HRBA@Tech

December 2022

Towards A Human Rights–Based Approach to New and Emerging Technologies: A Framework

Acknowledgments

This report was co-authored by URG and SAPI, with Professor Stephan Sonnenberg (SNU), Louis Mason (URG), Professor Yong Lim (SNU) and Tejaswi Reddy (URG) serving as the primary authors, each of them writing, rewriting, editing, structuring, and restructuring the report countless times. They were assisted by a small army of contributors, teaching assistants, research assistants, peer reviewers, editors, willing colleagues, and project managers, all of whom shared generously their wisdom and critical insights, and all of whose fingerprints appear throughout the report.

First and foremost, we owe an enormous debt of gratitude to Permanent Representative, Ambassador Taeho Lee, Ambassador Seong-mee Yoon, Counsellor Jeong A Yu, and Secretary Yena Yoo of the Permanent Mission of the Republic of Korea in Geneva, who have been so consistently supportive of this project and, more broadly speaking, have exerted efforts at the UN and elsewhere to shine a much-needed spotlight on this important topic. It has been a true joy and privilege to work across diplomatic, civil society, and academic boundaries, and the credit for allowing this to go forward goes to the Permanent Mission of the Republic of Korea in Geneva. In particular, we owe the Permanent Mission of the Republic of Korea in Geneva our thanks for hosting a dynamic and insightful roundtable discussion with experts from the United Nations, various diplomatic missions, think tanks, and academic institutions around Geneva. At that event, we had the honor to receive preliminary feedback and reactions from Ms. Anna Walch, Mr Ioannis Zafeiriou, Mr. David Fairchild, Ms. Chan Sze Zest, Ms. Anniken Enersen, Mr. Preetam Maloor, Ms. Sadhvi Saran, Dr. Sara (Meg) Davis, Mr. Andrés Zaragoza, and Mr. Felix Kirchmeier, to this report. To those who were able to participate at that event, our humble gratitude to you for offering us your insights. We hope we have managed to do your comments justice in the pages of this report.

We are also deeply grateful to Professor Buhm-Suk Baek, who is a leading voice on human rights and who served as the rapporteur for the Advisory Committee Report that gave rise to this effort in the first place. We are deeply indebted to him for his wisdom and insights in this report, and hope that our paper answers, perhaps in a small way, the call he and his colleagues in the Advisory Committee to the Human Rights Council made for further thinking in this area.

We would like to particularly acknowledge the tireless work of URG's small but dynamic team and particularly thank Joseph Burke for his significant contributions throughout the project, repeatedly reviewing, drafting, researching and supporting all elements of the project. We would also like to humbly thank Lola Sanchez and Amalia Ordoñez Vahi for their invaluable contributions and indefatigable team spirit. Special acknowledgement is also owed to Victor Ojeda Gallego for his unwavering professionalism in designing this report to tight deadlines. URG would also like to thank former colleague, Charlotte Marres, for having set the stage for this collaboration and, of course, a particular note of gratitude is owed to URG's Executive Director, Marc Limon, for always having faith in his team and for overseeing this project with his usual incisiveness, clairvoyance and strategic foresight.

On behalf of SAPI, Director and Professor Yong Lim (SNU) would also like to thank its team and affiliates who made this project a reality, at least from the Korean side of the collaboration. Without their patient support and constant substantive enrichment, this feat of transcontinental drafting would have never been possible. We owe a warm note of gratitude to Professor Sanghcul Park (SNU), and Dr. Dohyun Park both of whom contributed their valuable insights and spent many hours reviewing drafts of the report. We would also like to acknowledge and humbly thank the professional research assistance provided by Magdalena Perl, YeonJoo Kim, and Jacob Kovacs- Goodman, all of whom provided valuable background research for this project. As usual, Joonyoung Yoon and SAPI's dedicated staff, provided the necessary logistical and administrative support that made the project run smoothly.

SAPI is deeply grateful to Professor Yong Guk Lee (SNU) and Professor Joan Yoo (SNU), both of whom— perhaps willingly or perhaps not— were dragged into this project, but added layers of richness to our analysis. Professor Yong Guk Lee shared his experience as a longtime corporate lawyer and expert on ESG investing, and generously opened his address book to connect us to several venture capitalists and technologists in the Seoul area. Thanks to Professor Lee's efforts, we were able to learn from Hyun Kim Yong of Envisioning Partners (a Seoul-based Venture Capital firm focused on sustainable entrepreneurialism), Charles Rim of Access Ventures LLC, (a Seoul- based Venture Capital firm focusing on technology startups), Minki Synn of Maven Growth Partners, (a 2-year old Seoul-based Venture Capital firm), and Jung Wook and Jacob Cheon of Kakao Health Care (a subsidiary of Kakao that invests and innovates in the health and technology sector). We would also like to thank Watney Mark and Kim Hye-il of Kakao for their insightful comments about Kakao's impressive efforts to make their services more accessible to differently abled persons. These conversations, each in their own way, were fascinating, honest, deeply insightful, and extremely influential in the framing of this paper. We would also like to thank Amy Lehr, who formerly worked as a Senior Associate at the Center for Strategic and International Affairs and while there authored numerous reports on human rights and technology, for her good advice on how to approach this project.

Professor Joan Yoo contributed substantially to this report, facilitating and co-organizing a Chatham-House Rules round table of social workers, technologists, government specialists, and academics to discuss the themes in this report. Thanks to this roundtable, we were able to learn from Professor Sooyoung Kim, Dr. Youngjin Han, Ms. Jeong-eun Lee, Dr. Eunhee Han, and Dr. Dohyun Park. A warm note of gratitude also for Seyeon Lee and Seong Ha for their invaluable translation, facilitation, and documentation support for this meeting.

SAPI would also like to note that its ongoing collaboration with NAVER, which has led to the adoption of NAVER's AI Ethics Principles and recent roll-out of its new CHEC system for AI governance consultations. This collaboration has greatly informed this project and the content of this report. Without the insights and lessons that were gained from this academic-industry collaboration, this report could well have been divorced from the realities on the ground. We are particularly grateful to NAVER's Agenda Research team, currently led by Woochul Park, that have worked with us on the most difficult issues surrounding AI and data governance.

SAPI would like to further thank the students in Professor Sonnenberg's Human Dignity Clinic (Fall Semester 2022), who tirelessly worked on a variety of research assignments related to this project, often patiently enduring hours of discussion about technology and human rights that somehow – magically – came together at the end of the semester. A warm thank you to Lea Gambier, Josephine (Josi) Oehme, Ayazhan Kazybekova, Carlota Romeu, Oliva Nolte, Tom Dembski, Teun DeGraaf, Esther Park, and Konrad Hünnekens. We would also like to thank students in Professor Sonnenberg and Professor Yoo's class on Rights and Responsibilities for their invaluable contributions to the case study on AI and Social Work, extracts of which appear in chapter 6 of this report. Many of their ideas found their way directly into that chapter, and demonstrate the potential for a bit of creative brainstorming on how to 'nudge' technologies in the direction of human rights. A special thank you to all of the students in that course who worked so hard and contributed their thoughts during that case study.

Finally, SAPI would like to thank Professor James Cavallaro, Co- Founder and Executive Director of the University Network for Human Rights, for his counsel and for facilitating contact with a global network of students interested in this topic. It was through Professor Cavallaro that we were able to engage with a fantastic group of students at Yale University, all of whom were working pro-bono on this project. A warm and heartfelt thank you to Alissa Johnson, Hayoung Choi, Doga Uenlue, Zach Black, and Declan O'Briain for their invaluable research.

Without a doubt, this list is incomplete. For all those whom we forgot to mention, our humble apologies for that, and our deepest gratitude nonetheless for your kind support on this project.

Table of Contents

The Brief

Introduction, Methodology, Executive Summary

Introduction

The interrelationship between science, technology (i.e., the application of scientific knowledge to the world around us) and the fundamental human dignity of individuals and communities has been part of the modern human rights movement since its birth in the late 1940s. The urgency of that discussion was as important in the shadow of the mushroom clouds over Hiroshima and Nagasaki as it is today. As new and emerging technologies increasingly impact every aspect of human life, they offer the potential to strengthen the promotion and protection of human rights. However, they also pose complex risks and challenges that could hinder the full and effective enjoyment of these rights. The relationship between technology and human rights is a paradox: on the one hand providing opportunities for social innovators to advance the cause of human rights and accelerate sustainable development, while often simultaneously representing a constant threat to our enjoyment of human rights.

Given this paradox, how should the human rights community situate itself vis-à-vis new and emerging technologies? Must we make the uncomfortable choice between two ultimately fundamentalist positions: the 'luddite,'[1] or precautionary stance on the one hand, whereby we must reject all technological progress unless we are guaranteed that it can be deployed with no risk to existing human rights and societal structures, or the 'tech utopianist'[2] stance on the other, in which we embrace an understanding of what it means to be human where we as a species are constantly influenced, altered and enhanced by the very technologies that we bring into existence?

To answer this question, one needs first to answer the question of whether technology can be said to be neutral with regards to moral, ethical, or human rights values. It has often been argued that technology itself is neutral–a mere instrument for human beings to exercise their own personal agency. This is the so-called Value- Neutrality Thesis (VNT). The pithy phrase "guns don't kill, people kill,"[3] succinctly embodies the VNT. The VNT is often advanced by those who would resist efforts to regulate or limit the development and diffusion of new technologies. The thesis relies on the claim that technology on its own (the 'artifacts' of technology) do not exude empirically verifiable values, and that therefore the technology must be seen as being either neutral, or at worst infused with only trivial and largely imperceptible values.

We explicitly reject the idea of tech neutrality in this paper.

If it were any other way, Part II of this paper, in which we propose a strategy for infusing human rights values into NETs, would not have been necessary. If we considered technology to be the mere extension of human agency, then our recommendations on how to structurally 'nudge' technologies in the direction of human rights would be meaningless. In rejecting the VNT, we join a growing list of scholars, including the members of the Advisory Committee to the UN Human Rights Council,[4] who also believe that technology itself can be imbued with values.

Conducting an analysis of which values might be embedded within a technology requires a process of 'wondering, deliberating, and reasoning.'[5] This is an inherently subjective process that 'registers what is perceived in relation to categories, concepts, and classes that are socially produced.'[6] The HRBA@Tech model presupposes that one obvious - perhaps the obvious - assortment of such 'categories, concepts and classes' is the human rights corpus. Human rights, we argue, can and should be used as the barometer to thread a 'middle path' between the 'luddite' and the 'tech-utopian' stance with regard to NETs. According to this more holistic and moderating approach, the benefits of technology can and should be exploited for the equitable and universal advancement of human rights -- to make the world a better place. But such use should always take place within clear legal and normative limits, based on universally accepted human rights standards, that serve to ensure that the development and deployment of NETs will not violate individuals' fundamental human rights. It recognizes that while technological progress is rapid and exhilarating, there must always be an emergency brake as well.

This middle path posits a two-pronged strategy on how to think about and respond to NETs. The first, premised in the classical human rights philosophy of avoiding harm, posits that technologies should be designed and deployed in a manner to allow for the minimization and remediation of any potential negative social and human rights impacts of a new technology. The second, more positive approach, aims to encourage the development of promising technologies that maximize the likelihood that they will benefit humanity as a whole.

The international community has long wrestled with the implications of technological innovation. While some are optimistic about its potential to advance social progress, others are more cautious, focusing on possible negative societal impacts. The accelerated pace of digital transformation has intensified the need for a comprehensive governance framework to manage both the positive and negative effects on individuals and their rights. In this context, the international human rights system has been increasingly mobilized to provide guidance to all stakeholders on how to best address the human rights implications of new technologies. In 2019, the Human Rights Council notably adopted a resolution on 'New and Emerging Digital Technologies and Human Rights.' This resolution tasked the Advisory Committee with preparing a report to examine the impacts, opportunities, and challenges that such technologies present for human rights. The report aims to offer insights on how these issues can best be addressed within the international human rights framework.

The landmark report, presented to HRC47 in June 2021, highlights conceptual and operational gaps within the existing international human rights framework that act as barriers to harnessing the potential and benefits of such digital technologies while preventing associated risks. Conceptual gaps identified include challenges in adapting the existing international human rights framework to the current realities of the digital age. There is also a noticeable lack of cooperation and coordination between the human rights and technology communities, leading to an insufficient understanding of how the two fields intersect. Additionally, there is a selective focus on certain technologies and specific human rights harms, which can overshadow other equally important issues. Meanwhile, operational gaps identified include practical challenges resulting from innovation outpacing regulation and the fragmentation of regulatory initiatives that lead to governance gaps, as well as the lack of adequate resources to support existing human rights mechanisms. To tackle these challenges, the Advisory Committee recommended that stakeholders adopt a human rights-based approach driven by multi-stakeholder cooperation and based on three pillars-(1) developing a holistic understanding of technology; (2) developing a holistic approach to human rights; and (3) developing holistic governance and regulatory efforts.[7]

Against this backdrop, the present report aims at tracing the contours of such a working methodology of a human rights-based approach to NETs (the HRBA@Tech Model). Part I of the report lays out the intellectual and normative foundations for the HRBA@Tech model. Since this report is ultimately aimed at the development of a new normative approach to NETs, Chapter 1 begins with a historical overview of existing efforts to develop such norms, starting with the Universal Declaration of Human Rights in 1948. This Chapter shows how many of the debates we are having today have

1. The Luddites were a movement of angry workers who began smashing newly-invented textile machinery in factories around central England. Luddites are commonly described as being against new technologies, but as one scholar notes, the Luddites were less concerned about the new technology per se and more worried about how that new technology might be used to undermine labor protections of the time. "People of the time recognized all the astonishing new benefits the Industrial Revolution conferred, but they also worried [...] that technology was causing a 'mighty change' in their 'modes of thought and feeling. Men are grown mechanical in head and in heart, as well as in hand.' Over time, worry about that kind of change led people to transform the original Luddites into the heroic defenders of a pretechnological way of life." See Richard Conniff (2011), "What Luddites Really Fought Against." Smithsonian Magazine (March 2011), https://www.smithsonianmag.com/history/what-the-luddites-really-fought-against-264412/. (Quoting 19th Century Scottish essayist Thomas Carlyle).

2. Technological Utopianism consists of "an ironclad faith in technology and its ability to solve any kind of problem, which would enable people to live in a sort of utopia." See Alberto Tundo (2020), What is Technological Utopianism? Maize, https://www.maize.io/news/ technological-utopianism/

3. Joseph Pitt (2014). "Guns Don't Kill, People Kill": Values in and/or Around Technologies. In Peter Kroes and Peter-Paul Verbeek (eds) The Moral Status of Technical Artefacts. Philosophy of Engineering and Technology, vol. 17. Springer, Dordrecht, 90

4. Human Rights Council Advisory Committee (2021) "Possible impacts, opportunities and challenges of new and emerging digital technologies with regard to the promotion and protection of human rights", Art. II (4), UN Doc. (May 19, 2021), A/HRC/47/52, New York, NY: United Nations

5. Boaz Miller (2021) "Is Technology Value-Neutral?" 46(1) Science, Technology & Human Values 53-80, 55 (summarizing the argument put forward in Joseph Pitt, id., at 90).

6. Ibid., (quoting Helen Longino (2002). The Fate of Knowledge. Princeton, NJ: Princeton University Press, 100).

7. Human Rights Council Advisory Committee (2021) "Possible impacts, opportunities and challenges of new and emerging digital technologies with regard to the promotion and protection of human rights", at Art. VII UN Doc. (May 19, 2021) A/HRC/47/52, New York, NY: United Nations

precedent in earlier decades, suggesting that the lessons learned from those conversations in the past may still hold relevance for us today. Chapter 2 shifts from a normative towards a policy stance, illustrating the fundamental paradox of NETs by specifically highlighting four NETs (genetic engineering, internet-based technologies (including social media), geoengineering, and artificial intelligence). Chapters 1 and 2 collectively provide the context for the HRBA@Tech model.

Part II of this report presents the HRBA@Tech model. This is presented merely as a starting point for possible future work, to be taken forward by, for example, a newly created Special Procedures mandate of the Human Rights Council. It brings thinking from the world of human rights, sustainable development, and technological ethics into one framework that aims to be broad enough to be applicable not just to one specific technology, but rather a diverse set of new, emerging, and even as yet unimagined future technologies. The model does not therefore set forth a fixed set of universally applicable standards, but rather proposes a process by which technologists, human rights actors, and affected individuals and communities can jointly define and promote specific standards that are meaningful in the context of a given new technology. Chapter 3 sets out the basic principles of HRBA@Tech, as derived from international human rights law and technology ethics. It draws from previous models of human rights-based approaches, notably in the context of development, which it complements with current discussions surrounding ethics of technology. Chapter 4 breaks down how these principles might be integrated into the typical life cycle of any given technology. It describes the process through which technologists and human rights analysts might identify human-rights relevant intervention points within that life cycle. Chapter 5 focuses on the specific stakeholders or actors who can then mobilize (or be called into action) to make real the promise of the HRBA@Tech model, often working together in multi-stakeholder initiatives. Chapter 6 illustrates the HRBA@Tech model by hypothetically applying it in the context of a project involving the development and deployment of an Artificial Intelligence driven project to advance social justice. This Chapter demonstrates that many of the principles outlined in the HRBA@Tech model are already embraced as 'best practices' by many in the tech community.

Part III concludes the report with a series of recommended actions for the international community, individual States, and other relevant actors to take in order to continue advancing this agenda forward.

Methodology

This document is the result of a collaboration between the Universal Rights Group, the Seoul National University Artificial Intelligence Policy Initiative, supported by the Permanent Mission of the Republic of Korea to the United Nations in Geneva (MoFA-Genève), that aims to propose a working concept of a human rights-based approach to NETs, which can ideally serve as a starting for future human rights and technology experts—and ideally a Human Rights Council Special Procedure to be established in coming years.

The Universal Rights Group (URG) is an independent think tank dedicated to analyzing and strengthening global human rights policy. Its main office is in Geneva, Switzerland and it also has offices in New York City, Bogotá, Colombia and Nairobi, Kenya. URG sits at the center of a wide international network for academic and research institutions, human rights defenders, and NGOs committed to furthering human rights. URG supports and strengthens policy- making and policy-implementation at the international, regional and local levels by providing rigorous yet accessible, timely and policy-relevant research, analysis and recommendation. The Group also seeks to provide a forum for discussion and debate on important human rights issues facing the international community and a window onto the work of the Human Rights Council and its mechanisms – a window designed to promote transparency, accountability, awareness and effectiveness. The Group is registered as a not-for-profit association under Swiss law.

The Seoul National University Artificial Policy Initiative (SAPI) was launched in 2017 to study social and policy challenges that are likely to arise in response to the expected proliferation of data-driven artificial intelligence (AI) technology. SAPI conducts interdisciplinary research bringing together specialists from the fields of technology development, humanities, social science, and the law. SAPI sees itself as a 'Social Laboratory': a research platform where various disciplines from around Seoul National University, ranked as Korea's top university, come together to discuss these crucial issues. SAPI is leading research efforts in the field of artificial intelligence & law, and is actively engaged in joint research and cooperation with domestic and international experts and institutions in various fields, including human rights, AI & ethics, and compliance.

In mid-2022, the Permanent Mission of the Republic of Korea in Geneva generously commissioned URG and SAPI to begin work on a series of policy papers on Human Rights and New and Emerging Digital Technologies. This is the first of those policy papers. Subsequent installments in this series of policy papers will build upon this initial publication. They will apply the developed approach to specific issues in the field, such as human rights in the context of climate change and new technologies, accountability mechanisms, Agenda 2030, and more.

Several activities inform the contents of this paper. SAPI and URG both conducted extensive desk reviews of existing literature in this field, prioritizing English as well as Korean-language materials in that process. SAPI also conducted numerous key-informant interviews in Seoul, South Korea, specifically with technologists, financiers, scholars, and industry insiders, while URG broadly consulted Geneva- based stakeholders from diplomatic missions, academia, civwwwil society and United Nations entities. While some of these interviewees wished to remain anonymous, all focused on the broad conceptual task of designing a framework for the promotion of human rights that would prove acceptable to multiple stakeholders involved in the development and promotion of NETs. SAPI also hosted a closed-door roundtable of technologists, policy makers, academics, and social workers involved in an effort to develop an AI-driven tool to help improve social services in South Korea. This workshop provided SAPI with a unique insight into the kinds of discussions, and processes, that inform such an effort to harness the power of NETs to combat a high-profile (and high-stakes) human rights issue.

SAPI also leveraged its location in a world-class research institution to bring this project into a classroom, intentionally using those fora as workshops to refine and elaborate on the concepts described in these pages. Students enrolled in the Human Rights and Dignity Clinic at Seoul National University[8] spent the entire 2022 Autumn semester researching ten separate NETs.[9] They used an iterative process to refine and debate crucial aspects of this paper, contributing to its depth and rigor. Students and faculty in that class worked iteratively to refine the HRBA@Tech model presented in this thought piece, seeking to test its appropriateness for any NET—even those we cannot yet even imagine in 2022. Likewise, students enrolled in an interdisciplinary undergraduate course on Rights and Responsibilities spent three weeks working on a case study that closely mirrored the content of the closed-door SAPI roundtable on Social Work and AI. SAPI also partnered with a group of volunteer students from Yale University, facilitated by the University Network for Human Rights, who conducted additional interviews with technologists and human rights researchers, primarily in North America, that further informed the analysis in this thought piece.

