two-way flow of information and lessons learned. In its ideal form, information, knowledge-sharing, and innovation flow in both directions between the facilitator and the planning committee. The facilitator, to be sure, works to help helping groups design and customize new human rights-based models to ensure the safety and reliability of technology products. But just as importantly, the knowledge and wisdom generated within companies and organizations grappling with the implementation of the HRBA@Tech model would also flow back out to the consultant. This two-way learning process is central to the long-term dynamism of this model of consulting and separates it from other models that risk quickly growing "stale" in light of the rapid pace of technological innovation.

Future consulting protocols and contracting arrangements could systematize that two-way learning process while also preventing the leakage of proprietary or sensitive internal documents or discussions beyond the walls of a contracting organization.

## What sets this approach apart from a regular management consulting "product"?

This training is structured similarly to the kind of 'product' or 'service' one might see from a classical management consulting firm. Indeed, the skills a facilitator would need to bring to the table are not dissimilar from those that a classical management consultant might also need in order to succeed.

That said, several key features distinguish this approach from a classical management consulting approach, and it is very important for these distinguishing features to be noted prominently by the facilitator and also by the "client." Understanding and memorializing those distinctions will make a crucial difference in the long-term social impact and viability of this initiative.

Some of these distinctions *may* impact the profitability of assuming the management consulting or facilitation role. Classical management consultants profit handsomely from their engagements, but they are usually also subject to very rigorous **non-disclosure agreements** (NDAs) as a result. A consultant may, for example, engage with a technology company to help facilitate a change-management process with them but be barred from subsequently discussing what they experienced, observed, or even learned from that engagement to anyone outside of that technology firm. Such NDAs are common for management consultants. The existence of an NDA, however, makes it virtually impossible for a consultant to feed learning – even anonymized learning – *back into a general pool of communal knowledge*. Such consultants may *themselves* become wise and very experienced over time, but they rarely have the opportunity to contribute to a generally-accessible knowledge-bank that others can also draw upon. This model may increase the (marketable) knowledge capital of an individual consultant, but not the overall knowledge capital of a system beyond that one expert.

This HRBA@Tech manual presumes a different approach, as described below. It imagines a two-way learning process, where the facilitators prompt curiosity, inquiry, and open brainstorming conversation with their workshop participants, but also themselves engage in rigorous learning process. That learning must then be captured, processed, and fed back into a generalized knowledge bank for use not only by the individual consultant in future engagements, but by ANY consultant adhering to this similar methodology.

## Two-Way Learning Process

A crucial element running through out this third unit is that these are fundamentally *learning conversations*, not just for the participants but also for the facilitator. Over the course of conducting many of these workshops, the facilitator accumulates greater and greater awareness about the ways that real-world organizations achieve human rights-based approaches to the development of NETs. Facilitators will accumulate greater insights into what kinds of interventions prove to be attractive, as well as the arguments and metrics that help convince other decision makers within an organization to move forward with an idea. Facilitators rely on this situational background for their credibility conducting these kinds of workshops.

It is both unfair and unrealistic to expect, however, that the success or failure of this entire capacity building exercise must rest in the accumulated knowledge of a select-few workshop facilitators, however. For this reason, it is imperative that the learning process result in a "knowledge bank" that others can also draw upon. Workshop participants should commit – as a prerequisite for participation in this training – to allow generalized descriptions of the solutions they ultimately decide to implement to be fed into the HRBA@Tech knowledge bank. Optionally, the participants could also agree to have transcripts of their deliberations during the workshop (anonymized and without attribution to speaker, or the organization they represent). This stylized process descriptions, which can be anonymized if the organization wishes not to be publicly associated with the outcome of the workshop, should be drafted by the facilitator using the language and concepts of the HRBA@Tech model. Over time, this repository of outcomes can be used by other workshop facilitators to help in future HRBA@Tech diagnostic workshops. Together with the anonymized transcripts of the deliberations leading to the outcome, they can also help feed the HRBA@Tech Chatbot thereby ensuring that this ancillary tool also becomes trained – iteratively – on the full slate of creativity generated by these workshops.

## Confidentiality and Anonymity

A first crucial consideration of this model is that the ironclad NDAs customary to most consulting engagements needs to be relaxed in minor but important ways. Trade secrets, business strategy discussions, and other sensitive business information *must,* of course, continue to be protected. So too should the individual contributions of workshop participants be protected, similar to a "Chatham House Rules" norm protecting the identity of the individual contributors while still leaving the substantive content of their contributions for public review. Finally, the identity of a company or organization that decides to embrace – in good faith – the HRBA@Tech approach to NETs should be made public. Doing so will hopefully add to the reputation of the HRBA@Tech model, but also inure to the benefit of a company or organization as a public manifestation of its commitment to human rights and human welfare.

## Memorialization of Lessons Learned

Facilitators should not just be a facilitators, but also 'students' of their own engagement process. After the conclusion of their engagement, as a *penultimate* work product, the facilitator should draft a memorandum detailing the results and insights of a particular engagement. This memorandum is described as the facilitators' "penultimate" deliverable, because they would subsequently need to share that memorandum with the organizations that commissioned them and then subsequently revise that document to represent a broader set of insights and experiences.

The memorandum would be drafted specifically to echo and reinforce the concepts in the HRBA@Tech model, but also to make explicit any 'translations' that were necessary to make

the more generalized HRBA@Tech model applicable to the unique operating environment of the individual organization. Facilitators could detail why certain processes were appealing to a certain organization, why others were less appealing, and also how an organization decided to implement a particular HRBA@Tech process to maximize its beneficial impact within that organization.

The results of this memorandum could be compiled in a traditional format (in the form of case studies), but also used to feed the learning of the HRBA@Tech Chatbot, which could periodically use these memoranda as inputs to improve its 'knowledge' of how the HRBA@Tech model translates into reality.

## Follow-up

As alluded to above, part of the facilitator's commitment should include a guarantee of being able to convene, either virtually or in person, a group of key workshop participants a set period of time (6 months or 1 year) after the conclusion of the engagement. In advance of this follow-up meeting the consultant would have shared the lessons-learned memorandum (see above) with the former workshop facilitators, asking them specifically for feedback and insights that only they could deliver.

This final focus group will also allow the facilitator to generate valuable feedback about several additional elements:

- Their own performance as a facilitator, as well as the overall effectiveness of the workshop model;
- The long-term viability of the ideas that were generated as a result of the workshop;
- Any preliminary data demonstrating the successes of a particular process innovation (drawing on the Monitoring & Evaluation discussion above);
- Any changes to the process initiatives that emerged as a result of time and experience;
- Any suggestions for improvements to the underlying HRBA@Tech model, now that workshop participants had some time to reflect on it.

After the conclusion of this final phase, facilitators can update and finalize their lessons learned memoranda and feed them back into a communal and constantly growing "knowledge bank" to supplement the HRBA@Tech training manual.

## Setting the Stage: Entry & Contracting

Entry and contracting is a crucial element of any facilitator's engagement process. It is *especially* important if the facilitator is part of an effort to build and strengthen the HRBA@Tech model. As described above, the facilitator will need to relax traditional NDA requirement, be 'allowed' to capture learning about an engagement in a semi-public document (the lessons learned memorandum), and be required to follow up with workshop participants after a given period of time to inquire (among other topics) about the organization's follow-trough on their commitments to implement new or overhauled HRBA@Tech processes.

In exchange, the facilitator allows the company or organization to claim that they used the HRBA@Tech model to demonstrate their commitment to human rights objectives, and also (see below) to possibly have their efforts certified. It also would give them free access to the HRBA@Tech Chatbot, possibly well beyond the conclusion of the workshop itself.

These provisions would have to be memorialized clearly in a contracting process and made very clear to the management team involved in any decision to engage an HRBA@Tech facilitator.

## The HRBA@Tech Chatbot

Throughout this manual, reference has been made to the HRBA@Tech Chatbot. This chatbot is a custom-tailored tool that can both maximize the impact of this workshop model while also minimizing the hours a facilitator needs to spend with a group to engage in rigorous brainstorming. Just like the Large Language Model (LLM) chatbots that many of us might already be familiar with (ex. Claude by Anthropic, Llama by Meta, Gemini by Google, ChatGPT by OpenAI, or Grok by xAI), this chatbot allows users to interface using natural language with a cloud-based chatbot, which then returns natural-language answers that have been specifically trained to be consistent with the HRBA@Tech model.