On November 24, 2022, URG and SAPI jointly hosted a policy dialogue in Geneva. This closed-door session brought together 20 policymakers, including representatives of Permanent Missions, civil society, UN entities, international experts, and other relevant stakeholders to discuss the proposal for the HRBA@Tech developed in this report. The policy dialogue provided an opportunity to validate and critique initial findings and to solicit additional input from the assembled stakeholders.

This policy report is a work in progress. Its authors are by no means "done" with this analysis. We sincerely hope, however, that our analysis provides a robust starting point for the focused development of universal norms that speak to NETs and the challenge of harnessing them as a force for the protection and more widespread enjoyment of human rights worldwide. Further, we propose that the methods of multi-stakeholder and multi-disciplinary engagement, dialogue, and iterative conceptual refinement that we used to inform this report might also provide a template for that future process of normative refinement.

8. For more on the Human Rights & Dignity Clinic, see http://slcc.snu.ac.kr/page/legalclinic_list.php (accessed Nov. 12, 2022).

9. In addition to Artificial Intelligence, the Human Dignity Clinic also researched genetic engineering & genomics; geo engineering; extended, virtual & augmented reality; 3D printing; new energy solutions; robotics; quantum computing; blockchain; and 'internet-of-things' technologies.

Executive Summary

‘Technology is not neutral.’

This statement increasingly represents the position of technologists, human rights campaigners, ethicists, and political scientists. According to this emerging consensus, technology itself can in fact be instilled with certain values - either in line with or contradictory to human rights norms. This simple conclusion is not merely descriptive, however, but also empowering. It empowers us – in our capacity as diplomats, policy makers, technologists, corporate managers, entrepreneurs, innovators, bureaucrats, civil society activists, students, professors, consumers and regular citizens – to guide technology and bend the ark of its life cycle in the direction of social justice and the promotion and protection of human rights.

Oftentimes, a new or emerging technology holds obvious promise from the perspective of advancing human rights. Without modern technology, for example, the world would have never been able to keep so many of our schools and hospitals running during the COVID-19 public health emergency, a simple fact that helped millions of children and patients continue to enjoy their rights to education and health in the midst of a global pandemic. Modern technology promises to cure some of society’s most vexing problems, including many that have the direct potential to advance the cause of human rights, including by creating entire new economies with the potential to lift millions out of poverty.

Confusingly, however, some of those same new and emerging technologies can also be used to undermine human rights. Technologies that typically facilitate global social engagement can also be used for surveillance purposes or can inadvertently undermine social cohesion and democratic institutions. Genetic engineering promises to cure certain kinds of hitherto incurable diseases, but can likewise worsen existing inequalities. Technologies with undeniable individual benefits can result in negative social externalities, while others that clearly improve societies can do so at the expense of individuals and their rights. Technology can be fungible: designed to serve a virtuous purpose and then nonetheless subverted or used by others to serve more ethically ambiguous (or outright nefarious) purposes.

If technology is not neutral, and if there is this consistently foreseeable risk of bad or ignorant actors deploying technology for less-than-noble ends, the burning question becomes how can we do everything in our power to make it more likely than not that technologies are beneficial to individuals, communities and humanity, while minimising and countering some of their inherent potential to do harm? Is there a method by which technologies, especially new and emerging technologies, can be ‘hard wired’ or ‘genetically engineered’ to serve pro-social causes that respect, protect and fulfil human rights?

In this policy report, we propose such a method, which we are calling the Human Rights-Based Approach to New and Emerging Technologies (HRBA@Tech, for short).

It was developed jointly by the Universal Rights Group (URG), a human rights think tank dedicated to proposing research-based policy prescriptions to strengthen human rights policy, with offices in Bogota, Nairobi, Geneva and New York, along with the Seoul National University Artificial Intelligence Policy Initiative (SAPI), an interdisciplinary research laboratory at Seoul National University in Korea devoted to the interdisciplinary exploration of issues having to do with artificial intelligence. In developing this model, both URG and SAPI drew on their comparative sources of experience, building on their relationships with the diplomatic, academic, and technical communities to develop and refine the model.

The HRBA@Tech model integrates three different perspectives, each of which brings unique insights to the questions posed above. First, it explores the principles that are most relevant in the development of new and emerging technologies (“the What” of the HRBA@Tech model). Drawing from both human rights standards as well as classical technological ethics literature, the paper articulates seven interlocking principles, organized into two pillars. The first (the ‘do no harm’ pillar) postulates that new and emerging technologies should be accountable, secure, non-discriminatory, and grounded in international human rights law. The second (the ‘make the world a better place’ pillar) additionally specifies that new and emerging technologies should also be based on proactive representation of, transparency towards, and the empowerment of those who would likely be impacted by new and emerging technologies.

Each of these principles is associated with a list of discrete processes, for example ‘consultation,’ ‘human rights by design,’ or ‘capacity building.’ The articulation and study of these discrete processes – of which we argue there are a total of 24 – is the true added value of this HRBA@Tech model, since it suggests actionable ways to make real the lofty principles of the HRBA@Tech model in an applied, real- world setting.

Second, the paper describes how these principles and corresponding processes apply to new and emerging technologies along the course of a classical technology lifecycle, starting from the initial days of innovation all the way to a technology’s eventual irrelevance. We call this ‘the How’ of the HRBA@Tech model. The report shows that certain processes tend to be far more relevant at certain stages of the technology’s lifecycle than others.

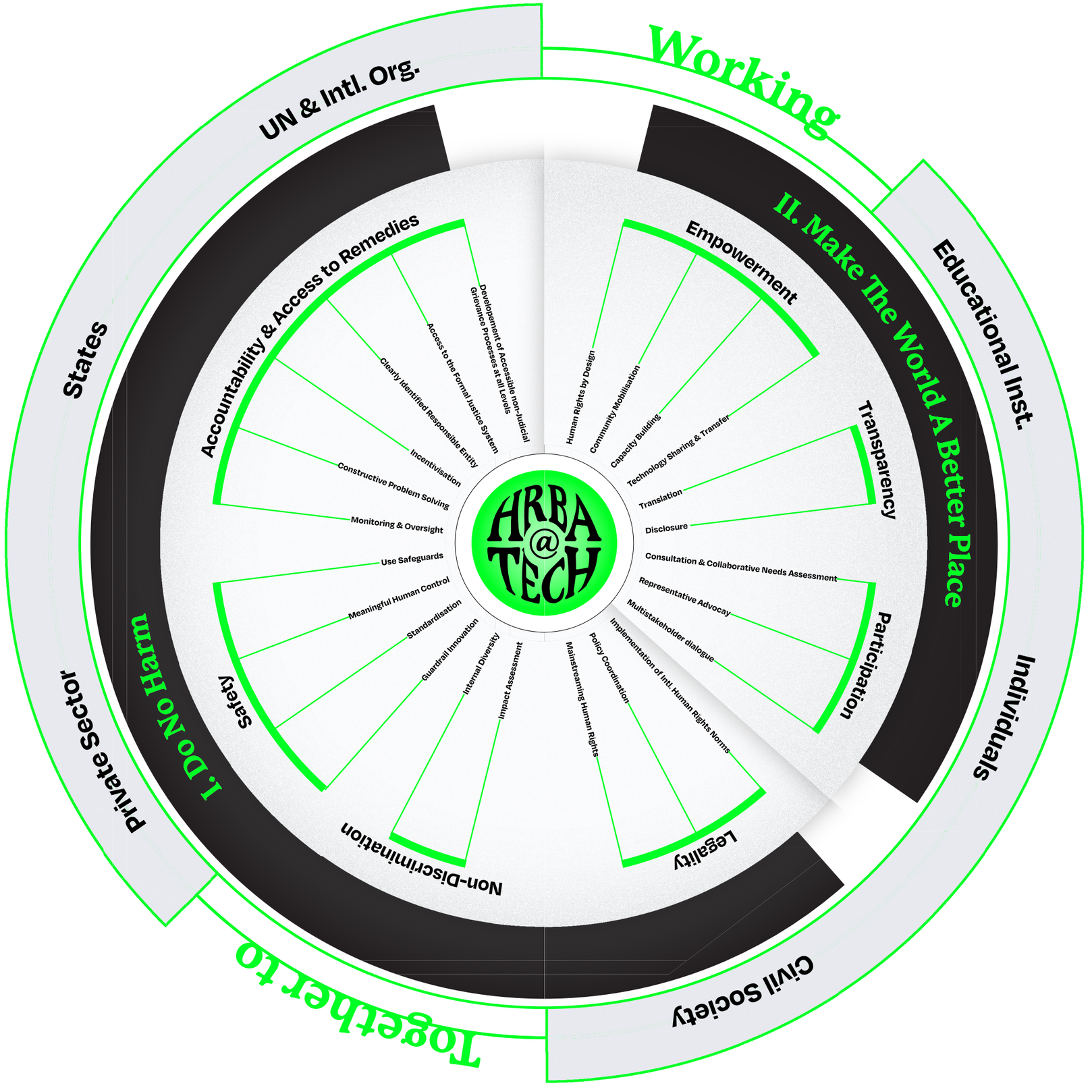
The third and final perspective is that of the stakeholders (‘the Who’ of the HRBA@Tech model). Here we ask what stakeholders are relevant, working in which coalitions, to drive the HRBA@Tech model. While these descriptions remain somewhat abstract, the discussion highlights how various stakeholders must learn to work not only in opposition to one another, but also join hands in collaborative problem solving to make headway on this issue.

The paper concludes by illustrating how the HRBA@Tech model would apply to artificial intelligence, highlighting how aspects of this approach are already being utilized by ethically-minded technologists, academics, government officials and corporate managers to nudge technologies in the direction of the promotion and protection of human rights.

The final Chapter of the paper distils those findings into a succinct list of recommendations.

Our intention in these pages is to bring together the various strands of technological ethics underneath the umbrella of a human rights- based approach, using accessible language to do so and highlighting specific processes that will be essential in any such efforts. We especially hope that this paper will contribute to efforts at the United Nations and elsewhere at the international level to use human rights mechanisms to systematically and progressively advance the issue of technology and human rights. Such a concerted effort is sorely needed, and we believe the diplomatic common ground exists to gradually develop implementable standards that can provide tangible guidance for all those who are involved in the process of designing and developing New and Emerging Technologies (NETs).

Model of the
HRBA@Tech



Part I

New and Emerging Technologies and Human Rights: Setting the Stage

> "The promise which science offers is understandably high but having invented and perfected the machine, is man going to become himself the slave of the machine or of those few in number who will be in the position to manipulate it?"
>
> *UN Secretary General U Thant*
> *Address to the Tehran Conference (1968)*

> "It [is] important not only to study the measures against intrusions into private life but also to guarantee other important human rights within the framework of the progress of science and technology."
>
> *Debate on the UN Human Rights Commission Resolution on Human Rights and Scientific and Technological Developments (1971)*

CHAPTER 1: HUMAN RIGHTS AND NEW AND EMERGING TECHNOLOGIES THROUGH THE YEARS: AN EXPANDING FIELD

Chapter Summary:

Though human rights issues arising out of new and emerging technologies do signify a certain modern and contemporary dimension, debates surrounding the relationship between human rights and technology have long been in existence, including the early and nascent years of the human rights field. This Chapter traces the interplay of human rights and technology through history, delves into the ways in which the field has expanded and adjusted its boundaries to accommodate technological evolutions, particularly advancements in digital technologies, and lastly, provides an overview of more recent developments with respect to new and emerging digital technologies. An understanding of this context offers useful insights to present-day human rights dilemmas related to new and emerging digital technologies and is foundational to the development of a holistic human rights-based approach, as recommended by the Advisory Committee in its seminal 2021 Report to address such issues.

The Universal Declaration of Human Rights (UDHR) was adopted in 1948 against the background of the devastation caused during the Second World War. That war, and the tremendous human suffering that characterized it, was made possible to a great extent by scientific and technological advancements, including the use of atomic bombs and weapons of mass destruction with the potential to jeopardize the very existence of humanity. In an embodiment of the profound lessons learnt during the Second World War, and in recognition of the need to share technologies equitably and prevent their monopolization by one nation or group of nations at the expense of other peoples, Article 27 of the UDHR provides for "the right freely to [...] share in scientific advancements and its benefits." A similar provision finds place in the International Covenant on Economic, Social and Cultural Rights (ICESCR), adopted in 1966. Article 15.1(b) proclaims "the right to enjoy the benefits of scientific progress and its application." The assumption and premise underlying these provisions was that science and technology (as an application of science) must only be used for the advancement of benefits and progress for humanity,[10] and that said advancement should benefit all people.

Despite this prominent incorporation of science and technology into the heart of the international human rights framework, these provisions are rarely invoked within the mainstream human rights movement, and "science [remains] one of the areas...to which states parties give least attention..."[11] The world today looks very different from the world in which most of these "core" international human rights instruments were adopted, and some of the human rights issues unique to NETs (such as AI, genetic engineering, geo- engineering, and modern ICT technologies) may not have been imagined by the original drafters, and yet the broadly worded guarantees of human rights enshrined in these core instruments, which are universal and in-temporal in nature still serve as a powerful starting point for addressing issues in the human rights and technology space even today. The international human rights system also evolved to reflect current technological developments over time. Many of these developments were specifically premised on the human rights and emerging digital technologies paradox, i.e., the realization that technological advancements pose both challenges and opportunities for human rights. Though these trends reflect a growing consciousness of the human rights implications of scientific and technological developments since the 1940s, initial efforts were narrow in scope and limited to specific aspects of human rights such as privacy, data protection, freedom of speech and expression and other such issues. In recent years however, efforts have also been directed towards addressing the human rights implications of digital technologies in a more general, overarching, and holistic manner.

In 1968, on the occasion of the 20th anniversary of the UDHR, the United Nations General Assembly (UNGA) convened an international conference on human rights in Teheran to reflect on the progress made in the field of human rights and also to formulate a program of action for the coming decades.[12] While the interrelationship between technology and human rights was certainly not at the forefront of the conference agenda, it was on this occasion that the human rights implications of scientific and technological developments, including the specific dangers posed by them, were for the first time specifically highlighted.[13] The resolution on 'Human Rights and Scientific and Technological Developments', adopted at the end of the Teheran Conference,[14] acknowledged that scientific and technological developments had opened new opportunities for economic, social and cultural progress, but also recognized that they posed complex human rights challenges. This duality (the paradox) highlighted at the Teheran Conference has remained central to the human rights and technology discussion ever since.[15]

10. Haochen Sun, Reinvigorating the Human Right to Technology, accessed at: https://repository.law.umich.edu/cgi/viewcontent.cgi?article=2090&context=mjil ; See Johannes Morsink, The Universal Declaration Of Human Rights: Origins, Drafting And Intent 217 (1999).

11. Committee on Economic, Social and Cultural Rights, General Comment No. 25-Article 15(1)(b), (2), (3) and (4): Science and Economic, Social and Cultural Rights (2020), (Article I.2).

12. UNGA Resolution XX (1965).

13. Christopher Weeramantry, Science, Technology and the Future of Human Rights, accessed at: https://www.jstor.org/stable/pdf/23001436.pdf?refreqid=excelsior%3Af9547523726ec13ba463bfd6b10e7881&ab_segments=&origin=&acceptTC=1.

14. Final Act of the International Conference on Human Rights, Teheran, 22 April to 13 May 1968 (New York, 1968), A/CONF.32/41, Res XI p. 12.

15. Steven L.B. Jensen, The 1968 United Nations Debate on Human Rights and Tech, Open Global Rights (30 August 2022); accessed at https://www.openglobalrights.org/the-1968-United-Nations-debate-on-human-rights-and-tech/; Kinfe Ilma, Emerging Technologies and Human Rights at the United Nations (2021), accessed at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3998498

Though the then-UN Secretary General U Thant in his opening address urged the conference attendees to find "the ways and means of turning science and technology from destruction to the enhancement of life",[16] the resolution adopted succeeded only in calling for further study of the issue. More specifically, the resolution emphasized four key focal points for this future research:

1. "Respect for privacy in view of recording techniques;
2. Protection of the human personality and its physical and intellectual integrity in view of the progress in biology, medicine and biochemistry;
3. The uses of electronics which may affect the rights of the person and the limits which should be placed on its uses in a democratic society;
4. More generally, the balance which should be established between scientific and technological progress and the intellectual, spiritual, cultural and moral advancement of humanity."[17]

Later that same year, the UNGA, on the basis of the recommendations adopted in the above resolution, and emphasizing the need for human rights standards in the area of scientific and technological developments, mandated the UN Secretary General (UNSG) to undertake a study into the four issues highlighted above as well as the potential impact of new technologies on national security.[18]

In the resulting report, which was issued in 1970,[19] the UNSG noted that the "explosion of scientific knowledge and of its technological application which has taken place has not been accompanied by an equally urgent and profound consideration of the implications thereof for human rights."[20] The report provided examples of some ways that governments addressed these challenges, predominantly highlighting legislative and regulatory efforts to ban or criminalize technologies that posed threats to privacy rights, security, and the right against self-incrimination, amongst other rights.[21] This focus reflected a general bias in favor of state action to prevent human rights abuses or violations. Other threats, including less tangible or unintentional threats posed by technological development, as well as issues related to the role of private actors and corporations, remained relatively unexplored.[22] The UNSG report mapped the human rights implications and relevant existing standards with regard to the four thematic areas highlighted by the Teheran Resolution, but also acknowledged situations where those existing standards may not be sufficient in their application. It concluded by calling for new standards to be developed that would better protect human rights. In light of the rapid evolution of technology at the time, the UNSG report also emphasized the need for flexible approaches to accommodate the pace of innovation, and called for the UN to serve a standard-setting role.

This early strand of thinking at the UN primarily focused on the "downsides" or risks of technology, positing technology as a potential threat to be controlled. This precautionary or defensive approach to technologies developed alongside and coexisted with another emerging strand of soft law that focused more on the "upsides" or opportunities of technology, specifically the progressive potential of new technologies, especially in regions that today we might describe as the "Global South." This approach had its landmark moment in 1963 with the UN Conference on the Application of Science and Technology for the Benefit of the Less Developed Areas.[23] In stark contrast to the tenor of the Teheran Conference, the chairman of the 1963 conference in his keynote address reminded the participants that "applied science can be the most powerful force in the world for raising living standards if action can be taken to harness it for that purpose-if the Governments and people of the world can find the means and the will."[24]

The 1963 conference led the UN Economic and Social Council (ECOSOC) later that same year decided to establish the Advisory Committee on the Application of Science and Technology to Development (ACAST)-which, as the name implies, couched the discussion of the positive and beneficial aspects of technology within a development framework.[25] A series of events following ACAST's advocacy and periodic reporting led to the UN Declaration on Social Progress and Development in 1969,[26] which embraced both the need to protect human rights as well as the desire to harness the dynamism of scientific collaboration and dissemination of technology as vital components of an emerging development agenda. It noted the potential of science and technology to contribute to and meet "the needs common to all humanity",[27] and called for its increased utilization for the benefit of social and economic development. It also specified that "[s]ocial progress and development shall be founded on respect for the dignity and value of the human person and shall ensure the promotion of human rights and social justice."[28]

A similar convergence was taking place in the UN Human Rights Commission, which in 1971 considered the report that had been previously submitted by the UNSG on Human Rights and Scientific and Technological Developments and a resolution purporting to take action on that report. "The UNSG report had faced withering criticism by some members of the Human Rights Commission, much of which remains relevant even today. Some voiced their "disappointment" that the report drew on "material collected only from limited areas of the world, especially the countries more developed from a technological point of view."[29] Those same critics lamented that the report focused only on technology's potential threat to human rights and not the "various other aspects of human advancement."[30] Critiques accused the UNSG of highlighting issues with "little or no social significance", relying on "too much speculative material."[31] The critics concluded that "science and technology should not restrict the rights of the individual, but also that science should be used in the interests of society as a whole, and not to increase social and property inequality or to intensify the exploitation of man by man."[32] "It was important", members noted, "not only to study the measures against intrusions into private life but also to guarantee other important human rights within the framework of the progress of science and technology."[33]

The amended resolution, which embraced both the potential risks and the potential opportunities that technologies posed for human rights, was adopted by the Commission on Human Rights at its 27th session on 18 March 1971.[34] The Commission called for presenting a balanced picture of the human rights challenges arising out of scientific and technological developments, including the ways in which they can be used for the benefit of mankind and interest of society as a whole, especially in developing countries. It urged the use of science and technology to improve living conditions and foster respect for human rights, and not to increase social inequality or intensify exploitation of man or to restrict human rights and fundamental freedoms. Interestingly, the Commission also noted that science and technology were in themselves "neutral"[35], and that the problems they pose or advantages they offer for mankind emanate from the use to which they

16. Address of the Secretary General, U Thant, Final Act of the International Conference on Human Rights, Teheran, 1968, at 37.

17. Resolution on Human Rights and Scientific and Technological Developments, Final Act of the International Conference on Human Rights, Teheran, 1968, at 12.