All chatbots, of course, are prone to hallucination, and this is no exception. The chatbot should thus never be used as a *replacement for* a human facilitator or human workshop participants! That said, the chatbot can serve a very valuable brainstorming role, helping individual workshop participants accomplish the following:

- Clarify workshop concepts that may not have been abundantly clear the first time they were presented.
- Brainstorm concrete ideas of how an organization might move forward with a reform idea in ways that fit within that organization's existing policy landscape.
- Prepare for a presentation to colleagues within their organization who are not part of the workshop to explain to them, in simple and intuitive terms, the implications of the ideas being discussed in the workshop.
- Upload and analyze an existing organization policy in terms that resonate with the 24 HRBA@Tech processes or 7 HRBA@Tech principles.
- "Play forward" the implications for an organization's overall operation if it were to implement a certain HRBA@Tech process, focusing on both potentially beneficial as well as potentially negative impacts (and then brainstorming ways to possibly mitigate those downsides).

Workshop participants could make use of the HRBA@Tech Chatbot either in small group settings *during* a workshop or during strategic breaks in the workshop for participants to do their "homework" on a given topic.

### Technical Specifications

The HRBA@Tech Chatbot was created on the basis of Meta's pretrained conversation model, LLaMA-2-chat. This open-source baseline model was then fine-tuned through a dataset of instructions-and-answers pairs either directly inspired by the HRBA@Tech 2022 report. By curating a robust set of these pairs, the HRBA@Tech Chatbot is highly capable and knowledgeable in topics discussed in the HRBA@Tech 2022 report, making it a useful as a supplemental tool alongside the HRBA@Tech consulting manual. After training the base Meta model on this dataset, numerous testing stages—ranging from simple prompting to direct

human reinforcement—took place to test the efficacy, efficiency, and the adeptness of the HRBA@Tech Chatbot.

### Ongoing Reinforcement Training

As with any language model, the HRBA@Tech Chatbot will not just rely in perpetuity on its initial fine-tuning. The HRBA@Tech Chatbot will need to adapt to changes and updates from the HRBA@Tech model as well as learnings from facilitators who will provide their insights as data to improve its underlying knowledge base. Feedback generated by future consulting engagements can be used to adjust the and strengthen the knowledge base feeding the HRBA@Tech Chatbot. Users of the tool will also have a means to reporting trivial as well as serious errors made by the Chatbot, as well as avenues to affirm any particularly insightful or useful traits of the HRBA@Tech Chatbot.

## Certification (?)

A final question – and one that goes well beyond the scope of this training manual – is whether any organization (for example a UN Agency such as OHCHR, or in the alternative a private or civil-society run entity) might be interested in using this approach as the basis of a process-based certification process that it might offer to companies or other organizations that – in good faith – complete one or more rounds of this HRBA@Tech change-management process.

# Summary

The following table summarizes and roughly-estimates the time requirement for the overall consulting and facilitation engagement.

| **Preparation** Setting the Parameters | | | |
|---|---|---|---|
| 1 | **Entry & Contracting** | | |
| | ~ 1-6 months | [Facilitator] | This workshop requires the commitment by an organization to at least seriously consider changing some aspect of how they approach the development and deployment of NETs. Of course, an organization may always decide against a potential plan of action or reform program, but there should be an element of "good faith" in the contracting arrangement. Facilitators should include in the agreement a willingness by the organization to allow the facilitator to document the lessons flowing from an engagement, not just during the engagement itself, but also one year out.<br><br>A payment arrangement would also have to be negotiated (assuming the facilitator is not employed by a CSO or an international organization) |

| | | | |
|---|---|---|---|
| 2 | **Pitching the Workshop** | | |
| | 2-3 hours | [Facilitator] | Potential facilitators will want to pitch the training in advance. Facilitators should prepare presentations of various lengths: 15 minutes, 30 minutes, 2 hours, etc…, depending on the degree of detail required. These pitches will be an important process of 'landing' an organization willing to commit to this series of workshops. |

| | | | |
|---|---|---|---|
| **Section 1:** | Introduction to the HRBA@Tech Model and Overarching Principles | | |
| **Total time requirement** | ~2.3 hour workshop<br>+ follow-on "homework" deliberations and decision-making | | |
| 1 | **Intro & HRBA@Tech Principles** | | |
| | 2-3 hours | [Facilitator] | Introductory session (slides 1-32). Necessary to get the workshop participants grounded for the work ahead and familiarized with the overarching HRBA@Tech model. |
| 2 | **HRBA@Tech Chatbot** | | |
| | 1-2 days | [workshop participants] | Working with the customized HRBA@Tech chatbot, participants will want to decide what scope they wish to embrace for their assessment. |