18. Ibid.

19. UN ECOSOC, Human Rights and Scientific and Technological Developments- Report of the Secretary General (26 February 1970) E/ CN.4/1028, E/CN.4/1023/add.1, E/CN.4/1023/add.2, E/CN.4/1023/add.3, E/CN.4/1023/add.3/Corr.1, & E/CN.4/1023/add.4

20. UN ECOSOC, Human Rights and Scientific and Technological Developments- Report of the Secretary General (26 February 1970) E/ CN.4/1028 at 5.

21. UN ECOSOC, Human Rights and Scientific and Technological Developments- Report of the Secretary General (26 February 1970) E/ CN.4/1028, E/CN.4/1023/add.1 & E/CN.4/1023/add.2.

22. UN ECOSOC, Human Rights and Scientific and Technological Developments- Report of the Secretary General (26 February 1970) E/ CN.4/1023/add.3, E/CN.4/1023/add.3/Corr.1, & E/CN.4/1023/add.4

23. Science and Technology for Development: Report on the United Nations Conference on the Application of Science and Technology for the Benefit of the Less Developed Areas; Volume VIII-Plenary Proceedings, List of Papers and Index, New York: United Nations, vii.

24. Ibid.

25. Questions relating to science and technology. UN. Economic and Social Council (36th sess. : 1963-1964: New York and Geneva), 1963

26. UNGA Res 2542 (XXIV), UN Declaration on Social Progress and Development (11 December 1969); accessed at: https://www.ohchr. org/en/instruments-mechanisms/instruments/declaration-social-progress-and-development

27. Ibid, Preamble

28. Id. Article 2

29. Commission on Human Rights, Report on the Twenty-Seventh Session (22 February-26 March 1971), (New York, 1971) E/4949, E/ CN.4/1068, par.181, p.41

30. Ibid., par.182, p.41

31. Ibid., par.183, p.42.

32. Ibid., par.189, p.43

33. Ibid., par.189, p.43.

34. Commission on Human Rights, Report on the Twenty-Seventh Session (22 February-26 March 1971), (New York, 1971) E/4949, E/ CN.4/1068

35. Ibid., par. 179, p.40.

are put. The Commission decided to retain the issue of technological developments as a standing item on its agenda to continue examining the human rights implications of scientific and technological developments. In recognition of the complexity of problems related to technology due to the rapid and unpredictable nature of developments, it called for "constant attention" to accommodate changing realities and the need to adapt accordingly.

These early discussions in the 1960s and 1970s encapsulated the tensions and fault lines that even today continue to define the debate on human rights and NETs. These include debates about tech neutrality (see introduction), as well as debates about whether to embrace a more precautionary stance towards the potential impact of technologies, how best to embrace their potential to facilitate human progress and development, whether and how to prioritize the promotion of civil and political rights vis-a-vis the progressive realization of economic, social and cultural rights, and how to deal with the potential social and economic disruptions caused by technological advancements (for example, the elimination of entire sectors in light of "creative destruction" by new technological developments).

The duality of the relationship between human rights and technologies continued to be noted over the years. The 1975 Declaration on the Use of Scientific and Technological Progress in the Interests of Peace and Benefit of Mankind,[36] for example, emphasized the need for international cooperation, the facilitation of greater transfer and exchange of technologies, capacity building in developing nations, and ensuring that the benefits of technology are enjoyed by all strata of society. In 1983, the Commission on Human Rights once again set out to identify concrete strategies to harness the potential of scientific and technological achievements in the realization and promotion of human rights and fundamental freedoms,[37] and again reiterated the duality of the relationship between human rights and technology.[38] The 1993 Vienna Declaration and Programme of Action[39] again noted the potential adverse human rights consequences of scientific and technological advances, notably in the field of information technology. The subsequent United Nations Millennium Declaration[40] announced a major new global push to harmonize international development standards and affirmed the need to ensure that the benefits of new technologies, especially information and communication technologies, be available to all.

Though the turn of the millennium was marked by rapid technological advancements, particularly in the fields of information and communication technology and in the health sector (see Chapter 2), the human rights implications of those technological developments were addressed only sporadically and intermittently, mostly in the form of soft law instruments. These efforts were also relatively narrow in their scope, addressing the impacts of specific technologies on specific aspects of human rights, such as the right to privacy, data protection, freedom of speech and expression, and health amongst others.

This report, and the HRBA@Tech model more generally, pertains to all NETs, and therefore it is worth briefly considering the nature of various NETs, their human rights implications, and an overview of the existing responses to these new and emerging technological innovation thereto, which is the subject of Chapter 2.

Digital Technologies and Human Rights

The term "digital technologies" encompasses electronic equipment and applications that are used to find, analyze, store, create, communicate and disseminate information (i.e., data), and is often used interchangeably with the term "Information and Communication Technologies (ICT)." ICT is an umbrella term that refers to the infrastructure that facilitates computing and is generally accepted to mean all devices, applications, networking components and systems that enable communication and management of information as well as network hardware and software, and associated services. It includes antiquated technologies such as telephones, radio and televisions but also more recent technologies such as cellular/mobile phones, computers, and satellite systems, amongst others. "Digitalization" is the process of using such ICT technologies to convert physical information into digital formats or computer readable language.

Digital technologies today are pervasive and deeply integrated in almost every aspect of human life and across areas of social, economic and political activity. Business models for the supply of products and services increasingly rely on digital technologies and the creation of such digital economies has altered patterns of production and consumption, reshaped systems of education, healthcare and public infrastructure. These technologies are also increasingly influencing public services, as states are heavily investing in the use of digital tools across the full range of decision-making processes. The trend towards digitization accelerated rapidly during the COVID-19 outbreak, where digital technologies and tools became essential to the pandemic response. They ensured continuation of access to healthcare, education, economic activity and other necessary services at a time when most other aspects of social and public life had come to a standstill.

Business models for the supply of products and services increasingly rely on digital technologies and the creation of such digital economies has altered patterns of production and consumption, reshaped systems of education, healthcare, and public infrastructure.

Digital technologies are not new; however, they have become more sophisticated over the years. This is due to an exponential increase in computing power ushered by the development of new and cheaper microprocessor technology. This has made it possible to execute complex tasks at higher speeds and scales than ever before. This greater capacity spurred the development of new devices and technologies offering enhanced capabilities and services, facilitated by the availability of vast amounts of data and the growth of the Internet and broadband networks.

These trends have variably been referred to as the third and fourth industrial revolutions. The "third revolution" typically refers to the explosion of electronic and ICT products and the advent of the internet, and the "fourth revolution" to the explosion of digital products and services building upon existing ICT products and taking the innovations of the "third revolution" to new levels in ways that have a transformative impact across sectors and areas. The fourth industrial revolution is characterized not just by the increasing sophistication of individual digital technologies, but also their cumulative societal effect.

New and Emerging Digital Technologies: Work of the HRC and the 2021 Advisory Committee Report

Recent years have seen many efforts within the international human rights system and the UN system in general to address the human rights implications of the innovations associated with the third and fourth industrial revolutions.

Most recently, this includes the work of the Human Rights Council (HRC) on "new and emerging digital technologies." In 2019, the HRC adopted resolution 41/11, 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 noted that the full extent of possible human rights impacts is still poorly understood, and requested the 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.

After undertaking a detailed study of the issue, surveying existing initiatives, and considering the inputs from a range of stakeholders, the Advisory Committee presented its landmark report to the HRC in June 2021. The report defines "new technologies" as the technological innovations that transform the boundaries between the virtual, physical and biological spaces, and also includes in that definition new technologies and techniques for the datafication (the process of transforming subjects, objects and practices into digital data), data distribution and automated decision-making. Examples of such novel techniques include Artificial Intelligence (AI), the Internet of Things, blockchain technology and cloud computing, amongst others. These processes are characterized by the synchronization of online and offline spaces in

36. Commission on Human Rights, Resolution 108(XXXIII), 11 March 1977

37. Commission on Human Rights, Resolution, 1983/41

38. Commission on Human Rights, Resolution, 1984/27

39. Vienna Declaration and Programme of Action, World Conference on Human Rights in Vienna (25 June 1993), accessed at: https://www.ohchr.org/en/instruments-mechanisms/instruments/vienna-declaration-and-programme-action

40. UNGA Res 55/2, United Nations Millennium Declaration (8 September 2000); accessed at: https://www.ohchr.org/en/instruments-mechanisms/instruments/united-nations-millennium-declaration

a process described by the Advisory Committee[41] as the physical-digital-physical loop or the datafication cycle which is spread across a series of three stages: (1) a datafication stage where physical events are transformed into data and stored online; (2) a distribution stage where that data is shared and distributed into larger datasets; and (3) a decision-making stage where those digital datasets are used to design policy or make decisions that have an impact on people in the real world through algorithmic or automated systems, with or without human oversight.

The Advisory Committee report highlights the technology paradox, and notes also the important role played by private actors, such as tech companies, in the development and deployment of new technologies. The report highlights various conceptual and operational gaps within the existing human rights framework, which effectively act as barriers to addressing human rights issues arising out of new and emerging digital technologies and prevent the development of a unified approach. These conceptual gaps include the challenges of adapting existing international human rights standards to the current realities of the digital age, a lack of cooperation and coordination between the human rights and the tech communities, and the selective emphasis on some technologies and some human rights harms as opposed to a more balanced approach that embraces also the potential 'upsides' of technological innovations in the achievement of human rights. The report also highlighted several operational gaps, including practical challenges due to the nature of innovation, which usually outpaces any credible attempts to regulate new technologies, and the fragmentation of regulatory initiatives leading to governance gaps. Moreover, the report notes a lack of resources for human rights mechanisms. The report concludes by calling for a "human rights-based approach" to NETs based on three pillars: (1) a holistic understanding of technology; (2) a holistic approach to human rights; and (3) holistic governance and regulatory efforts.

Previous efforts at the international level to deal with human rights and ICT technologies often proceeded in disciplinary silos. The Advisory Committee report, in contrast, focuses on overarching and holistic solutions. The Advisory Committee Report also explicitly rejects the VNT (the assertion that technology is neutral), which had been central to previous discussions of the human rights implications of new technologies at the UN level. The Advisory Committee report dismissed the VNT as an oversimplification and observed that technologies can and do embody the values and biases of the people or entities that make them. The Advisory Committee noted that such biases (whether intentional or unintentional) can result in discriminatory outcomes, especially in cases of AI-based decision- making. Accordingly, the Advisory Committee recommended strategies to be elaborated that would function to seek out and counter bias not just from individual user(s) of technology, but also the technology itself.

Second, the Advisory Committee Report recognized the futility of focusing on the human rights impacts of any one technology taken in isolation, and how a singular technology cannot be distilled from the cumulative impacts of that technology as part of an interconnected ecosystem of technologies. This again represents a paradigmatic shift from previous approaches, many of which focused on one technology in isolation. According to the Advisory Committee's approach, technology (or rather technologies) are woven together into evolving technological ecosystems that collectively impact human interactions, and can therefore only be addressed by means of an integrated and holistic approach.

After the Advisory Committee issued its report, the HRC issued Resolution 47/23, in which it reaffirmed the need for a human rights- based approach to new and emerging digital technologies and the need for a holistic, comprehensive and inclusive approach with an emphasis on multi-stakeholder cooperation.

Bringing the Private Sector into the Equation

In parallel to these developments, a separate yet closely related strand of normative evolution was emerging in the form of the business and human rights movement. This normative strand of thought is relevant to new and emerging digital technologies, since many of these technologies are developed and deployed by private (corporate) actors.

The classical human rights framework was traditionally designed with the State at the center of obligations towards individual rights holders. Over the years, however, much work has been done to articulate precisely how this framework applies also to private actors, in recognition of the seminal role that businesses play in the realization (or non-realization) of human rights. Although States remain the primary duty bearers and thus the primary actors responsible for enforcing human rights norms within their respective jurisdictions, businesses are increasingly also understood to be key stakeholders in the promotion and protection of human rights.

In the year 2000, at the initiative of the UN Secretary General, the voluntary Global Compact (which brings together private sector actors to promote sustainable development and act in the service of broader UN initiatives) promoted a soft law approach to regulate the activities and operations of transnational corporations. It set out ten non-binding principles that companies could mainstream across their business operations, with the key focal points being human rights, labor protections, environmental sustainability, and anti-corruption. The Global Compact lacks a robust enforcement mechanism or recourse in case of non-adherence to these principles, but nonetheless served an important role in the articulation of globally-applicable norms. With time, these norms gradually found their ways into corporate boardrooms and ethics policies around the world.

In 2011, the UN Special Rep. of the Secretary General on Human Rights and Transnational Corporations and other Business Enterprises (SRSG on Business) presented the UN Guiding Principles on Business and Human Rights (UNGPs)[42] to the Human Rights Council. The UNGPs represented a landmark moment in the ongoing efforts to promote a sense of corporate social responsibility and respect for human rights, and were unanimously endorsed by the UN HRC.[43] The UNGPs provided a three-pillar framework:

1. **The State's duty to protect human rights.**
   States, as the primary duty bearers, retain the obligation to protect human rights. This includes the obligation to enforce laws requiring businesses operating within their jurisdiction to respect human rights. States must ensure that their domestic laws and policies, including corporate law, do not constrain but rather enable business to respect human rights. States are also obligated to provide effective guidance to businesses on how best to respect human rights throughout their business operations, and encourage (and where appropriate) require businesses to communicate those efforts publicly.

2. **A corporate responsibility to respect human rights.**
   This corporate responsibility to respect human rights entails a conscientious effort by businesses to not directly or indirectly cause or contribute to adverse human rights impacts through their activities and to address, mitigate and remedy such impacts if they nonetheless do occur. Accordingly, businesses are required to articulate their corporate strategy or commitments to respect human rights. In addition, they are required to develop human rights due diligence processes designed to identify, prevent, mitigate and account for any human rights impacts of their business operations. Finally, they are obligated to develop legitimate grievance processes that enable the remediation of any adverse human rights impacts they cause or to which they may have contributed.

3. **Victims access to an effective remedy.**
   Lastly, victims of business-related human rights abuses, as individual rights holders, must have access to effective remediation mechanisms. This includes state and privately administered grievance processes, all of which should embody a common set of minimal standards to be considered legitimate.

Though the UNGPs constitute "soft law" and are not legally binding, they have nonetheless become authoritative standards for corporate responsibility to respect human rights, and have already been adopted into various binding legal and policy frameworks in national jurisdictions. They continue to guide further work in the area of business and human rights, and are becoming increasingly accepted standards even in jurisdictions where they do not currently constitute "hard law." The UNGPs speak to the responsibilities of all private corporations, including today's technology companies, whether they be start-ups or multinational corporations. In many ways the UNGPs are ideally suited for technology corporations, since they often wield immense political and financial power and are sometimes able to defy jurisdictional control while also exerting significant influence on society and policy makers. Some of these corporations are even performing core governance functions that previously might have been handled exclusively by sovereign States, such as providing public services, facilitating the adjudication of disputes, and even holding human rights violators to account for their

41. Human Rights Council Advisory Committee (2021) "Possible impacts, opportunities and challenges of new and emerging digital technologies with regard to the promotion and protection of human rights", Art. II (7)a-c, UN Doc. (May 19, 2021) A/HRC/47/52, New York, NY: United Nations

42. John Ruggie (Special Representative of the Secretary-General), Rep. on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises: Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, U.N. Doc A/HRC/17/31 (Mar. 21, 2011)

43. Endorsed in HRC Resolution 17/4, 2011

misdeeds in "courts of public opinion" (as exemplified by the #MeToo or the #BlackLivesMatter movements).

Following its adoption of the UNGPs, the Human Rights Council replaced the Special Rapporteurship with a more robust and multidisciplinary Working Group on Business and Human Rights, comprised of five members. Together, these experts remain responsible for the promotion of the UNGPs and to make recommendations for their further implementation. The Working Group has also examined the impact of technological developments on the realization of the UNGPs, discussing, for example, the human rights challenges posed by novel and disruptive technologies such as AI and blockchain technology, while also pointing out their promise in realizing some of the goals of the UNGPs—for example, by enabling the more efficient tracking and assessment of complex supply chains,[44] or in helping to reduce barriers for vulnerable social groups such as women and minorities[45] to enter into the labor force.

Building upon the UNGPs, several human rights mechanisms have examined the role of business in the context of specific rights. The Committee on Economic, Social and Cultural Rights, for example, issued General Comment No. 24 in 2017[46] wherein it noted the key role of businesses—including tech companies—in the realization of economic, social and cultural rights such as the rights to health,[47] housing,[48] food,[49] water,[50] social security,[51] and right to work and just and favor favorable conditions of work,[52] amongst others. The ICESCR has also noted the challenges of holding businesses accountable for human rights violations when they operate extraterritorially, and in the context of NETs has specifically observed that many inequalities are strongly linked to the capacity of businesses to access, store and exploit massive amounts of data.[53] The Committee on the Rights of the Child issued General Comment No. 16 in 2013 on the impact of businesses on the rights of children,[54] noting the growing impact of the business sector on child rights owing to globalization, privatization of State functions, and technological advancement. The Committee on the Rights of the Child also highlighted the role of technology companies with regard to online crimes against children, online sexual abuse and exploitation, and exposure to harmful content, emphasizing the need for cooperation between States and the ICT industry and also the need for those corporations to conduct child rights impact assessments and due diligence.[55] The Committee on the Rights of the Child subsequently focused specifically on the rights of children in the digital environment in its 2021 General Comment No. 25.[56]

Other Developments on New and Emerging Technologies

In parallel to the work going on under the auspices of the Human Rights Council, the UN Secretary General (UNSG) has also been actively promoting efforts to address the impacts of NETs. In 2018, the UNSG released its Strategy on New Technologies, an internal, overarching guide for the UN system to define how it will use new technologies to accelerate its efforts in the achievement of its mandate. This strategy draws on the 2030 Agenda (the Sustainable Development Goals) that advances five principles: the protection and promotion of global values; the fostering of inclusion and transparency; the ideal of working together in partnership with other stakeholders; building on existing capabilities; and adopting a learning mindset. In the same year, the UNSG also established the High-Level Panel on Digital Cooperation to strengthen international and multi-stakeholder digital cooperation and provide recommendations for the international community to optimize the benefits of digital technologies while mitigating risks. In 2019 the Panel published its report titled "The Age of Digital Interdependence", which included a series of recommendations on digital cooperation, including the building of an inclusive digital economy and society; the development of human and institutional capacities; the protection of human rights and human agencies; the promotion of digital trust, security and stability; and efforts to foster global digital cooperation. On the basis of this report, the UNSG in 2020 launched "A Call to Action for Human Rights" and a "Roadmap for Digital Cooperation", setting the agenda to reflect upon the actual and potential implications of digital technologies on human rights and recognizing them as important instrumentalities of a fair, safe, and dignified future for humanity and the eventual achievement of Agenda 2030. This was also followed by the establishment of the Office of the Envoy of the Secretary-General on Technology. Meanwhile, efforts to negotiate a Global Digital Compact to ensure an open, free and secure digital future for all remain ongoing.

These initiatives by the UNSG are complemented by parallel efforts undertaken by the Office of the High Commissioner on Human Rights (OHCHR) on digital technologies and human rights. OHCHR's "B-Tech Project" seeks to provide an authoritative roadmap for applying the UNGPs to the development and use of digital technologies. It promotes an inclusive and participatory consultation and research process involving key stakeholders in order to better understand the cross-cutting impacts of NETs on the enjoyment of human rights, and a search for practical solutions that build on existing initiatives, good practice and expertise. Its work comprises practical guidelines and public policy recommendations for the realization of a human rights- based approach to the development, application and governance of digital technologies.

The Human Rights Council, in its 2021 Resolution 47/23, requested the OHCHR to prepare a report on the practical application of the UNGPs to tech companies. The OHCHR published its report in 2022, wherein it noted that despite a wealth of initiatives and efforts within the international human rights system to deal with the duality of human rights implications of new and emerging digital technologies, existing regulatory frameworks remain fragmented and unclear. The OHCHR's conclusions echoed those reached by the HRC's Advisory Committee, arguing that the UNGPs provide the most compelling starting point for tech companies and States to mitigate risks associated with digital technologies while also fostering innovation and creating a fair, level playing field for all. The OHCHR issued a series of detailed recommendations for applying the UNGPs to the activities of tech companies.

In recent years, the UN has also embraced digital technologies as an integral part of its operational strategy. Acknowledging its role as a global platform for discourse on issues related to NETs, the UN is taking steps to modernize its own approach to technology. Its various organs and specialized agencies are increasingly utilizing digital technologies and innovations to improve their operations. The UNSG, for example, has established an innovation lab to promote technological innovation, share best practices, and promote innovative solutions to accelerate implementation of the SDGs. It has also led to the establishment of the Global Pulse Platform to leverage AI and big data in efforts to promote peace and development.

This review of the UN's existing efforts to develop standards related to human rights and NETs barely scratches the surface of the extensive work that characterizes this field. It does not cover, for example, the voluminous work taking place at national policy-making levels, in regional and other international organizations, within the corporate sector, under the auspices of multi-stakeholder industry associations, in civil society, and in academic circles. A full discussion of those efforts goes beyond the scope of this paper.

Certainly, this paper, and its proposed HRBA@Tech Model, is neither the first such proposal, nor will it likely be the last. We identified more than 200 existing proposals for standards on how to manage the human rights impacts of technology. The authors of this paper seek merely to bring together as many strands of thought on this issue as possible. We propose a common way forward, built on a recognition of the fundamental duality of technology and oriented towards the development of a concrete, solution-driven model that can be embraced by human rights actors and technologists alike as they jointly work to 'nudge' technologies in the direction of human rights."

44. A/73/163 (2018); A/HRC/41/49 (2019); A/75/212 (2020)
45. A/HRC/41/43 (2019)
46. CESCR, General Comment No. 24, 2017
47. CESCR, General Comment No. 14, 2000, paras. 26 and 35
48. CESCR, General Comment No. 4, 1991, para. 14
49. CESCR, General Comment No. 12, 1999, paras. 19 and 20
50. CESCR, General Comment No. 15, 2002, para. 49
51. CESCR, General Comment No. 19, 2007, paras. 45, 46 and 71
52. CESCR, General Comment No. 18, 2005, para. 52. See also CESCR General Comment No. 23, 2016, paras. 74.
53. CESCR, General Comment No. 25-Article 15(1)(b), (2), (3) and (4): Science and Economic, Social and Cultural Rights (2020)
54. CRC, General Comment No. 16 on State obligations regarding the impact of the business sector on children's rights (2 March 2021), CRC/C/GC/25
55. CRC, General Comment No. 17 on the right of the child to rest, leisure, play, recreational activities, cultural life and the arts (art. 31) CRC/C/GC/17 (2013); CRC, General Comment No. 20 on the implementation of the rights of the child during adolescence CRC/C/GC/20 (2016)
56. CRC, General Comment No. 25-Children's Rights in Relation to the Digital Environment (2 March 2021), CRC/C/GC/25

CHAPTER 2: THE HUMAN RIGHTS AND NEW AND EMERGING TECHNOLOGIES PARADOX

Chapter Summary:

This Chapter focuses on the paradox of new technologies, namely that the same technologies can often have both positive and negative impacts on human rights. NETs can enhance our collective enjoyment of human rights (e.g., access to information, the right to quality health care and education, or efforts to assist persons living with disability to participate equally in civic life); improve public health and welfare; improve inclusive education and youth welfare activities; promote efforts to monitoring human rights situations (e.g., by facilitating secure communication among human rights activists, remote sensing, satellite imagery, data forensics, protecting human rights defenders, etc.). At the same time, NETs can also cause potential and actual human rights harms (e.g., by facilitating discrimination based on race, gender, or other protected characteristics, by enabling discriminatory surveillance, through biased algorithms, by spreading hate speech and by allowing online sexual harassment and other crimes to proliferate in difficult-to-regulate forums, etc.). These two elements–the positive and the negative; the human rights promoting as well as the human rights threatening elements of NETs– would be addressed together in order to move away from a polarising dichotomy in which this issue is often framed, and to move instead towards a more nuanced, holistic and comprehensive approach to the issue. The Chapter highlights a diverse set of four NETs and considers the particular implications of these technologies for the enjoyment of human rights. It briefly addresses some of the past attempts to address these technologies from a human rights perspective, with a view to setting the stage for the presentation of the HRBA@Tech model in subsequent Chapters.

The analysis in Chapter 1 revealed two insights. First, it has shown that the paradox of NETs is no novel discovery. The duality of NETs has likely been with us since the origins of technological innovation itself. Thus, our challenge is to find a middle path between those two extremes. The second insight is that while what is considered a "new and emerging" technology may change from one day to another, and yet the paradox remains a constant.

To illustrate this, we initially selected eight NETs – or at least technologies that were described as "new or emerging" in 2022 – as the fuel that would drive the elaboration of the HRBA@Tech model (described in Chapter 3). We were keenly aware that each NET would likely pose new and unique particularities, and hoped that this strategy might help us reality-test our emerging ideas about what a universal HRBA@Tech model would look like not just in light of one particular technology, but in light of all of them in a general and overarching way.

For each such technology we first conducted a rudimentary analysis of that technology's "promise" from the perspective of protecting and promoting human rights. Second, we conducted a similar analysis of the potential harms or "risks" that might flow from that technology, again from the human rights perspective. Finally, we discussed efforts underway to 'nudge' each of those technologies into the direction of human rights. We chose four of the above-mentioned technologies to highlight in this chapter, based on their diversity and also the degree to which they are currently considered "hot topics" at the intersection of human rights and technology.

Each of those sub-Chapters illustrates the paradox of technology, and highlights the kaleidoscope of different stakeholders coalescing at different points of a technology's life cycle, in common pursuit of different principles.

HR & NETs	Brief Description
Genetic Engineering	Genetic engineering is a method of artificially changing the human genome by means of a new technology that allows scientists to adapt, replicate, change, or block certain parts of the human DNA genome. This technology promises to allow scientists to edit human genomes, potentially curing previously incurable diseases. The technology also opens the possibility, however, for scientists to make changes to the human genome that can be passed from one generation to the next (infinitely) thus potentially indefinitely altering the human genome. It also opens up at least the theoretical possibility of scientists creating "designer babies" with certain non-medically necessitated alterations (e.g., certain hair or eye-color, gender, or other physical characteristics that might be considered "desirable" by the parents but that might also perpetuate harmful social stereotypes).
Internet- Based Technology including Social Media	The internet and social media are not new technologies, and yet they continue to be active spaces for innovation. This technology has connected billions of people on single platforms, opening up hitherto unimaginable opportunities for communication and the exchange of ideas. This same technology has also opened the door for predatory behavior to flourish, for example criminal efforts to harass, exploit, or humiliate individuals via online channels. New developments in this field are exploring a so-called "Metaverse" in which technologists hope we will spend even more of our time immersed in virtual realities, both professionally as well as socially.
Geo- Engineering	Geo-engineering is a controversial technology, often criticized by environmentalists as well as technologists as untested, unethical, and as a diversion from more pressing discussions on tackling the climate crisis. However, with increasing and accumulating evidence that global efforts to 'mitigate' and/ or 'adapt' to climate change will likely not succeed in keeping global temperatures within the target of 1.5°C of pre-industrial averages, scientists are again exploring whether it might be possible to "engineer" the climate back within safe limits, either by removing carbon dioxide from the atmosphere or by deflecting sunlight away from the earth's surface. This technology is unproven, but to prove it one risks causing irreversible harms that cannot – by definition – be contained in a laboratory. The impacts of these harms also risk being unevenly distributed globally, raising concerns that single countries might initiate such a scheme without regard for the harms it might be causing in other communities.

HR & NETs	Brief Description
Artificial Intelligence	Artificial Intelligence has already revolutionized how we work with data and make decisions. From health care to national security, artificial intelligence is increasingly changing how we perceive, reason and engage with our world. A new generation of machine learning is making these decisions autonomously, often without meaningful human control or oversight. AI is an enabling technology – speeding up and rendering infinitely more powerful various decision making processes. The technology promises to revolutionize virtually any process-those designed to maximize profit but also those designed to promote human rights. The flipside is also true, however, in that AI can also undermine human rights, including exacerbating discrimination, threatening privacy and stifling free speech amongst others. Far from extreme scenarios of killer robots and machines that can "feel" human emotions, the much more present-day threat of AI is its ability to subtly replicate social biases and 'sterilize' them under the guise of quantified objectivity.
3D Printing	3D printing is rapidly evolving to allow regular individuals to print three-dimensional objects at home, using printing instructions that can be freely downloaded online. Simple 3D printers allow users to print using one material, but more sophisticated printers can conceivably print in a variety of materials, including even human tissues. This technology can enable tinkerers in literally any corner of the world to create dynamic physical objects, and can potentially be used one day to "print" replacement organs for human patients. At the same time, the technology risks undermining intellectual property regimes, and could even threaten sanctions regimes – a tool often used by the human rights community to encourage human rights compliance. Moreover, 3D printing risks rendering irrelevant entire sectors of craftsperson, small manufacturing operations, and cottage industries.

EXAMPLE 1:
Human Rights & Genetic Engineering

Genetic engineering, in some form, has always been a part of human history. Native Americans living in present-day Mexico, between 8000 and 4000 BCE used selective breeding methods to cultivate modern-day corn (maize) from teosinte, a type of wild grass native to the area.[57] Similarly, horse breeders in the latter half of the 18th century aggressively bred horses for agility and power, thus dramatically reducing the genetic diversity that had existed prior to that point among the worldwide horse population.[58] With the discovery of the 3-dimensional structure of deoxyribonucleic acid (DNA) and the realization that embedded within these DNA molecules were the "blueprints" for biological life, however, the field of genetics transformed from a theory about selective breeding into the study of a concrete DNA molecule. By 2003 the human genome had been mapped in its entirety[59] and by 2014, scientists at the Massachusetts Institute of Technology had developed a "genome editing" technology known as CRISPR (named after the "Clustered Regularly Interspaced Short Palindromic "Repeats" that form part of the body's natural defenses against bacteria), that could be used to correct, deactivate, or replace targeted parts of the DNA molecule with alternative sequences.

Promises

CRISPR technology has made it relatively easy for scientists to experiment with genetic engineering. Proponents of this kind of research anticipate finding new treatments for a number of diseases such as cancer, various types of ocular, hematological, immunological, cardiovascular, and neurodegenerative diseases, among others.[60] Some diseases, for example Cystic fibrosis, Huntington's chorea, Duchenne muscular dystrophy, and sickle cell anemia result from mutations in only one gene in the human genome.[61] Patients suffering from such diseases might be cured of their diseases by means of a CRISPR-based cure, thus restoring their regular biological functions. Such treatments are referred to as "somatic cell modification" therapies "reflecting the traditional approach to disease mitigation" since the impact of the intervention is limited to the patient being treated only.[62] Genetic engineering can also alter the cells used to reproduce, however, in which case the altered genetic code would be passed from one generation to the next, in perpetuity. This type of genetic engineering, which is already common in agriculture and animal experiments, is called "germline genome editing." It holds the radical promise of eliminating certain diseases permanently from the human population.

Risks

In 2018, a Chinese scientist announced that he had used germline genomic engineering to produce two viable human embryos (twins): girls named Lulu and Nana. This caused an immediate condemnation of the scientist's actions from around the world, including strong legal and regulatory disciplinary action in China itself. Arguments against germline genomic engineering are many, and have mostly to do with still unanswered questions about the relationship between the human genome and the physical traits that manifest based on that DNA. As a result, scientists can often only guess at whether a particular change in the human genome will have the desired therapeutic effect. Moreover, scientists may also not be able to predict unintended side- effects that might result from a particular genetic alteration. If such an unintended negative side-effect happens because of somatic cell modification therapy, the impacts will be limited to that one (presumably consenting) adult patient. If the unintended negative impacts happened in cells used for reproduction, on the other hand (germline genetic editing) the impacts could be profound, altering humanity in perpetuity, and obviously without the consent of those future generations who would be accordingly altered.

Risks of unintended consequences aside, some have also argued that genetic engineering risks permanently altering what it means to be human itself, and thus constitutes a violation of basic human dignity, especially for future generations who will have been robbed of what today we might describe as our full human experience.[63] Still others worry that by offering the possibility of "correcting" for traits that today may be associated with common disabilities, genetic engineering risks re-stigmatizing persons who already live with such disabilities, undoing years of hard-fought progress in combatting stigma. Analysts have also warned about the potential for "designer babies" to create a new class inequality, reinforced not just socio-economically but now also biologically. These critiques, while powerful, are not universally shared, and many scientists feel that the risks of genetic engineering are more than outweighed by their potential therapeutic value, and moreover that they can be effectively managed.

Proposed Solutions

Numerous attempts have been made to regulate genetic engineering. Initially, many of those efforts have proposed blanket bans on genetic engineering, especially those interventions focused on germline editing. With time, however, such stances have softened, driven (perhaps) by a re-examination of the comparative risks and benefits associated with this technology, and no-doubt informed by greater scientific familiarity with the underlying risks involved.

In addition, industry-efforts are also underway to limit the degree to which genetic engineering will be prone to exploitation by "rogue scientists."[64]

At the international level, discussion began to focus on genetic engineering around the turn of the millennium. In 1997, the Council of Europe issued the Convention for the Protection of Human Rights and Dignity of the Human Being with Regard to the Application of Biology and Medicine (commonly

57. National Science Foundation (2005) "Scientists Trace Corn Ancestry from Ancient Grass to Modern Crop" News Release 05-088 (May 27), https://www.nsf.gov/news/ (accessed Nov. 16, 2022)

58. Katherine Wu (2019) "Domestic horses have existed for millennia. Modern breeding reduced their genetic diversity in just a few hundred years," NOVA (May 3), https://www.pbs.org/wgbh/nova/article/horses-genetic-diversity/ (accessed Nov. 16, 2022).

59. National Human Genome Research Institute at the National Institutes of Health (2022) "the Human Genome Project" (last updated September 2, 2022), https://www.genome.gov/human-genome-project (accessed Nov. 16, 2022).

60. Ning Guo, Ji-Bin Liu, Wen Li, Yu-Shui Ma, and Da Fu (2022), "The Power and the Promise of CRISPR/Cas9 Genome Editing for Clinical Application with Gene Therapy," 40 Journal of Advanced Research 135-152

61. Jennifer Doudna (2020) "The Promise and Challenge of Therapeutic Genome Editing," Nature 578, 229–236. https://doi.org/10.1038/ s41586-020-1978-5

62. Ibid., at 233

63. Ibid., at 345-49 (discussion, including also the authors retort to such claims).

64. R. Alta Charo (2019), "Rogues and Regulation of Germline Editing," 380(10) New England Journal of Medicine 976-980, 977

known as the Oviedo Convention).[65] This convention stated that the "interests and welfare of the human being shall prevail over the sole interest of society or science," and explicitly distinguished three types of genetic engineering:

1. predictive genetic tests, which are permissible as long as they are motivated by "health purposes," (Article 12);
2. genetic engineering, which is permissible for "preventive, diagnostic or therapeutic purposes only" and only for somatic- cell modifications, not germline interventions (Article 13); and
3. a complete prohibition on sex-selective genetic interventions, except when motivated by a desire to avoid a serious inheritable disease (Article 14).

This early convention language, which was drafted almost two decades prior to the invention of CRISPR technology, established a strong normative predisposition against germline genetic engineering in all cases, regardless of its potential therapeutic value. The convention grounded this blanket prohibition in this understanding of the human genome as a constituent part of human dignity and human identity (Article 1).

Later that same year, UNESCO hosted a process culminating in the Universal Declaration on the Human Genome and Human Rights.[66] This declaration again highlighted the human genome as the "heritage of humanity." It highlighted fundamental issues related to the dangers of unethical scientific and technological research in the fields of biology and genetics, and proposed a three-pronged approach to protect against those dangers:

1. "States and competent international organizations [should] identify [] such practices and tak[e], at [the] national or international level, the measures necessary to ensure that the [practices which are contrary to human dignity, such as reproductive cloning of human beings, shall not be permitted]" (Article 11);
2. The benefits of scientific research should "be made available to all" (Article 12.a); and
3. A reaffirmation that freedom of research derives from the fundamental human right to freedom of thought, but also that "[t]he applications of research, including applications in biology, genetics and medicine, concerning the human genome, shall seek to offer relief from suffering and improve the health of individuals and humankind as a whole." (Article 12.b)

As is true also in the HRBA@Tech model (described in Chapter 3 below) in this paper, the declaration goes beyond a "do-no-harm" approach to human rights in that it places a proactive normative obligation on scientists to advance human well-being by means of their research.

Eight years later, in 2005, UNESCO oversaw the drafting of the Universal Declaration on Bioethics and Human Rights,[67] which set forth universally recognized principles in the field of bioethics, anchored in human rights and the need to safeguard human dignity, respect for privacy, autonomy, confidentiality, non-discrimination, informed consent, the need to maximize the benefits while also minimizing the harm of advancements in scientific knowledge, respect for human vulnerability and personal integrity, the desire to share the benefits from scientific research and its applications, and the protection of future generations – all issues that remain central to the debate on human rights and technology more broadly even today.

In parallel to these efforts at the international level, many countries also implemented national legislation to regulate genetic engineering. A discussion of those various legislations goes beyond the scope of this paper.

Other stakeholders have also been busy supplementing these regulatory and legislative efforts. This so-called "'ecosystem approach' will arguably do more to control and guide [] technology than a moratorium or formal ban."[68] This ecosystem of actors includes institutions at the international level (for example the World Health Organization, which can create specialized committees to promulgate advice and guidance), the national level (for example national licensing authorities responsible for approving new therapeutic treatments) as well as at various private and academic institutions (for example insurance providers, research oversight boards, funders, publishers of respected academic journals, and even professional medical licensing bodies).[69] Each of these stakeholders has an important role to play. Only when actors at all these levels are appropriately sensitized about the need to protect human rights (and the modalities of doing so) can this ecosystem function effectively.

EXAMPLE 2:
Human Rights & Internet based Technology (INCL. Social Media)

The internet first emerged in the 1960s as a physical connection between the computers of various research institutions in the United States. In 1983, a new and standardized communications protocol was unveiled that allowed computers to "speak" with one another. This is considered to be the birth of the modern internet as we know it today, where any computer can communicate with any other computer as long as they are both connected to the same communication network. The internet allows for the sharing of files, exchange of emails and electronic bulletin boards (blogs, distribution listservs, etc.), and also the creation of the countless websites on which individuals or corporations can post information for the rest of the world to access.

The way in which information is displayed, shared and communicated on the web is also evolving, largely enabled by changes in technology, communication trends, and hardware capabilities. In what has been called Web 1.0, websites were largely static information "broadcasting" tools. The owners of those websites would assemble the contents, whereupon users (usually using desktop or laptop computers) could consult that information. This earliest version of the internet fueled a hitherto non-existent ecosystem of corporations that have since become household names (not to mention multibillion-dollar enterprises), including companies that manufacture the hardware devices consumers use to access the internet (e.g., Samsung, Apple, Lenovo, Fujitsu), companies that specialize in selling products to customers online (e.g., Amazon, Coupang, Alibaba), and companies that help users orient themselves within the internet using simple and intuitive interfaces (e.g., Google, Naver, Baidu).

Beginning in the mid-2000s, a new generation of technology firms emerged, offering more interactive services over the internet. These services are now commonly referred to as "social media" platforms. These more interactive web applications are commonly referred to as "Web 2.0" websites (including YouTube, Facebook, Instagram, Naver, and countless others). Web 2.0 sites emphasize interactivity between the user and the owner of the website, or typically amongst the users themselves. Such engagement-oriented websites (often described collectively as "social media") have led to an explosion of various consumer and entertainment products and allowed the owners of these sites to gather increasing amounts of data about their users, and eventually market that data as a lucrative "product" to sell to those seeking more targeted advertising opportunities.

The growth of Web 2.0 applications, and especially their integration into the daily routines of so many individuals—increasingly not just in the Global North but also in the Global South—is the result of a symbiotic relationship between those web-based products and services and the explosion of the market for handheld computing devices (mobile phones) coupled with major technological advances in wireless telecommunications technology. The process of interacting with the web has undergone a dramatic transformation since the 2000s. What once required sitting at a desktop computer and navigating a slow internet connection via a web browser has now been streamlined to mere seconds using the mobile phones many of us carry constantly. Additionally, the volume of data we can process with these technologies has expanded from kilobytes to megabytes. The widespread use of these mobile devices, along with the interactive features of Web 2.0 applications, has led to the emergence of various web-facilitated services that have real-world applications. Examples include delivery services, ride-sharing platforms, and e-government services that streamline interactions between citizens and public institutions. This is the web that – by and large – has become a familiar feature of modern life at the time this report went to press in late 2022.

Technologists are currently speculating about a third generation of web applications, dubbed as Web 3.0 applications. What precisely distinguishes these applications from their 2.0 predecessors is still somewhat loosely defined. Some have mentioned that Web 3.0 applications will be decentralized, relying not on centralized servers storing mass volumes of data but rather decentralized blockchains storing data that can be accessed from anywhere. Others have commented on the increasing prevalence of AI enabled functions within Web 3.0 applications. AI-enabled online applica-

65. Council of Europe, Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine (Oviedo Convention), ETS No.164, April 4, 1997

66. General Conference of UNESCO, Universal Declaration on the Human Genome and Human Rights (11 November 1997); accessed at: https://www.ohchr.org/en/instruments-mechanisms/instruments/universal-declaration-human-genome-and-human-rights

67. Universal Declaration on Bioethics and Human Rights, SHS/EST/BIO/06/1, SHS.2006/WS/14, https://unesdoc.unesco.org/ark:/48223/pf0000146180

68. R. Alta Charo (2019), "Rogues and Regulation of Germline Editing," 380(10) New England Journal of Medicine 976-980, 976

69. Ibid.

tions and services allow users to interact with websites using natural (human) language as inputs, and also allow for the automated screening and analysis of large volumes of data in ways that are increasingly indistinguishable from how a human might manage the same process (except for its vastly improved speed). Other technologists are predicting that Web 3.0 applications will transform Web 2.0 social media experiences into truly immersive social spaces, often dubbed the "Metaverse", where – for example – colleagues from around the world can collaborate in virtual office spaces despite being physically located on different parts of the planet, perhaps even communicating across language barriers using AI-enabled translation apps, and collaborating in virtual spaces that appear as though they were the real thing by means of sensory-enhancing virtual reality technologies.