| | | | |
|---|---|---|---|
| **Section 2:** | Prioritizing Processes | | |
| | **Section 2a:** Big Picture Perspective | | |
| | **Total time requirement** | ~6 hour workshop<br>+ follow-on "homework" deliberations and decision-making | |
| 1 | **Welcome Back** | | |
| | ~15 min. | [facilitator] | Reorientation and agenda-setting. |
| 2 | **Reporting Out** | | |
| | ~30-45 min. | [workshop participants] | Workshop participants report out what they learned / discussed, possibly with the help of the HRBA@Tech Chatbot. |
| 3 | **Comprehensive Heatmaps** | | |
| | 15-30 minutes per process (x24 processes total)<br><br>3 hours total *(assuming 4 parallel groups of 3-4 and pre-work)* | [workshop participants, working together with facilitator] | This is a significant time investment. Pending sufficient participant numbers, workshop facilitators can streamline this process by:<br>a. dividing groups into smaller sub-groups of 3-4 participants each, and<br>b. giving groups preparatory "homework" to research the functioning of that process in society, perhaps with the support of the HRBA@Tech Chatbot + independent research. |

| | | | |
|---|---|---|---|
| 4 | **Brainstorming** | | |
| | 1 hour | [workshop participants working in small groups] | Groups will collectively brainstorm areas where the system as a whole could stand to be strengthened. |
| 5 | **Action Planning for needs-based capacity building** | | |
| | 45 minutes | [workshop participants working together with facilitator] | Groups will collectivize their analysis and finalize a list of those processes they wish to prioritize for Phase 3 of the workshop. |
| 6 | **HRBA@Tech Chatbot** | | |
| | 1-2 days | [workshop participants] | Working with the customized HRBA@Tech Chatbot, participants will brainstorm possible system-wide reforms that might build the collective societal capacity to 'nudge' NETs in the direction of human rights. |

| | | | |
|---|---|---|---|
| **Section 2b:** | Organizational Perspective | | |
| **Total time requirement** | ~7 hour workshop<br>+ follow-on "homework" deliberations and decision-making | | |
| 1 | **Welcome Back** | | |
| | ~15 min. | [facilitator] | Reorientation and agenda-setting. |
| 2 | **Reporting Out** | | |
| | ~30-45 min. | [workshop participants] | Workshop participants report out what they learned / discussed, possibly with the help of the HRBA@Tech Chatbot. |
| 3 | **Organizational Heatmapping** | | |
| | 15-30 minutes per process (x total number of remaining processes)<br><br>2-3 hours total *(assuming 4 parallel groups of 3-4 and pre-work)* | [workshop participants, working together with facilitator] | This is a significant time investment. Pending sufficient participant numbers, workshop facilitators can streamline this process by:<br>(1) dividing groups into smaller sub-groups of 3-4 participants each, and<br>(2) giving groups preparatory "homework" to research the functioning of that process in society, perhaps with the support of the HRBA@Tech Chatbot + independent research. |
| 4 | **Brainstorming** | | |
| | 1 hour | [workshop participants | Groups will collectively brainstorm areas where the system as a whole could stand to be strengthened. |

| | | | |
|---|---|---|---|
| | | working in small groups] | |
| 5 | **Capacity \| Mandate \| Organizational Appetite Assessment** | | |
| | 45 minutes | [workshop participants working together with facilitator] | Groups will collectivize their heatmap, identifying (a) processes in which the organization currently is not engaged, and (b) processes where the organization could be engaging more efficiently.<br><br>Next, the group should conduct a *subjective* analysis of those processes flagged where a greater investment would likely fall within the organization's existing mandate. |
| 6 | **Action Planning for needs-based capacity building** | | |
| | 45 minutes | [workshop participants working together with facilitator] | Groups will collectivize their analysis and finalize a list of those processes they wish to prioritize for Phase 3 of the workshop. |
| 7 | **HRBA@Tech Chatbot** | | |
| | 1-2 days | [workshop participants] | Working with the customized HRBA@Tech chatbot, participants will want to use each other and or the HRBA@Tech Chatbot to ideate possible process innovations or reforms the organization could undertake. |