Finally, a new generation of interconnected machines promise to extend the internet beyond our web browsers, and beyond even our mobile phones, into various connected devices in our homes, workplaces, and lived environment. This so-called Internet of Things (IoT) might, for example, allow our refrigerators to "know" when we are running low on eggs, automatically add an item to a shopping

list that will be curated and delivered by a delivery service to our doorstep the following morning. Each step in this scenario could conceivably unfold without the homeowner even being consulted once to approve the various transactions required to make this ecosystem work. It might also allow our children's dolls to "talk" to our children using child-friendly language, enabled by AI language recognition programs. Similarly, our internet-connected watches might sense irregularities in our heartbeats and automatically call an ambulance, just as our internet enabled "smart speakers" may pick up auditory signs of domestic violence in a household and call the police to investigate.

Promises

From a human rights perspective, the internet has enabled an entire industry of "liberation technologies" built on expanding the fundamental right of free speech and expression. Human rights activists have used the internet to effectively spread information about rights abuse in ways that would have been difficult if not impossible in a pre-internet era. This has changed the way traditional human rights organizations do their work.[70] This potential has given birth to a new generation of human rights organizations dedicated to the use of technology as a tool to advance human rights.[71] Human rights activists have coordinated across borders and worked to effectively raise awareness about pressing human rights challenges. This has transformed community activism, human rights fact-finding, and strategies for human rights awareness-raising. It has enabled, for example, laborers and communities located along the upper reaches of a supply chain to engage directly with consumers of the products that rely on those raw inputs, transforming the potential for fundraising, community empowerment, and agency, and eliminating or lessening some of the distortions that plagued traditional advocacy efforts.

The internet has also dramatically expanded our collective access to information. Individuals at one moment can post information to the internet for the entire world to see, and a moment later they can access the thoughts, ideas, and reactions of virtually anyone else in the world. This has allowed interest-communities to emerge that might previously never have had the occasion or resources to interact with one another. This has been particularly important for marginalized, persecuted and vulnerable communities. Movements such as #MeToo, Black Lives Matter, #StopAsianHate, and various LGBTQI+ movements did not suddenly "discover" new human rights problems in societies, but they did successfully mobilize diverse coalitions of individuals with similar experiences and similar convictions in ways that would have been unimaginable without Web 2.0 platforms to facilitate such efforts.

The internet has also proven to be a tremendous enabler of social and economic rights, in ways too numerous to recount here. Rural communities in the Global North and the Global South have been able to dramatically improve the quality of their schools through the use of digital technology. During the COVID-19 pandemic, many of these technologies experienced accelerated and forced adoption, enabling children worldwide to continue their education despite mandatory school closures. This highlighted the critical role of digital platforms in maintaining essential services and social functions. While post- pandemic research has shown that online technologies still fall short of their offline alternatives in terms of effectiveness,[72] they undoubtedly still represent a vast improvement over no education at all.

Digital technologies also promise to revolutionize medical care. Internet-based technologies are enabling medical professionals to extend their reach far beyond their offices and even across national borders.[73] Such 'telemedicine' promises to extend high quality specialist healthcare deep into previously underserved regions, often also at significantly lower costs. Furthermore, a slew of wearable medical devices and AI-driven services promise to revolutionize early detection of potentially curable or preventable diseases.

> The Human Rights Council in the early 2010's embraced a theme of "Promotion, Protection and Enjoyment of human rights on the internet," which led to a number of corresponding resolutions in recent years.

Risks

Despite the technology sector's best efforts to reinforce the security of these various internet-based technologies, it is also well known that all systems remain prone to hacking. This poses profound national security threats, which go beyond the scope of this paper. It also raises numerous human rights issues. State and non-state actors, for example, can exploit technical or human weaknesses to compromise the data of individuals or organizations whom they consider to be adversaries. This information can be used to directly harm those adversaries, subvert their activities, or extort them to cease their activities. Numerous prominent human rights activists have found themselves on the receiving end of such attacks.

Moreover, state and private actors have also used internet-based technologies to effectively subvert regular democratic processes. The internet, and especially Web 2.0 type platforms, have proven to be fruitful playgrounds for small groups of activists to spread certain messages, perhaps with the intention of influencing public opinion in a certain way. Using the facilitated flow of communication that these platforms enable, state propagandists have found a low-cost, high- impact means to spread their messages, often in ways so subtle they can no longer be traced to their original authors.

In some ways, online spaces are merely continuation of offline realities.[74] In this sense, the internet can be said in some instances to merely reflect and perhaps amplify existing social, economic, or political inequalities and discrimination. This is particularly true for vulnerable groups such as racial, ethnic, and linguistic minorities, women, children, older individuals, and persons with disabilities, among others. In other instances, the "digital divide" (unequal access to the internet or digital illiteracy) creates a new vector for old inequalities and discriminatory social structures to manifest again in the form of unequally distributed resources and access. In yet other situations, hate speech and harassment, which obviously predated the advent of the internet, nonetheless proliferate and flourish in volumes and scale that would have been unimaginable prior to the digital era.

In addition, the architecture and logic of social media sites often does more than merely allow existing social biases to continue to proliferate. Social media is often criticized for exacerbating social and political polarization in societies around the world.[75] Algorithms designed to keep users engaged (and thus maximize advertising revenue for the technology platform) have been found to promote ever more salacious content, gradually pushing individuals into more and more self-rein-

70. See, for example, Amnesty Tech (https://www.amnesty.org/en/tech/), seeking to "Bolster social movements in an age of surveillance, [challenge] the systemic threat to our rights posed by the surveillance-based business model of the big tech companies; [ensure] accountability in the design and use of new and frontier technologies; [and encourage] innovative uses of technology to help support our fundamental rights." (accessed Nov. 17, 2022)

71. See, for example, the Sentinel Project (https://thesentinelproject.org/), Benetech (https://benetech.org/what-if/), and the Electronic Frontier Foundation (https://www.eff.org)

72. Meckler, Laura, "Scores fall coast to coast, especially in math, under pandemic's toll," (Oct. 24, 2022) Washington Post, https://www. washingtonpost.com/ (accessed Nov. 17, 2022)

73. Baek Byung-yeul, "KT targets Vietnam's digital healthcare market," (April 13, 2022) Korea Times https://www.koreatimes.co.kr/ (accessed Nov. 17, 2022)

74. Sida (2022), "HRBA and a Free, Open and Secure Internet", https://cdn.sida.se/app/uploads/2022/05/03093124/10205933_Sida_TN_HRBA_Secure_Internet_webb.pdf

75. Roose, Kevin, "Rabbit Hole" (podcast) New York Times, April 22, 2020, available at https://www.nytimes.com/column/rabbit-hole

forcing extremist political communities, rather than exposing them to a diversity of opinions and topics.[76] Moreover, the ability to post information to a diverse group of readers without having to internalize the sometimes significant impacts of those postings on the readership has emboldened many users of Web 2.0 technologies to post increasingly incendiary, racist and intentionally offensive messages. Social media "trolls" sometimes engage in causing offense as a form of sport, seeking the thrill of creating social turbulence as its own reward, often without facing any significant social repercussions. In response to such threats, most social media platforms have evolved their approach. Initially adopting a largely permissive stance, these platforms have shifted towards more proactive measures, focusing on purging their sites of terrorist or state-sponsored propaganda, hate speech, and blatant disinformation or misinformation.

Another problem, from a human rights perspective, is that the business models of many Web 2.0 platforms are specifically built on efforts to gather as much personal data about its users as possible. This data enables a host of ever-more-narrowly tailored advertising campaigns, which can be monetized by the technology platforms to generate vast profits. But these caches of data, held by private corporations, also represent a massive potential erosion of personal privacy, especially if those data start being used to discriminate against certain groups of individuals as a result of the datafication cycle (when accumulated datasets get used to inform decisions with real-world impacts).

The new Web 3.0 type IoT ecosystems, which often make direct use of such datasets, pose a whole new set of potential privacy issues. It is not hard to imagine how a malignant actor with sufficient technological hacking expertise could seek to gather information directly from the listening devices which – thanks to IoT technology – will be present in millions of households, phones, and wearable devices around the world—all conveniently connected directly to the web. Documented instances of privacy violations include hackers successfully compromising baby monitors, altering the functionality of cardiac devices, accessing the video feeds of security cameras, and even taking control of moving automobiles. These incidents underscore the pressing need for robust cybersecurity measures to protect against the vulnerabilities inherent in interconnected devices and systems.[77] Such hacks promise to become ever-more prominent as the ecosystem of different IoT devices begin to proliferate.

Proposed Solutions

For nearly half a century, the international community has been wrestling with the human rights implications of what we now refer to as ICT and digital technology. As early as 1977, the Commission on Human Rights tasked a newly- created Special Rapporteur to study the human rights implications of computerized personal files.[78] The Rapporteurship led the UNGA, 13 years later, to adopt the UN Guidelines for the Regulation of Computerized Personal Data Files.[79] This document set forth ten succinctly-stated principles ("orientations") that national authorities should use to guide their efforts to regulate the collection and retention of personal data files. In 1988, the Human Rights Committee issued a General Comment on the Right to Respect of Privacy, Family, Home and Correspondence that clarified key terms and specified the States' responsibility to ensure that such privacy rights are respected.[80]

As described above, since that time, ICT developments have transformed the ways in which humans communicate. This transformation has led to renewed efforts to bring ICT technologies in line with human rights priorities and principles.[81] Various international human rights bodies, including many Treaty Bodies and Special Procedures, have addressed both the positive as well as the negative role of the internet and ICT technologies in the realization of all human rights. Examples include the Committee on Economic, Social and Cultural Rights General Comment No. 21 on the right to take part in culture through the internet,[82] the Committee on the Elimination of Racial Discrimination General Comment No. 29 regarding online hate speech,[83] and various efforts to promote the rights of children online.[84]

A series of massive protests and uprisings across the Middle East and North Africa region, often known as the "Arab Spring," famously highlighted the tremendous potential for the internet and social media to play a role in the promotion of international human rights as a tool of mobilization and awareness raising, but also – in a subsequent wave of repression and reaction by many regimes in the region, as a tool of suppression, infiltration,[85] and a means to suppress independent speech and thought.

Partially in response, many international bodies emphasized the paradox of technology described above. Many noted, for example, how the internet and internet-based technology facilitates freedom of speech and expression[86] and freedom of assembly[87] The Committee on Economic, Social and Cultural Rights, for example, which is mandated to protect the right to enjoy the benefits of scientific progress and its applications,[88] has highlighted the role that technology, including ICT, plays in advancing the right to education[89] and health[90] among other economic and social rights.[91]

The Human Rights Council in the early 2010's embraced a theme of "Promotion, Protection and Enjoyment of human rights on the internet," which led to a number of corresponding resolutions. In 2012,[92] the HRC issued a resolution under that heading, which acknowledged that the rapid pace of technological development had enabled individuals all over the world to use ICT and recognized the internet as an enabler of human rights. The resolution underscored the internet's global and open nature, highlighting its role as a catalyst for progress worldwide. It emphasized the importance of ensuring that individuals enjoy the same human rights online as they do offline.

A number of other resolutions focused also on the ways in which our increasing reliance on ICT-enabled communication made us more vulnerable to human rights abuse. In light of striking revelations about the extent of global mass surveillance in the United States and elsewhere, a number of UN resolutions were drafted to protect the right to privacy in the digital space. The UNGA, for example, issued five separate resolutions on the Right to Privacy in the Digital Age,[93] supplemented by four more issued by the Human Rights Council (HRC).[94]

76. Paul M. Barrett, Justin Hendrix, J. Grant Sims (2021), Fueling the Fire: How Social Media Intensifies U.S. Political Polarization — And What Can Be Done About It" NYU Stern Center for Business and Human Rights, https://bhr.stern.nyu.edu/polarization-report-page (accessed Nov. 17, 2022)

77. Kilpatrick, Harold (2018) 5 Infamous IOT Hacks and Vulnerabilities, IOT Solutions World Congress, https://www.iotsworldcongress.com/5-infamous-iot-hacks-and-vulnerabilities/ (accessed Nov. 17, 2022)

78. Special Rapporteur on the Study of the Relevant Guidelines in the Field of Computerized Personal Files (1980), "Guidelines for the Regulation of Computerized Personal Data Files", https://digitallibrary.un.org/record/43365?ln=en

79. UNGA Res 45/95, Guidelines for the Regulation of Computerized Personal Data Files (14 December 1990).

80. Human Rights Committee, General Comment No. 16-Article 17: The right to respect of privacy, family, home and correspondence, and protection of honour and reputation (1988).

81. Rifkin, Jeremy (2011), The Third Industrial Revolution: How Lateral Power is Transforming Energy, the Economy, and the World, New York, NY: Palgrave MacMillan

82. Committee on Economic, Social and Cultural Rights, General Comment No. 21: Right of everyone to take part in cultural life (2009), E/C.12/GC/21

83. Committee on the Elimination of Racial Discrimination, General Comments No. 29 (2002)

84. Committee on the Rights of the Child, General Comment No. 7 (2006), General Comment No. 9 (2007); Special Rapporteur on the sale of children, child prostitution and child pornography, E/CN.4/2005/78 (2005)

85. See contrast between Blake Hounshell, "The Revolution Will Be Tweeted: Life in the vanguard of the new Twitter proletariat." (June 20, 2011) Foreign Policy, https://foreignpolicy.com/2011/06/20/the-revolution-will-be-tweeted/, and Malcom Gladwell, "Small Change: Why the revolution will not be tweeted." (Sept. 27, 2010) the New Yorker, https://www.newyorker.com/magazine/2010/10/04/small-change-malcolm-gladwell).

86. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression. General Assembly, 66th session, 10 August 2011, http://daccess-dds-ny.un.org/doc/UNDOC/GEN/N11/449/78/PDF/N1144978.pdf

87. Report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association: The rights to freedom of peaceful assembly and of association in the digital age, 17 May 2019, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/141/02/PDF/ G1914102.pdf?OpenElement

88. ICESCR, art. 15(1)(b)

89. Committee on Economic, Social and Cultural Rights, General Comment No. 13-Article 13: The Right to Highest Attainable Standard of Health (2000); Committee on Economic, Social and Cultural Rights, General Comment No. 22-Article 12: The right to sexual and reproductive health (2016)

90. Committee on Economic, Social and Cultural Rights, General Comment No. 14-Article 12: The Right to Education (1999)

91. Committee on Economic, Social and Cultural Rights, General Comment No. 25-Article 15(1)(b), (2), (3) and (4): Science and Economic, Social and Cultural Rights (2020)

92. Human Rights Council Resolution 20/8 (16 July 2012), "The promotion, protection and enjoyment of human rights on the internet" adopted by the Human Rights Council in July 2012

93. UNGA Res 68/167-The right to privacy in the digital age (18 December 2013); UNGA Res 69/166-The right to privacy in the digital age (18 December 2014); UNGA Res 71/199-The right to privacy in the digital age (19 December 2016); UNGA Res 73/179-The right to privacy in the digital age (17 December 2018); UNGA Res 73/218-The right to privacy in the digital age (20 December 2019)

94. HRC Res 28/16 (March 2015); HRC Res 34/7 (March 2017); HRC Res 37/2 (March 2018); HRC Res 42/15 (September 2019).

The United Nations Office of the High Commissioner of Human Rights (OHCHR) drafted three separate reports on the right to privacy in the digital age,[95] and in 2015 mandated the creation of a new Special Procedure on the right to privacy.[96]

Various human rights mechanisms and advocacy organizations have pointed out how ICT technologies can be used to facilitate online content censorship (or moderation),[97] intentional disinformation campaigns,[98] restrictions on freedom of speech and expression, and online surveillance that directly interferes with the right to privacy.[99] Other mechanisms have addressed the ways in which the internet and internet-based technologies can amplify gender-based violence,[100] or the tendency for online hate speech to disproportionately impact women, LGBTQI+ communities, and racial, ethnic, linguistic or religious minorities.[101] As technologies mature, some of these pronouncements require periodic updating, giving the impression that efforts to craft meaningful human rights norms at the international level often lag years or decades behind the development of the technologies themselves. The Human Rights Committee's General Comment No. 34, 2011 in fact replaced its earlier General Comment No. 10 on the freedoms of opinion and expression and urged state parties to take into account the extent to which developments in information and communication technologies, such as internet and mobile based electronic information dissemination systems, have substantially changed communication practices.[102] Similarly, the HRC, in its General Comment No. 37 (2002), also discussed how the implications of the right to peaceful assembly have changed over time in light of evolving ICT capabilities.[103]

EXAMPLE 3:
Human Rights & Geoengineering

Geoengineering is the "deliberate large-scale intervention in the Earth's natural systems to counteract climate change."[104] There are generally two categories of geoengineering solutions. The first focuses on Solar Radiation Management (SRM), consisting of strategies to reflect part of the sun's radiation back into space, thus counteracting the rise in global temperatures. The most promising such geoengineering technology involves seeding the earth's stratosphere with sulphate particles that would emulate the effect of a volcano on the earth's upper atmosphere, blocking (temporarily) a fraction of the sun's energy and thus cooling the climate.[105] A second strand of research focuses on Greenhouse Gas Removal (GGR), which involves somehow removing carbon dioxide from the earth's atmosphere, either directly by carbon removal technology or indirectly, possibly by stimulating the oceans' capacity to absorb CO^2.

Promises

The promise of geoengineering technologies is substantial. Climate scientists are increasingly pessimistic about the chances of our world escaping the worst ravages of climate change. Each successive year, scientists issue ever-direr warnings about what will happen if current policies do not change, and every year policy-making efforts fall shorts of the desired targets. Increasingly, geoengineering technologies, which have been often criticized by mainstream climate scientists as unproven and potentially reckless (given their propensity to reassure policy makers that a simple technological "fix" to the climate crisis might be close at hand, thus obviating the urgency to make costly emissions reductions in the short-term), have been re-embraced as possibly the last chance humanity has to save itself from the worst effects of unchecked climate change. This would carry clear positive implications for the enjoyment of all human rights that depend on the environment, most notably the right to enjoy a clean, healthy and sustainable environment, as well as the right to life and development for potentially millions of people who are increasingly living in the path of accelerating climate change fueled natural disasters.

Risks

The risks of geoengineering are also substantial. The UN Special Rapporteur on human rights and the environment, in a 2021 report to the UNGA, warned that geoengineering technologies "could have massive impacts on human rights, severely disrupting ocean and terrestrial ecosystems, interfering with food production and harming biodiversity."[106] The Intergovernmental Panel on Climate Change (IPCC) has consistently warned of geoengineering's risks to people and ecosystems, which remain poorly understood. There are

inherent uncertainties involved in almost any geoengineering project. Some worry that SRM technologies may alter weather patterns, for example the South Asian monsoon, thereby potentially disrupting the livelihoods of millions of people in parts of the world that depend on those seasonal weather patterns, while leaving largely unaffected those living in other parts of the world. From a human rights perspective, the inherent inequality of who is likely to bear the most substantial risk (and who the benefit) from the hypothetical use of such technologies is highly concerning. Moreover, human rights activists have raised concerns that most solar geoengineering experiments are currently planned or implemented on Indigenous territories with often inadequate provisions for securing the free, prior, and informed consent of impacted communities and with only insufficient public oversight.

Proposed Solutions

Those who advocate for at least considering geoengineering technologies as part of a global response to climate change are well aware of these ethical and human rights concerns, and have proposed several solutions they claim will protect against harm. Oxford University's Geoengineering Program, for example, has proposed a set of guiding principles for the governance of geoengineering projects "from early research to the point where they may be available for eventual deployment."[107] These principles hold that (1) geoengineering should be regulated as a public good, meaning that private corporations can and should play a role in their delivery, but that the governance of that technique should always remain with public authorities; (2) public participation should inform the decision making process; (3) geoengineering research must be made public; (4) the impacts of any geoengineering research should be independently audited, and (5) geoengineering research should not take place before any governance structures are put in place.

In the absence of any national or regional regulatory guidance, academic institutions are also beginning to think seriously about the ethics of approving open-air geoengineering experiments. Harvard University, for example, in 2019 set up an ethics committee specifically for geoengineering, tasked to "ensure that researchers take appropriate steps to limit health and environmental risks, seek and incorporate outside input, and operate in a transparent manner."[108] Critics of the Harvard approach stressed the potential for the oversight board to be insufficiently independent from the university, and also highlighted the potential for fierce social backlash against such research if done without a necessary national consensus.[109]

Human rights organizations have also expressed concern about the potential risks of geoengineering

95. OHCHR, Report on the right to privacy in the digital age (2014); OHCHR, Report on the right to privacy in the digital age (2018); OHCHR, Report on the right to privacy in the digital age (2020).

96. OHCHR, Special Rapporteur on the Right to Privacy, https://www.ohchr.org/en/special-procedures/sr-privacy

97. Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression: Report on content regulation, 6 April 2018, https://www.ohchr.org/en/calls-for-input/report-content-regulation

98. Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression: Report on disinformation, 13 April 2021, https://www.ohchr.org/en/calls-for-input/report-disinformation

99. Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression: Report on the adverse effect of the surveillance industry on freedom of expression, 28 May 2019, https://www.ohchr.org/en/calls-for-input/report-adverse-effect-surveillance-industry-freedom-expression

100. Report of the Special Rapporteur on violence against women, its causes and consequences on online violence against women and girls from a human rights perspective, 18 June 2018, https://www.ohchr.org/en/documents/thematic-reports/ahrc3847-report-spe- cial-rapporteur-violence-against-women-its-causes-and

101. Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression: Report on online hate speech, 9 October 2019, https://www.ohchr.org/en/documents/thematic-reports/a74486-report-online-hate-speech

102. Human Rights Committee, General Comment No. 34-Article 19: Freedoms of Opinion and Expression (2011), CCPR/C/GC/34

103. Human Rights Committee, General Comment No. 37-Article 21: Right of Peaceful Assembly (2020), CCPR/C/GC/37

104. Oxford Geoengineering Programme, "What is Geoengineering?", http://www.geoengineering.ox.ac.uk/www.geoengineering.ox.ac.uk/ what-is-geoengineering/what-is-geoengineering/

105. Wake Smith and Gernot Wagner (2018) Environmental Research Letters 13 124001

106. OHCHR (2019), "Safe Climate: A Report of the Special Rapporteur on Human Rights and the Environment"https://www.ohchr.org/ sites/default/files/Documents/Issues/Environment/SREnvironment/Report.pdf org/

107. Oxford Geoengineering Programme, "What is Geoengineering?", http://www.geoengineering.ox.ac.uk/www.geoengineering.ox.ac.uk/ oxford-principles/principles/

108. Temple, James, "Geoengineering is very controversial. How can you do experiments? Harvard has some ideas." (July 29, 2019) MIT Technology Review

109. Ibid. Winickoff, David, and Mark Brown (2013) "Time for a Government Advisory Committee on Geoengineering Research" 24(4) Issues in Science and Technology, https://issues.org/time-for-a-government-advisory-committee-on-geoengineering-research/

and called for governance frameworks to be grounded in the right to a healthy environment. Substantively, this requires recognizing the interdependence of the natural environment with basic human rights, and procedurally it requires the development of robust public participation processes. The Human Rights Council also expressed concern about the potential human rights impacts of geoengineering, and – in its 2021 Resolution 48/14 establishing a Special Rapporteur on human rights and climate change, also tasked its Advisory Committee "to conduct a study and to prepare a report, in close cooperation with the Special Rapporteur [for submission] to the Council at its fifty-fourth session."

Outside of the few voluntary codes of conduct and university review processes described above, few credible governance structures currently exist that would regulate or guide such research, much less the potential deployment of any real geoengineering interventions. Efforts to create such governance models at a global level are only just getting underway. In 2013, David Winickoff and Mark Brown proposed a series of standards designed to build a trusted institution that can regulate and oversee research in this area as well as (in the future) make decisions about the potential deployment of any geoengineering solutions.[110] Pascal Lamy, former Director General of the World Trade Organization and former EU Trade Commissioner is currently leading an effort to develop and promote various governance structures relevant to geo-engineering.[111] Lamy predicts that these governance structures will likely need to operate at the international level, thus avoiding the temptation for any one government or leader to take the task of regulating the global climate into his or her own hands. Futurists, on the other hand, ominously predict that precisely this risk poses a significant threat to a stable and coordinated global response to climate change, especially if and when climate-change continues to cause increasingly costly natural disasters in certain countries with the technological capacity to implement such geoengineering projects.[112]

EXAMPLE 4:
Human Rights & Artifical Intelligence

The Organisation for Economic Cooperation and Development (OECD) defines an Artificial Intelligence (AI) system as a "machine- based system that is capable of influencing the environment by producing an output (predictions, recommendations or decisions) for a given set of objectives." Such AI systems use "machine and/or human-based data and inputs to (i) perceive real and/or virtual environments; (ii) abstract these perceptions into models through analysis in an automated manner (e.g., with machine learning), or manually; and (iii) use model inference to formulate options for outcomes. AI systems are designed to operate with varying levels of autonomy."[113] This broad definition moves on from first generation rule-based systems, which are no longer considered to be "new or emerging" by technologists, to encompass modern machine learning systems.

As an initial matter, AI systems should be distinguished from simple algorithms. An earlier generation of AI in the 20th century, often called rule-based expert systems, tried to realize AI through pre-defined logic. The algorithms that would be used in these expert systems would instruct a computer what to do with a predefined set of inputs.

An algorithm, for example, might instruct a computer program to take a certain entry in a dataset and multiply it by two. Why two? – it would be the task of a computer programmer to determine the correct parameter (which in this case would be "2"). In more modern machine-learning AI systems, on the other hand, the AI system itself automatically 'learns' a model (a set of parameters) from data without being explicitly programmed. The AI system on its own determines the appropriate parameter to use, for example by conducting a complicated analysis of large volumes of data to determine that value. AI systems are built using algorithms, but they distinguish themselves from simple algorithms by having autonomous learning and decision-making capabilities. While rule-based expert systems are inductive and logical, machine learning AI systems are deductive, statistical, and data-driven.

Alan Turing, as early as 1950, suggested that machines that could fool another human into believing the machine was actually a human should be considered as "intelligent" machines.[114] While this so- called "Turing Test" has remained controversial ever since it was first proposed, it is safe to say that for an increasing number of human functions, machines are rapidly approaching or even exceeding human levels in terms of their performance.

Machine learning derives its models from past data (or experiences) and using those insights to predict future behavior. This type of AI can self-adjust its predictive modelling logic based on new data inputs and is therefore capable of 'learning' and 'predicting,' not just 'doing.' In particular, multi-layered perceptron, or "deep learning" is framed on "artificial neural networks," in which data (inputs) are processed not all at once (a single layer of analysis) but rather across a series of interim analytical processes (including "forward feeding" and "backpropagation"). What humans do is only to design a basic structure (for example, setting the numbers of layers and nodes) and then feed data to the system. It is then the machine's task to calculate the degree at which each node affects subsequently connected nodes (which are called weights and biases, or collectively "parameters"). As such, deep learning is often called "end-to-end learning" or a "black box." This technology, which was rapidly improved since the early 2010s,[115] is particularly useful in analyzing non-linear, unstructured data sets (for example natural human language, images, or the natural environment as seen through the lens of a digital camera). Such AI is opening the doors on NETs like autonomous driving cars,[116] robots that can maneuver autonomously in the built environment,[117] speech recognition,[118] toxicology testing for new drugs,[119] medical image analysis,[120] and even as a replacement for traditional social science opinion surveys.[121]

More recently, deep learning AI technologies have themselves evolved into Convolutional Neural Network (CNN), Recursive Neural Network (RNN), and Transformer Network (TN) models of AI. The TN model of AI, which is built on a self-attention algorithm, is currently regarded as "state-of-the-art", and has recently been applied to various downstream tasks such as language models including GPT-3 and BERT or AI-powered painting. The Human Centered AI Program at Stanford University (Stanford HAI) recently described the TN model the "foundation model" and warned that it might generate homogenous social harms while being used for multiple tasks.[122]

To fully pass the Turing Test, future AI systems may also need to develop increasingly sophisticated theory-of-mind strategies (the ability to emulate human "common sense" or emotions) as well as self-awareness.[123] While recognizing how such an idea might cause concern among lay observers, technologists also warn that by focusing too much on this issue one might inadvertently distract attention from more immedi-

110. Winickoff, David, and Mark Brown (2013) "Time for a Government Advisory Committee on Geoengineering Research" 24(4) Issues in Science and Technology, https://issues.org/time-for-a-government-advisory-committee-on-geoengineering-research/

111. Harvey, Fiona, "Climate geoengineering must be regulated, says former WTO head," (May 17, 2022) The Guardian, https://www.theguardian.com/environment/2022/may/17/climate-geoengineering-must-be-regulated-says-former-wto-head (accessed Nov. 13, 2022)

112. Kim Stanley Robinson (2020), Ministry of the Future, New York, NY: Orbit).

113. https://oecd.ai/en/ai-principles. There are many different definitions of AI, depending on who is defining them and the purposes for that definition, and thus it is a quite fluid concept. See e.g., Carlos Ignacio Gutierrez, Anthony Aguirre, Rosto Uuk (2022), "The European Union Could Rethink its Definition of General Purpose AI Systems (GPAIS), Nov. 7, 2022, OECD.AI Policy Observatory, https://oecd.ai/en/wonk/eu-definition-gpais (last accessed Nov. 30, 2022)].

114. Alan Turing (1950) Computing Machinery and Intelligence. 49 Mind 433-460

115. House, Bryan (2019), "2012: a Breakthrough Year for Deep Learning," Deep Sparse, https://medium.com/neuralmagic/2012-a-break- through-year-for-deep-learning-2a31a6796e73, (accessed Nov. 19, 2022). The first multi-layered perceptron in history was invented by Werbos in 1974

116. Barla, Nilesh (2022), "Self-Driving Cars with Convolutional Neural Networks (CNN)," MLOps Blog, https://neptune.ai/blog/self-driving-cars-with-convolutional-neural-networks-cnn (accessed Nov. 19, 2022).

117. Rajan, Kanna and Alessandro Saffiotti (2017), "Towards a science of integrated AI and Robotics, 247 Artificial Intelligence 1-9

118. Amberkar, Aditya, Parikshit Awasarmol, Gaurav Deshmukh, and Piyush Dave (2018), "Speech recognition using recurrent neural networks." In 2018 international conference on current trends towards converging technologies (ICCTCT), pp. 1-4

119. Ciallella HL, Zhu H. Advancing Computational Toxicology in the Big Data Era by Artificial Intelligence: Data-Driven and Mechanism-Driven Modeling for Chemical Toxicity. (2019) 15:32(4) Chem Res Toxicology 536-547

120. Ohad Oren, Bernard J Gersh, Deepak L Bhatt (2020) "Artificial intelligence in medical imaging: switching from radiographic patholo- gical data to clinically meaningful endpoints" 2 The Lancet Digital Health e486-488, https://doi.org/10.1016/S2589-7500(20)30160- 6 (accessed Nov. 19, 2022)

121. Argyle, Ethan Busby, Nancy Fulda, Joshua Gubler, Christopher Rytting, and David Wingate (2022). "Out of One, Many: Using Language Models to Simulate Human Samples," https://www.researchgate.net/ (accessed Nov. 19, 2022)

122. Several authors, "On the Opportunities and Risks of Foundation Models", Center for Research on Foundation Models (CRFM) Stan- ford Institute for Human-Centered Artificial Intelligence (HAI), https://arxiv.org/pdf/2108.07258.pdf

123. Tegmark, Max (2017), "Consciousness" in Max Tegmark Life 3.0, New York, NY: Vintage Books, 281-315 (Describing the concept of "singularity," and how tremendously difficult it is to even conceptualise how a machine learning AI system may ever develop a sense of consciousness, at least according to current AI knowledge.)

ately relevant (and no less urgent) human rights issues related to existing AI technologies.[124]

As computing power has continued to grow in recent years, AI systems have become increasingly able to outperform humans, even highly trained specialists, at certain narrowly defined tasks (often described as "Narrow AI" or "weak AI"). While there is no consensus as to when exactly this might occur (or if it perhaps has already occurred), experts are confident that AI systems will soon attain higher degrees of freedom and be able to meet or exceed human-level abilities within the next few decades.[125] Once this threshold is crossed, scientists and philosophers will have to deal with the prospect of machines that can "out-think" even the most brilliant human minds, using logic that may exceed the capacity of humans to easily comprehend. So far, this is the domain of science fiction, and yet, as one CEO of an AI company recently put it: "[if] this type of AI is successfully created, no one knows what the impact will be."[126]

Promises

As described above, AI is not one specific technology, but rather a technology that can be embedded within other strategies to render them more efficient, more precise, and more effective. Thus, AI holds the promise to improve any technologically-enabled strategy to promote and protect human rights. This is especially true for economic and social rights, which are often difficult for states to guarantee due to the costs of realizing them. Bringing those costs down by means of AI-enabled technologies would subsequently make it more likely that states could progressively realize those rights, as defined in the ICESCR. One study, built on a "consensus based expert elicitation process" found that AI had the potential of advancing 82% of the indicators in the Sustainable Development Goals (SDGs) related to social outcomes, 70% of those related to economic outcomes, and 93% of those related to environmental outcomes.[127]

"[In] SDG 1 on no poverty, SDG 4 on quality education, SDG 6 on clean water and sanitation, SDG 7 on affordable and clean energy, and SDG 11 on sustainable cities, AI may act as an enabler for all the targets by supporting the provision of food, health, water, and energy services to the population. It can also underpin low-carbon systems, for instance, by supporting the creation of circular economies and smart cities that efficiently use their resources."[128]

This finding mirrors the historical pattern described in Chapter 1 whereby technology has always been embraced by the development community as one of the primary drivers of progress and improved human security.

AI can also be deployed to protect and promote civil and political human rights. NAVER (the South Korean technology company) for example, has used AI as part of a charitable campaign designed to raise awareness about the lives of people suffering from cerebellar atrophy, a rare and unfortunately incurable neurological disorder.

In an effort to increase awareness about the disease, Naver used its Cloud-Based Virtual Assistant (CLOVA) to analyze the handwriting of a patient suffering from this disease to create a font named "Let's Walk Together" that could be freely downloaded through its philanthropic service HappyBean, prompting (so far) over 6,000 donations to support patients with this and other rare and incurable diseases.[129]

Similarly, researchers at MIT's Lincoln Laboratory have been "developing machine learning algorithms that automatically analyze online commercial sex ads to reveal whether they are likely associated with human trafficking activities and if they belong to the same organization." Using natural language processing, researchers are rendering visible transnational human trafficking networks and passing that information along to law enforcement authorities. This information is indexed and sorted into "three major buckets — text, imagery, and audio data. These three types of data are then passed through specialized software processes to structure and enrich them, making them more useful for answering investigative questions." Using facial recognition algorithms the researchers can "identify additional victims and corroborat[e] who knows whom." Finally, researchers can "allow investigators to partially transcribe and analyze the content of [jail phone calls from suspects who are awaiting trial, for indications of witness tampering or continuing illicit operations]."[130]

Public authorities are also using AI to enhance and streamline their services. Using AI methods, national security agencies are increasingly deploying "predictive analytics for terrorist activities; identifying red flags of radicalization; detecting mis-information and disinformation spread by terrorists for strategic purposes; moderating and taking down harmful, terrorist or extremist online content; countering terrorist and violent extremist narratives; and managing heavy data analysis demands."[131] All of these strategies, while traditionally thought of as national security strategies, also inure to the State's obligation to protect the right to life of its citizens.

Some governments, notably the Republic of Korea, are also developing so-called "welfare technology" to improve their social welfare system,[132] integrating data from across numerous administrative agencies and combining them with IoT gadgets to better target social services and resources to previously 'invisible' populations. Other technologists are developing AI systems that provide customized solutions to differently abled individuals.

Finally, by its very nature as a data-driven technology, machine learning (assuming it operates without bias) could foreseeably be more predictable and more consistent than human-based discretion and cognition, which of course also suffers from the very human traits of getting tired, making simple mistakes, and the more insidious phenomenon of unconscious implicit bias that most of us have even without knowing it.[133]

Risks

For many people, AI conjures visions of killer robots, perhaps animated by cinematic dramas about machines having reached self- consciousness and turning on humanity. Such dystopian scenarios, which campaigners draw upon to great effect,[134] often serve as a distraction from the debate about technology and human rights. AI is not yet "fully autonomous," nor is it clear to most technologists what that would necessarily entail, even if it were technologically conceivable. Such scenarios, therefore, while certainly not inappropriate as topics of discussion and research, are not yet imminent present-day threats. Short of worrying about killer robots, there are numerous other human rights issues that also deserve our urgent and more immediately actionable attention.

One of the biggest technological problems associated with AI is that the datasets used to 'train' the AI systems often reflect significant biases within them that then get replicated in the resulting AI models.[135] Thus, for example, AI systems designed to predict crime in urban areas are trained using historical crime data risk simply amplifying and reinforcing the racial and socio-economic biases that informed that training data. Machine learning cannot distinguish legitimate data patterns from illegitimate or illegal biases, and thus risks silently perpetuating human biases while also 'sanitizing' them in the guise of the 'objective' but inscrutable logic of AI. This has made some AI models the silent modern-day enforcer of age-old stereotypes, biases, prejudices and inequalities.

AI ethicists point out that such examples illustrate two interrelated (but separate) types of harms. The first are so-called "allocative" harms, which occur "when oppor-

124. Will Knight "Forget Killer Robots – Bias is the Real AI Danger" Oct. 5, 2017, Business Insider, https://www.businessinsider.com/killer-robots-biases-artificial-intelligence-ai-2017-10?r=US&IR=T.)

125. Bostrom, Nick (2014), Superintelligence: Paths, Dangers, Strategies, Oxford, UK: Oxford University Press, 23 (describing four sepa- rate surveys of technologists, conducted between 2011 and 2013, that found that these individuals collectively estimated that by 2022 (10% likelihood), 2040 (50% likelihood) or 2075 (90% likelihood) technologists would have likely succeeded in developing Human Level Machine Intelligence (HLMI)

126. Betz, Sunny(2022), "4 Types of Artificial Intelligence," BuildIn (Aug. 25), https://builtin.com/artificial-intelligence/types-of-artificial-intelligence (accessed Nov. 19, 2022).

127. Vinuesa, Ricardo et. al (2020), "The role of artificial intelligence in achieving the Sustainable Development Goals" 11:233 Nature Communications 1-10, https://doi.org/10.1038/s41467-019-14108-y (accessed Nov. 19, 2022).

128. Ibid., at 2

129. NAVER-SAPI AI Report (2022), https://www.navercorp.com/navercorp_/research/2022/20221128101249_2.pdf)

130. Foy,Kylie,"Turningtechnologyagainsthumantraffickers"(May6,2021),https://news.mit.edu/2021/turning-technology-against-human-traffickers-0506 (accessed May 19, 2022)

131. UN Office of Counter-Terrorism (UNOCT) and UN Interregional Crime and Justice Research Institute (UNICRI) (2021), Countering Terrorism Online with Artificial Intelligence: an Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia, New York, NY: United Nations

132. Soyun Choi et. al (2022) "Ethical Use of web-based Welfare Technology for Caring Elderly People who Live Alone in Korea: A Case Study, 21(4) J. of Web Engineering 1239-1264

133. https://implicit.harvard.edu/implicit/takeatest.html See also Bonezzi, A., Ostinelli. M., & Meltzner, J. *(2022) "The Human Black-Box: The Illusion of Understanding Human Better than Algorithmic Decision-Making. 151(9) J. Exp. Psychol. Gen. 2250-2258.].

134. Stop Killer Robots Campaign, https://www.stopkillerrobots.org (last accssed Nov. 19, 2022).

135. O'Neil, Cathy (2016) Weapons of Math Destruction, UK: Penguin

tunities or resources are withheld from certain people or groups".[136] Allocative harms are the ones we read about in the newspaper, and the ones we can attempt to measure (for example, by comparing the outcomes of AI decisions based on race, gender, age etc.) These allocative harms are often compounded by a second type of harm – so called "representational" harms, which occur "when certain people or groups are stigmatized or stereotyped."[137] Representational harms imprint themselves in the psyches of victims and observers, manifesting in subtle and often impossible-to-detect ways, such as when a child growing up in a

minority neighborhood takes to heart the message, reinforced and "sanitized' by a biased AI system, that "kids in my neighborhood don't make it to college," for example. Such representational harms are far more difficult to measure, and perhaps far more difficult to purge even after the problem has been identified.

Another example of representational harms comes from the process by which modern-day natural language processing (NLP) AI systems "learn" to emulate human speech. Such AI systems "train" using content found on the internet, and can therefore result in the system generating prejudiced, offensive or otherwise inappropriate language.[138] Researchers have found, but been unable to pinpoint exactly why, more modern AI systems, which were trained on bigger datasets, tended to generate more toxic and stereotyped language than their older predecessor systems that had been trained using smaller datasets.[139]

Others have pointed to AI as a modern-day threat to privacy. Especially those AI systems that depend on users providing their personal data in exchange for online services, serious privacy risks, since users are often not aware of how easy it is to "dox" or re-identify a person from supposedly 'anonymous' or 'pseudonymized' data.[140] NLPs have also been found to be susceptible to information retrieval attacks, where sophisticated users can use targeted prompts to get an AI system to reveal personal information embedded deep within its training datasets, including addresses, phone numbers, etc.[141] Such privacy violations can then easily lead to other forms of human rights abuse, especially in the hands of governments intent on denying those rights or criminal private actors. Governments can use AI systems, for example, to efficiently target certain minorities, political opponents, or other vulnerable communities. This can compromise a host of human rights across the civil, political, economic, social and cultural spectrum.

Given the human rights implications of the spread of AI technologies, most countries in the world are putting in place specific laws and policy frameworks to regulate the use of AI systems.