| | | | |
|---|---|---|---|
| **Section 2c:** | Product Perspective | | |
| **Total time requirement** | ~7 hour workshop<br>+ follow-on "homework" deliberations and decision-making | | |
| 1 | **Welcome Back** | | |
| | ~15 min. | [facilitator] | Reorientation and agenda-setting. |
| 2 | **Reporting Out** | | |
| | ~30-45 min. | [workshop participants] | Workshop participants report out what they learned / discussed, possibly with the help of the HRBA@Tech Chatbot. |
| 3 | **Pinpointing a Technology's Progress along a TLC** | | |
| | ~15 minutes | [workshop participants working together with facilitator] | This may be fairly obvious following the previous discussions. |
| 4 | **Checklists & heatmaps (what are we currently doing?)** | | |
| | 30-45 minutes per technology lifecycle "phase"<br><br>1 hour total | [workshop participants, working together with facilitator] | This will be a much narrower discussion, since it focuses on products or technologies that are already in existence, so groups should be relatively efficient. |

| | | | |
|---|---|---|---|
| 5 | **Brainstorming** | | |
| | 1 hour | [workshop participants working in small groups] | Groups will collectively brainstorm areas where the system as a whole could stand to be strengthened. |
| 6 | **Capacity \| Mandate \| Organizational Appetite Assessment** | | |
| | 45 minutes | [workshop participants working together with facilitator] | Groups will collectivize their heatmap, identifying (a) processes in which the organization currently is not engaged, and (b) processes where the organization could be engaging more efficiently.<br><br>Next, the group should conduct a *subjective* analysis of those processes flagged where a greater investment would likely fall within the organization's existing mandate. |
| 7 | **Action Planning for needs-based capacity building** | | |
| | 45 minutes | [workshop participants working together with facilitator] | Groups will collectivize their analysis and finalize a list of those processes they wish to prioritize for Phase 3 of the workshop. |
| 8 | **HRBA@Tech Chatbot** | | |
| | 1-2 days | [workshop participants] | Working with the customized HRBA@Tech chatbot, participants will want to use each other and or the HRBA@Tech Chatbot to ideate possible process innovations or reforms the organization could undertake with regard to a particular product. |

| **Section 3:** | Prioritized List of HRBA@Tech Process Options for Discussion & Planning | | |
|---|---|---|---|
| **Total time requirement** (for each selected process) | ~7 hour workshop<br>+ follow-on "homework" deliberations and decision-making | | |
| 1 | **Welcome Back** | | |
| | ~15 min. | [facilitator] | Reorientation and agenda-setting. |
| 2 | **Reporting Out** | | |
| | ~30-45 min. | [workshop participants] | Workshop participants report out what they learned / discussed, possibly with the help of the HRBA@Tech Chatbot. The focus here will be on concrete process-related ideas participants might like to discuss. |
| 3 | **Product Lifecycle Map** | | |
| | ~30 min. | [workshop participants] | Facilitator asks the group to brainstorm in small groups, and then reports out. |
| 4 | **Stakeholder Role Map** | | |
| | ~30 min. | [workshop participants] | Facilitator asks the group to brainstorm in small groups, and then reports out. |

| | | | |
|---|---|---|---|
| 5 | **Finalizing the Dashboard** | | |
| | ~30 min. | [facilitator w/ input from participants] | The facilitator will bring together all that the groups have learned and discussed. |
| 6 | **Decision time: More Forward or No?** | | |
| | Variable:<br><br>either ~30 min. if during workshop or 1-2 weeks if necessary to consult with colleagues and/or managers. | [participants] | |
| 7 | **Decision Maker Systems Analysis** | | |
| | ~1 hr. | [workshop participants working in small groups] | Stakeholder Mapping process: Part 1. Identifying key stakeholders as well as the influence pathways between those decision makers and other stakeholders in a dynamic system. |
| 8 | **Action Mapping & Consensus Building** | | |
| | ~1-2 hrs. | [workshop participants working in small groups] | Stakeholder Mapping process: Part 2. Designing a unique 'negotiation approach' to interact with key stakeholders in this system. |
| 9 | **Action** | | |
| | Variable | [workshop participants] | During this phase, the workshop participants on their own 'negotiate' an action plan within their organization. |

| | | | |
|---|---|---|---|
| 10 | **Monitoring & Evaluation** | | |
| | ~1.5 hrs. | [workshop participants + facilitator] | Groups will identify measurable indicators that they can use as 'indicators' as success that their process intervention is working as planned. |
| 11 | **Policy Formulation & Capacity Building** | | |
| | 1-2 days | [workshop participants] | Groups will develop necessary paperwork to document how the new process innovation is intended to work, as well as an organizational training and capacity building approach to ensure that all relevant stakeholders are informed of the new process and empowered to play an appropriate role in its function and improvement. |