Other well-documented human rights impacts of AI might involve autonomous systems such as self-driving cars, drones, or robots navigating in crowded areas that inadvertently cause harm to humans in the vicinity, for example when a self-driving car fails to respond to a preventable accident, or when robots (including automated wheelchairs) maneuver in crowded pedestrian areas.[142] Such accidents cause an obvious threat to human life and well-being, even if arguably the non-AI powered alternatives to such systems might cause statistically even more harm.

A more systemic impact –one that harkens back to the 19th century luddite protests against improved weaving technologies, in which traditional weavers rose up in revolt against the technologies they accused of robbing them of their traditional livelihoods is the threat of AI systems to entire professional vocations. While technologies have long been eroding certain types of jobs, those have often tended to be the jobs of less socio-economically powerful groups. As AI systems gradually become more and more capable, however, they are increasingly able to replace higher-skilled workers, including doctors, lawyers, psychologists, accountants, stock brokers, artists, researchers, journalists, and others. As fewer and fewer professions become "safe" from predation by increasingly efficient and indefatigable AI systems, the risk of displacement, accompanied by the incumbent social disruption and poverty, becomes difficult to avoid.

At its most existential level, some critics have pointed out that AI systems still lack the humanity that is essential to certain care professions.[143] An AI-enhanced restaurant that replaces its wait staff with AI-based robots, for example, may be missing the intangible human element of care (a smile, a joke, an understanding glance from one parent to another) that makes such interactions more dignified when conducted between humans.

Proposed Solutions

Given the human rights implications of the spread of AI technologies, most countries in the world are putting in place specific laws and policy frameworks to regulate the use of AI systems. These national frameworks demonstrate a range of options on how to deal with AI, starting from those that focus heavily on regulation and accountability to others oriented more towards encouraging research and development in AI products through a lighter regulatory touch.[144]

A few such policy guidelines include various policy recommendations or guidelines designed to promote so-called "trustworthy AI", designed to bring AI more in line with many of the important human rights concerns described above.[145] Various other 'hard law' (enforceable) laws also exist, including regulations designed to govern automated decision-making[146] Similarly, there are also numerous legislations designed to curb the use of facial recognition technology, including those in the state of Virginia, and the cities of Boston and San Francisco. The European Union similarly is preparing more comprehensive regulatory frameworks of such technologies.

Researchers have found that developed economies are often far better at implementing legislation than those in the Global South, where merely implementing good law is rarely tantamount to effectively regulating a new technology such as AI that is almost always controlled and deployed by corporations operating from outside of those jurisdictions.

136. Suresh, Harini and Guttag, John (2021), "Understanding Potential Sources of Harm throughout the Machine Learning Life Cycle", https://mit-serc.pubpub.org/pub/potential-sources-of-harm-throughout-the-machine-learning-life-cycle/release/2

137. Ibid.

138. Victor Silva "The Hidden Dangers of Language Models: How Language Models Came to be 'Stochastic Parrots,' (Dec. 30, 2020) The Startup, https://medium.com/swlh/the-hidden-dangers-of-language-models-980ee0ccb5b4]

139. Khari Johnson, "The Efforts to Make Text-Based AI Less Racist and Terrible," (Jun. 27, 2021) Wired, https://www.wired.com/story/efforts-make-text-ai-less-racist-terrible/].

140. Lomas "Researchers spotlight the lie of 'anonymous' data" (July 24, 2019) Tech Crunch, https://tcrn.ch/2MbcDIu (accessed Nov. 19, 2022) (demonstrating that it is possible to de-anonymize "99.98% of individuals in anonymized data sets with just 15 demographic attributes.")

141. Nicholas Carlini, "Privacy Considerations in Large Language Models," (Dec. 15, 2020), Google Research Blog, https://ai.googleblog.com/2020/12/privacy-considerations-in-large.html.]

142. Pericle Salvini, Diego Paez-Granados & Aude Billard (2022), "Safety Concerns Emerging from Robots Navigating in Crowded Pedes- trian Areas" 14 International J. of Social Robotics 441-462]

143. Santoni, Filippo, and Aimee Van Wynsberghe (2016) "When Should we use Care Robots? The Nature-of-Activities Approach. 22(6): Science and Engineering Ethics 1745–1760. DOI 10.1007/s11948-015-9715-4

144. In Korea, for example, the government launched a "Digital New Deal" designed to encourage AI innovation and bring the tools of AI into the private sector. This policy approach has been focused on making public data sets available and putting in place legislation specifically designed to facilitate AI innovation. See Kyunhee Song, "Korea is leading an exemplary AI transition. Here's how." OECD.AI (March 10, 2022), https://oecd.ai/en/wonk/korea-ai-transition (accessed Nov. 19, 2022). China, which has also been an epicenter of AI innovation, has three concurrent policy approaches to AI in place, some of them designed to push technology developers to make AI systems more explainable and accountable. In this way, China—which is often disregarded as a center of AI governance innovations—is perhaps poised to push technology providers to develop new and innovative (and rights-conforming) strategies to satisfy Chinese regulators in ways that will also prove beneficial elsewhere. See Matt Sheehan, "China's New AI Governance Initiatives Shouldn't Be Ignored" Carnegie Endowment for International Peace (Jan. 4, 2022), https://carnegieendowment.org/ (accessed Nov. 19, 2022). The European Union is also hard at work harmonizing regulations across the Eurozone pertaining to artificial intelligence, especially to high-risk AI systems. See European Commission, Proposal for a Regulation laying down harmonised rules on artifi- cial intelligence (Apr. 21, 2021), Brussels: European Commission, https://digital-strategy.ec.europa.eu/en/library/proposal-regula- tion-laying-down-harmonised-rules-artificial-intelligence. The USA also released a proposed AI Bill of Rights, which highlights five principles that should hold true when deploying AI technologies. See Blueprint for an AI Bill of Rights, Washington DC: The White House, https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (accessed Nov. 19, 2022)

145. Examples include the U.S. Presidential Executive Order 13960 (2020) to "Promote the Use of Trustworthy AI in the Federal Govern- ment", the White House's 2022 draft AI Bill of Rights, the US National Institute of Standards and Technology (NIST)'s 2022 draft AI Risk Management Framework, and the Korean government's 2020 AI Ethics Guidelines.

146. Examples of such enforceable legislation includes the EU's General Data Protection Regulation (Art. 22), Korea's Credit Information Act (Art 36-2), California's Privacy Rights Act of 2020 (Sections 1798.185(a)(16), and the proposed Algorithmic Accountability Act of 2022 (H.R. 6580), which was proposed in the US Congress in 2022.

Corporations engaging in AI research also are actively promulgating AI Codes of Ethics. Some prominent examples include the AI ethics codes of Google,[147] Microsoft,[148] IBM,[149] Kakao,[150] and Naver,[151] to mention just a few.

One meta-study analyzing various proposed governance principles on AI put forward by various corporations, civil society organizations, governments, international organizations, and or multi-stakeholder collaborations found that they emphasized eight separate principles, and that there was an increasing convergence trend among these many documents.[152] These are:

1. **Privacy.**
   "AI systems should respect individuals' privacy, both in the use of data for the development of technological systems and by providing impacted people with agency over their data and decisions made with it.
2. **Accountability.**
   There should be "mechanisms to ensure that accountability for the impacts of AI systems is appropriately distributed, and that adequate remedies are provided."
3. **Safety and Security.**
   "AI systems [should] be safe, performing as intended, and also secure, resistant to being compromised by unauthorized parties."
4. **Transparency and Explainability.**
   "AI systems [should] be designed and implemented to allow for oversight, including through translation of their operations into intelligible outputs and the provision of information about where, when, and how they are being used."
5. **Fairness and Non-discrimination.**
   "AI systems [should] be designed and used to maximize fairness and promote inclusivity."
6. **Human Control of Technology.**
   "Important decisions [should] remain subject to human review."
7. **Professional Responsibility.**
   Individuals play a vital role "in the development and deployment of AI systems [and should consider it as their professional duty to ensure] that the appropriate stakeholders are consulted and long-term effects are planned for.
8. **Promotion of Human Values.**
   "The ends to which AI is devoted, and the means by which it is implemented, should correspond with our core values and generally promote humanity's well- being."

Finally, numerous serious research efforts are under way to pre-emptively address the still unanswered question of how to constrain AI if and when it ever becomes technologically possible to create super-intelligent, general AI,[153] such that dystopian fictions about robots dominating humanity remain strictly in the realm of hypothetical science fiction, not reality.

Conclusion

Chapters 1 and 2 have shown that the time is ripe to propose a holistic, multidisciplinary, and action-oriented approach to ensure that NETs serve as a force for progress and human rights in society, rather than the opposite. We are not the first generation to struggle with the impact of NETs on our societies, nor will we likely be the last. And yet we are faced with several NETs that promise to fundamentally disrupt the way our society works. In fact, many technologists pride themselves for developing so-called 'disruptive' technologies. Given the still tenuous foothold that modern human rights norms – barely 75 years old at the time this report goes to press – have in so many of our societies, we ought to be careful to avoid new technologies from "disrupting" our slow but steady progress towards their ultimate realization. As such, perhaps the time has come for us to discuss what it would take to 'nudge' NETs in the direction of actively promoting and protecting human rights. In other words, far from considering technology to be a morally neutral concept, we should actively seek out concrete strategies (broken down into individual processes) to structurally render new and emerging technologies into a force for good.

147. Google Corporation, Artificial Intelligence at Google: Our Principles, https://ai.google/principles/ (accessed Nov. 19, 2022)

148. Microsoft Corporation, Responsible AI, https://www.microsoft.com/en-us/ai/responsible-ai (accessed Nov. 19, 2022).

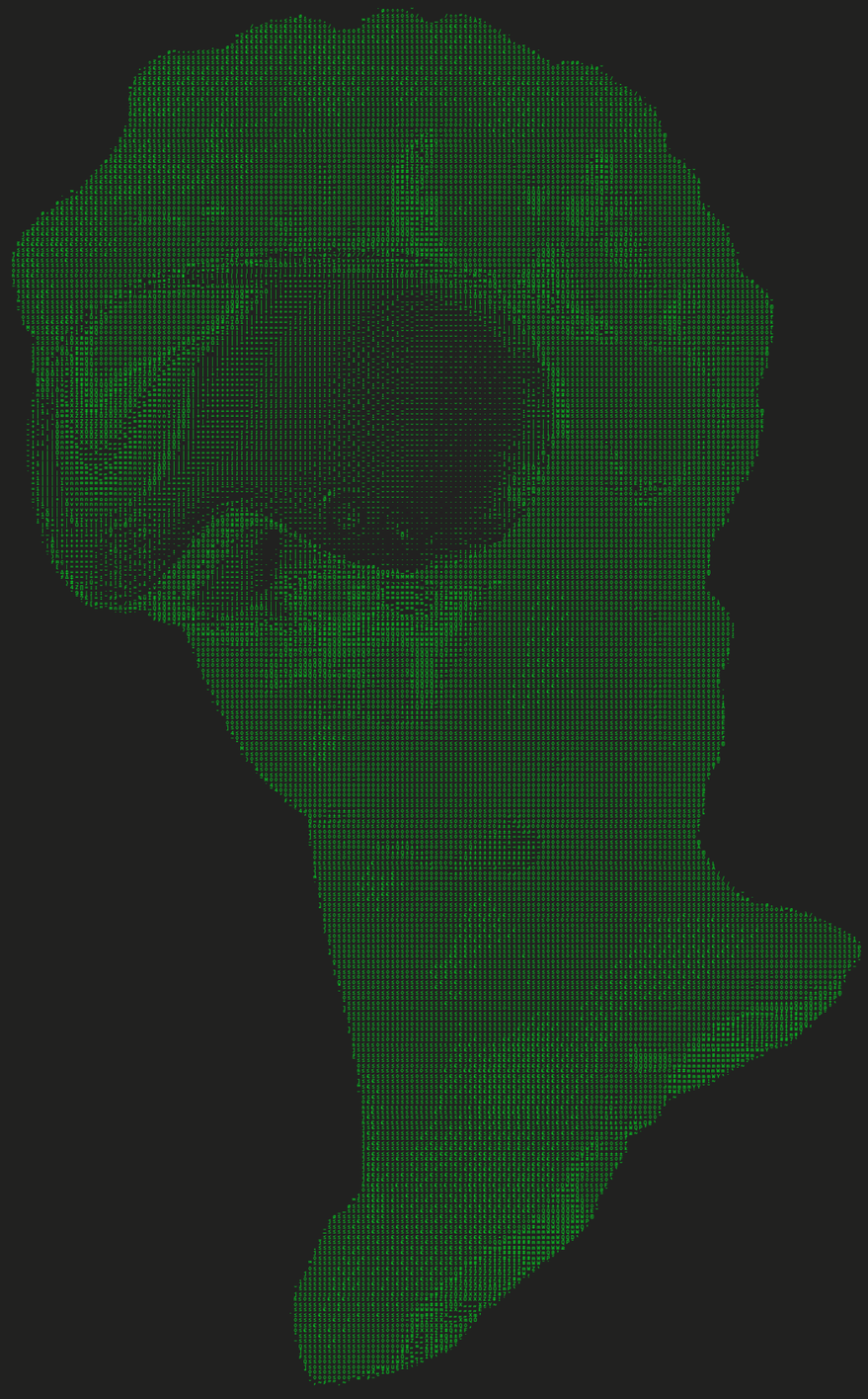
149. IBM Corporation, AI Ethics, https://www.ibm.com/artificial-intelligence/ethics (accessed Nov. 19, 2022)

150. Kakao Corporation, AI Ethics: Creating a healthy digital culture with technology and people, https://www.kakaocorp.com/page/res- ponsible/detail/algorithm?lang=ENG&tab=all

151. Naver Corporation, AI Ethics Principles, https://www.navercorp.com/en/value/aiCodeEthics (accessed Nov. 19, 2022)

152. Field, Jessica, et. al., "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI," Berkman Klein Center for Internet and Society Reesarch Publication Series No. 2020-1, https://cyber.harvard.edu/publication/2020/principled-ai (accessed Nov. 19, 2022)

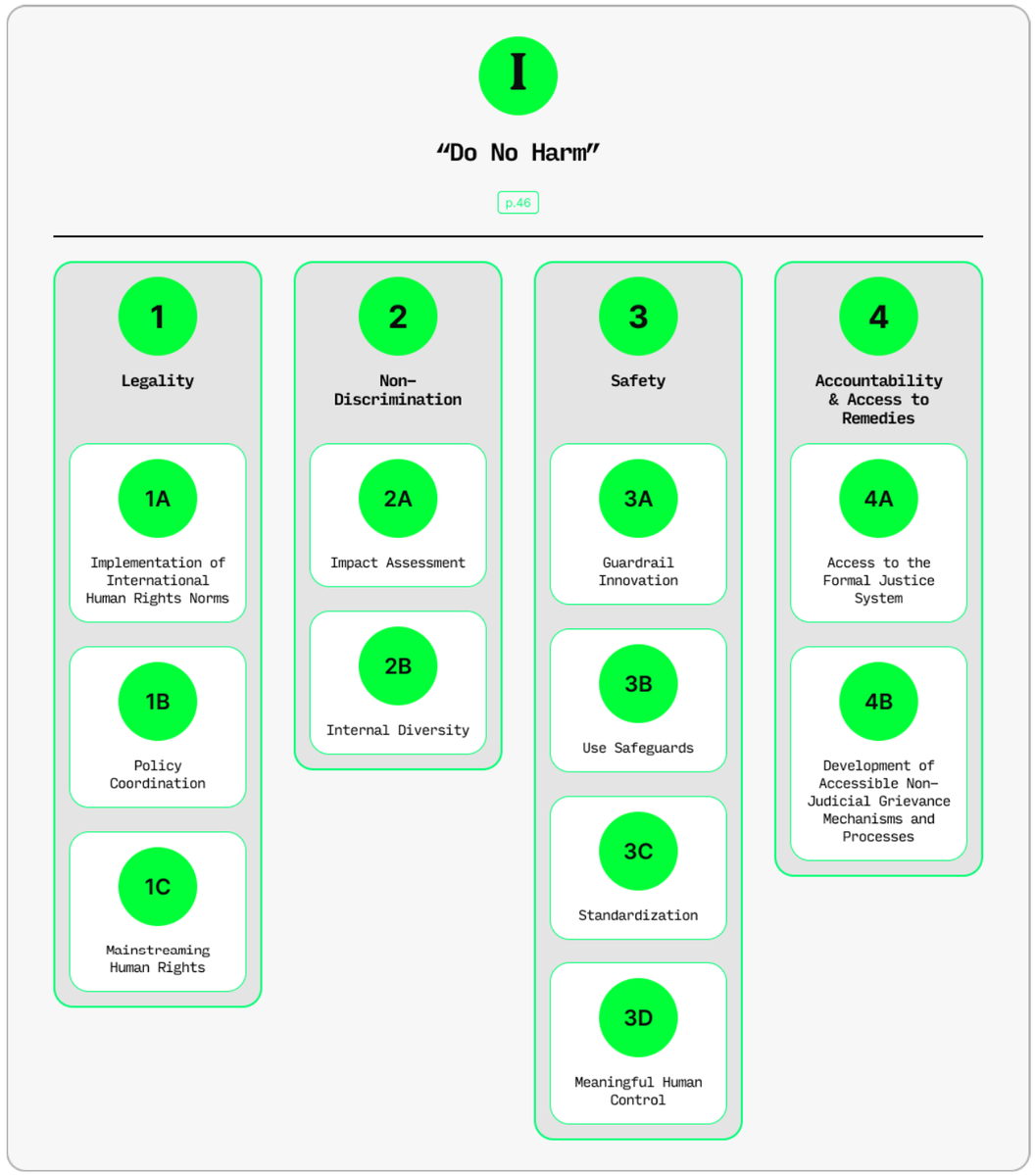
153. See e.g., Future of Life Institute, https://futureoflife.org (accessed Nov. 19, 2022).

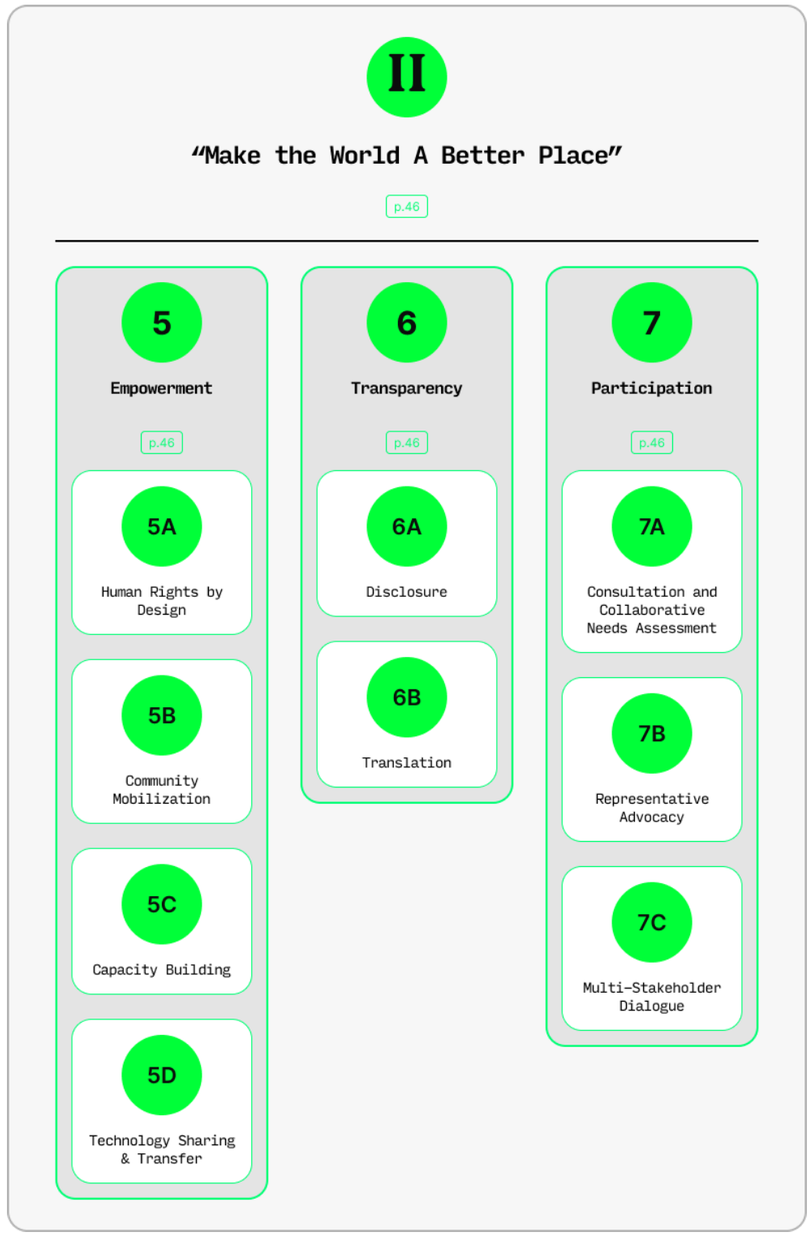


Part II

A Human Rights Based Approach To New And Emerging Technologies

The HRBA@Tech Model





The second part of this policy report proposes a way forward. The discussion so far has made clear that the question of how to nudge technological and scientific innovation towards a better, more equitable, more just, and more dignified world has been with us since at least the start of the industrial revolution, if not since the dawn of human history. The international community has been grappling with how best to conceptualize technological innovation from a human rights perspective since at least the 1960s. We have also seen, however, that many of these discussions end by merely reaffirming the 'paradox' of technology and human rights, namely that new technologies can be used either for good or for bad purposes. This duality often leaves one with the impression of having to choose between a tech utopian or a tech-phobic approach, corresponding with the vision of a deregulated libertarianism on the one hand and a much more precautionary approach on the other. When neither of those two extremes seems satisfying from a human rights perspective, many prior attempts to speak with more nuance on this issue have ended with a humble call for 'further study' of the issue, or perhaps a more targeted effort to regulate only one particular technology.

What follows is what we are calling the Human Rights Based Approach to New and Emerging Technologies (HRBA@Tech). It brings together two distinct strands of thinking that were apparent already in the debates of the 1960s and the 1970s. The first is the commitment to the gradual improvement of human society, articulated best by the bullish optimism of the international development sector. This approach embraces and celebrates the potential of NETs to solve many of our world's problems and "make the world a better place." It looks to technology and scientific innovation to help us find solutions to some of society's most vexing problems: global climate change, governance, the challenge of entrenched discrimination, social alienation and exclusion from mainstream society, public health threats, market inefficiencies, etc. The approach also, however, embraces the more precautionary (some might argue constructively realist) approach one might associate with the human rights movement, namely an insistence that "no one should be left behind."[154] In the push to make the world a better place, this second more cautionary approach also reminds us that individuals and vulnerable communities cannot be sacrificed at the altar of societal and technological progress. To make this a reality, some of the time- tested human rights strategies, built on accountability mechanisms and concrete incentives to nudge actors towards decisions that advance the cause of human rights, remain a necessary component of an overall strategy.

The human rights-based approach, originally devised in the context of development cooperation, is a framework for policy making and development programing anchored in international human rights standards. It provides a flexible model to guide technologists and policy-makers on how to uphold human dignity throughout the lifecycle of technology. Under the international human rights framework, States are and remain the ultimate duty bearers for the protection and promotion of human rights. Nonetheless, the human rights-based approach also provides a practical and flexible framework that applies also to non-state actors such as private companies, civil society actors, and academia, all of which are central to the NET space. The human rights-based approach remains conceptually tethered to international human rights norms, and yet operationally it is a broader framework of principles and processes that go well beyond the traditional human rights toolbox. The HRBA@ Tech model places human rights at the center of policy making, both as legal compliance standards and as an ethical frame of reference. It is malleable enough to encompass a range of governance and policy measures, undertaken by a wide variety of stakeholders, all designed to 'nudge' NETs in the direction of human rights. While the approach is grounded in human rights, the HRBA@ Tech model also proceeds beyond a mere articulation of principles and standards to identify a multidisciplinary range of core processes–some of them familiar to classical human rights actors and some of them less so.

The HRBA@Tech is presented in three separate Chapters. The first (Chapter 3) presents the framework itself. This Chapter can be seen as "the What" of the HRBA@ Tech – a vision of a future world in which advancements in technology and science are intentionally designed, produced, and deployed in ways that ensure a greater respect for the inherent human dignity of all who might be impacted by those technologies. This vision is broken down into two pillars, which themselves are sorted into sub-components, each of them corresponding to a discrete area where processes can be designed to make real on the promise of the HRBA@Tech model. This can be thought of as the aspirational vision of the HRBA@Tech model.

Chapter 4 focuses on "the How" of the HRBA@ Tech model. It proposes a general approach to understand and analyze the human rights implications of any new and emerging technology. It is structured to highlight a non-exhaustive list of potential "intervention points" along the classical Technology LifeCycle (TLC) of a NET, where various coalitions of motivated actors can begin to 'nudge' new scientific or technological innovations towards human rights or human dignity reaffirming outcomes.

Chapter 5, finally, focuses on "the Who" of the HRBA@ Tech approach. It explores the types of stakeholders that typically can collaborate— sometimes joining together in innovative and unusual coalitions—to design and implement the procedural safeguards to ensure that new technologies and scientific innovations lead to a better world.

The challenge in writing these three Chapters is to leave them sufficiently abstract that the framework can apply to any new or emerging technology (even those we cannot yet even imagine today), while also being specific enough to still offer the human rights and technology communities some concrete guidance on how to proceed.

154. United Nations, Department of Economic and Social Affairs, "The 17 Goals", https://sdgs.un.org/goals

CHAPTER 3: ELEMENTS OF THE HRBA@ TECH MODEL AND HOW THEY TRANSLATE TO NEW AND EMERGING TECHNOLOGIES

Chapter Summary:

This Chapter introduces the HRBA@Tech model. This Chapter describes The What of the HRBA@ Tech model, drawing inspiration from international human rights law as well as parallel discussions taking place in the field of technology ethics. It breaks these principles into two broad categories (or pillars) that collectively support the HRBA@Tech model. The first, the "do no harm" pillar, is further broken down into four constituent principles. The second, the "make the world a better place" pillar, is broken down into three constituent principles. Together, these seven principles provide the normative basis for the HRBA@Tech model.
These seven principles are associated with an exhaustive list of 24 core processes that, we argue, constitute the complete toolbox of strategies and methods that together can be used to 'nudge' technology in the right direction. If technology is not neutral (as we argued at the outset of this paper), these 24 processes constitute the proposed methodology we believe should use to ensure that new and emerging technologies are structurally biased in favor of human rights priorities.

The HRBA@Tech model can be thought of as a house built on two columns, with interlocking support beams to lend stability to the edifice. The first pillar focuses on the well-established obligation to "do no harm." The second focuses on how to "make the world a better place." To breathe life into these two foundational pillars of the HRBA@Tech Model, each pillar is broken down into several aspirational principles that provide the normative framework for the HRBA@Tech model. By thinking through the practical applications of these principles for different stakeholders, the HRBA@ Tech model moves beyond the aspirational language of principles and norms, and turn to the very practical business of bringing those norms to life. It proposes concrete processes that, taken in the aggregate, will ensure that the development and deployment of any new and emerging technologies can be consistent with (and in fact supportive of) the human rights agenda. The Chapter ends with an illustrative chart listing the 24 identified processes that might be used by various stakeholders to achieve the objectives associated with the HRBA@ Tech model.

Fundamental Principles of the HRBA@Tech Model		
1	**Legality**	States must enact laws to promote and protect human rights in the context of the development and deployment of NETs. Private companies and other stakeholders should fully respect human rights and take steps to support their full and effective realization.
2	**Non-Discrimination and Equality**	The use of new and emerging technologies must not intentionally or inadvertently discriminate against any persons or groups, even if doing so might (purportedly) allow for other persons or groups to enjoy an enhanced quality of life.
3	**Safety**	Safety concerns and adequate safeguards or "guardrails" must be integrated into the development of technology so that its deployment can adhere to intended use.
4	**Accountability & Access to Remedies**	Systems and mechanisms must be put in place to ensure that those responsible for the development and deployment of NETs face costs for not respecting human rights, while ensuring rights holders have an avenue to secure remedy for grievances.
5	**Empowerment of Vulnerable Populations**	Any new or emerging technology should be designed to make the vulnerable better off than they were before that technology existed. The best way to achieve this is through their empowerment.
6	**Proactive Transparency**	It is the duty of the technologists to disclose relevant information and to make a new or emerging technology understandable for non-technologists, policy makers, potential users of those technologies.
7	**Proactive Representation in Design and Implementation**	The only way to earn the trust of communities that stand to be affected by a new or emerging technology is to proactively involve them (or their representatives) in the design and implementation of that technology.

I

Do No Harm

The first pillar of the HRBA@Tech model is to "do no harm." Mary Anderson pioneered the "do no harm" concept in the field of humanitarian assistance[155] but of course this concept is also well known to medical professionals in the form of the Hippocratic Oath. In the context of NETs, this principle can be derived from the efforts of ethicists who see technology as neutral -- a mere instrumentality of humans acting with agency. For those thinkers, the focus should be on those human actors who intentionally or inadvertently deploy technologies for nefarious or non-human rights compliant purposes. This prong of the HRBA@Tech model seeks to raise awareness primarily among the users of a technology, incentivizing them to use it responsibly.

The 'do-no-harm' principle posits four principles, broken down further into nine constituent processes, that together serve to minimize and remedy those human rights harms as best possible.

This pillar draws heavily from the existing business and human rights framework described above (p.16). Just like the UNGPs, the framework speaks to both state and private (corporate) actors.

1

Legality

The principle of legality is an obvious and well-established corollary of the international human rights regime. According to classical human rights theory, States enter into mutually binding human rights agreements with each other or become subject to obligatory human rights norms as a matter of customary international law. Individual citizens, however (i.e., rights-holders who are the ultimate intended beneficiaries of those human rights protections), are only actually guaranteed those rights once States act on their duty to first recognize those human rights as legally enforceable entitlements within their domestic legal systems, usually when they adopt laws and policies for their implementation.

Since the adoption of the UDHR, an elaborate system has evolved to monitor State compliance with these universal human rights principles and to recommend policy prescriptions designed to improve the alignment of domestic laws and practices with international human rights norms. The principle of legality entails that States: 1) bind themselves to human rights obligations by becoming States Parties to the nine core human rights treaties;[156] 2) translate these international norms into domestic laws and policies for the effective enjoyment of human rights; and 3) implement recommendations from the various international human rights mechanisms to constantly improve their compliance with relevant standards.

The legality principle also requires domestic lawmakers to remain alert to potential threats (as well as opportunities) arising from NETs. This can be challenging, since lawmakers usually have a hard time anticipating the impacts of such technologies, and once potential human rights implications are more apparent it may often already be too late to retroactively 'legislate away' the damage. Thus, lawmakers have a unique obligation to onboard technical human rights and technology experts who can help 'translate' the human rights implications of new and emerging technology into policy language.

Notably, given the prominent role that private corporate actors play in the development, design and promotion of NETs, the principle of legality also requires States to take measures to protect against human rights abuse within their territory and/or jurisdiction by third parties, including business enterprises. In turn, this implies that businesses must fully comply with national laws and regulations protecting human rights.

The technical challenges of the State properly regulating NETs in line with applicable human rights norms under its jurisdiction should never serve as an excuse for private actors to ignore human rights principles when they begin to develop and deploy NETs. A private actor's responsibility to respect human rights 'exists independently of States' abilities and/or willingness to fulfil their own human rights obligations' and 'over and above compliance with national laws and regulations protecting human rights'.[157]

The UN Global Compact and the UN Guiding Principles on Business and Human Rights (UNGPs) are especially relevant in this regard. The Global Compact sets out principles for proactive engagement of businesses with regards to human rights (i.e., to respect and support human rights), while the UNGPs provide an authoritative, conceptual, and operational framework for States and companies to do so with respect to their human rights obligations and responsibilities.

The principle of legality also requires private corporations to articulate codes of conduct or ethics codes that 'make real' the company's commitment to protect and promote human rights. These policies should not merely enumerate aspirational goals, but also include concrete governance provisions designed to ensure that the company lives up to its aspirations.

New technologies illustrate the interdependence and indivisibility of the universal human rights corpus. As was demonstrated above, a technology designed to advance the right to education cannot at the same time undermine the right to privacy, even though the former is typically articulated in the ICESCR and the latter in the ICCPR. Since so many NETs aim to advance (i.e., "progressively realize") our ESC rights, and so many of the unintended impacts of new technologies threaten to impact our CP rights, the interdependence of all human rights is more apparent with regard to NETs than is often the case in other human rights discussions. Jurisdictions that have been hesitant to embrace certain human rights should realize that when it comes to regulating new technologies, it must be done in light of the entire corpus of human rights. The principle of legality does not presuppose that there will never be potential contradictions between competing rights claims. The right to free speech, for example, may be in tension with the right of women, girls, and sexual minorities to be free from revenge porn and harassment. When maneuvering such challenges, stakeholders must seek out nuanced approaches, balancing of interests of different individuals and communities within the context of a rights-based approach to NETs. The UN Office of the High Commissioner for Human Rights' Rabat Plan of Action, for example, provides one example on how to balance between freedom of expression (Article 19 of the ICCPR) and the right to be free from incitement of discrimination, hostility or violence (Article 20 of the ICCPR). Such standards can provide concrete guidance, for example, to content moderators on social media platforms when deciding on what kinds of speech to allow or disallow on their websites.

This further entails that international organizations and entities, such as the Office of the UN High Commissioner for Human Rights (OHCHR), the International Telecommunication Union (ICU), and the international human rights mechanisms, have key roles to play in clarifying the application of international human rights standards to particular technologies. These organizations and entities (along with others) are central to any efforts to ensure global policy coherence in the field of technology and human rights. These organizations are uniquely mandated to overcome governance and regulatory fragmentation by coordinating diverse international actors, as well as by monitoring compliance with international human rights norms, helping to build the capacities of various other stakeholders (including UN country teams, domestic policy-makers and rights holders), and by mainstreaming human rights throughout international, regional and national systems.

Finally, regional human rights instruments and standards can prove instructive for upholding the legality principle to the HRBA@Tech. Overarching human rights instruments such as the African Charter of Human and People's Rights, the American Declaration on the Rights of Man, the Arab Charter on Human Rights, the ASEAN Human Rights Declaration, and the European Convention on Human Rights, can help mainstream a HRBA@Tech model, and importantly, integrate crucial regional perspectives. Additionally, regulations such as the EU's General Data Protection Regulation (GDPR), or the US' proposed AI Bill of Rights, both of which provide guiding frameworks for privacy and data

155. Mary Anderson (2010), Do No Harm: How Aid Can Support Peace–or War, Boulder, CO: Lynne Rienner Publishers)

156. The nine core human rights treaties: Universal Declaration of Human Rights (1948), International Convention on the Elimination of All Forms of Racial Discrimination (1965), International Covenant on Economic, Social and Cultural Rights (1966), International Covenant on Civil and Political Rights (1966), Convention on the Elimination of All Forms of Discrimination against Women (1979), Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (1984), International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families (1990), Convention on the Rights of the Child (1989), Convention on the Rights of Persons with Disabilities (2006)

157. OHCHR, "Guiding Principles on Business and Human Rights", https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf

protection of EU and US citizens, can breathe further regionally relevant life into the principle of legality.

Process Associated with Legality		
1A	**Implementation of International Human Rights Norms**	Give domestic effect to international human rights norms. Ensure a balanced focus on both CP as well as ESC rights. This process depends on legislative reform efforts to give domestic effect to human rights, as well as capacity building and awareness raising activities to ensure that a nation's institutional framework is prepared to respect, protect and fulfil human rights.
1B	**Policy Coordination**	Develop coherent technical expertise at the United Nations to provide non-binding advice to member states on new and emerging technologies, their potential human rights implications and also comparative best practices on how to manage relevant risks and opportunities, and increase policy coordination amongst the various UN organs and agencies. This activity also involves standard setting, especially at the international and industry-wide levels.
1C	**Mainstreaming Human Rights**	This obligation applies to all stakeholders, namely to ensure that human rights norms and principles are mainstreamed throughout all relevant stages of the development and deployment of new and emerging technologies.

2

Non-Discrimination and Equality

The principle of equality and non-discrimination is a foundational principle of the international human rights framework. It posits that the inherent equal dignity of all human beings entitles them to enjoy all universally accepted rights and freedoms 'without distinction of any kind, such as race, color, sex, language, religion, political or other opinion, national or social origin, property, birth or other status' (Art. 2 UDHR). The right to equality and non-discrimination entails that every individual is entitled to be treated equally before the law. In other words, in equal circumstances two persons are to be treated in the same way.

Early on, the modern international human rights movement recognized that in order to fully realize equality, measures would have to be taken to rectify and counter past and existing discriminations. In this sense, the principle of equality and non-discrimination requires not just formal equality before the law but also a focus on the particular situation of vulnerable and marginalized populations, as well as an understanding of how a given measure (e.g., a law or a policy) can have structurally discriminatory effects even despite its application on equal terms. In other words, blind justice without an awareness of existing inequalities can constitute a harm by entrenching discrimination.

Artificial Intelligence tools, for example, have often been accused of discriminating against some individuals (for example racial minorities) in the name of efficiency. An example might be an algorithm that considers race (or information that acts as proxy for race) as a factor in calculating a customized life insurance policy, with the result that minorities (who have historically been discriminated against in such situations) may receive higher-priced product offers whereas non- minority customers receive better and more cost-effective services. This in turn, becomes a data point for future determination of life insurance policy for other individuals, creating a negative feedback loop and perpetuating a vicious cycle of discrimination by reinforcing existing structural inequalities or historical discrimination. Even if hypothetically speaking the aggregate impact of such a technology across all members of society might be positive, and even if this might be a potentially very lucrative market opportunity, this still would be an unjust use of such a technology.

For States, adherence to the principle of equality and non- discrimination in relation to new and emerging digital technology therefore requires more than just establishing general prohibitions on discriminatory treatment on the basis of protected characteristics (e.g., through constitutional protections or sector specific protections) - though this is a necessary starting point. It requires proactive consideration of how certain laws, policies and practices. In line with their commitments to the SDGs, States should collect disaggregated data, allowing them to better track and assess discriminatory impact (including multiple and inter-sectoral forms of discrimination) on particular groups. Such monitoring can inform national reports to international human rights mechanisms. Human rights impact assessments are especially important when public authorities use NETs, such as in the provision of social services or national security initiatives. The collection of such disaggregated data for purposes of assessing discrimination must be accompanied by adequate safeguards to ensure that the information gathered does not itself constitute a threat to privacy and data protection, perhaps at the hands of unscrupulous private or state-sponsored hackers.

As a corollary of their duty to protect against abuses from private entities, States must take effective measures to ensure that the principle of equality and non-discrimination is fully upheld by business enterprises. At an initial level, this means prohibiting intentional direct discrimination by companies, either in their internal policies or in their design and use of technologies. State protective measures should also help companies identify and prevent indirect discrimination by mandating or encouraging human rights impact assessments and due diligence measures. To be effective, impact assessments should be conducted regularly throughout the lifecycle of technology. Those who develop and deploy NETs, working either on their own or in partnership with other stakeholders, can also devise relevant methods, standards, and technical benchmarks to assess potential bias or discriminatory impact. And it goes almost without saying that once evidence of any such bias or discriminatory impact is discovered, this should lead immediately to a remediation process, even if not prompted by a concrete "grievance" lodged by an affected individual or community.

In a commercial or business setting, this assessment should start from a calculus of the impact of a NET on classic categories of 'vulnerable' individuals and groups. These categories might change depending on the technology, and yet there are certain categories that are good to start with as a 'baseline,' including women, children, older persons, persons living with disability, ethnic or racial minorities, sexual minorities, etc. Assessment should then move on to consider the potential negative effects on individuals and groups in local contexts that could be disproportionately affected. In the case of a "disruptive" technology. Let us imagine, for example, a ride-sharing app that may threaten to displace traditional taxi services in an urban area. In such a scenario the impact assessment should focus not only

on the impact of the new technology on the potential consumers of the proposed service, but also those whose livelihoods stand to be made redundant by the new technology. The principle of non- discrimination applies also to situations where a new technology promises to bring little benefit to anyone other than the investors who fund the new technology.

This is not to say, of course, that the responsibility to ensure this non- discrimination should rely solely on one actor (see below, Chapter 5). In the case of the disruptive ride-sharing technology, for example, the corporations have some responsibility as they develop their business model, the municipal and regulatory authorities have their responsibilities to manage social change and minimize the negative social and economic externalities of businesses operating in their jurisdiction, the individual consumers have a responsibility, and also those whose livelihoods are at stake have an obligation (or perhaps a strong incentive) to gradually re-skill and retool—presumably with the support of other actors who provide those opportunities—so that they too can continue to thrive in a changing and modernizing community.

Two other strategies, both having to do with diversity, additionally help those who would seek to develop or deploy NETs to avoid discrimination being inadvertently 'hardwired' into that technology. The first such strategy is to constantly work towards high levels of internal workplace diversity, not just as an overall statistic but also within the individual teams tasked with developing and deploying NETs. Cultivating workplace diversity is typically considered to be a human resources challenge and not a technological challenge, and yet it has a major if indirect impact on the integrity of NETs, primarily by ensuring that a plurality of views are systematically represented throughout an NET's technology lifecycle. Such efforts to cultivate internal workplace diversity can be further strengthened by the proactive participation measures described below, which fall under the "Make the world a better place" pillar of the HRBA@Tech Model. Non-discrimination requires diversity and representation, especially of vulnerable and marginalized communities, in the teams developing NETs, as well as at all other levels of an organization. It may also entail diversification across supply chains including representation in the choice of suppliers and vendors. In order to be truly meaningful,

this diversity must also be accompanied by an openness to free communication and engagement, thus ensuring that with the greater workplace diversity and representation comes also a corresponding diversification of viewpoints, perspectives, concerns, and departure points for creative brainstorming that flows naturally from such diversity.

Processes Associated with Non-Discrimination and Equality		
2A	**Impact Assessment**	**Anticipatory Impact Assessments (Vulnerability Assessments)** Those responsible for developing and deploying NET must assess the potential bias or discriminatory impact of the NET throughout all stages of the technology lifecycle. This includes assessing the impact of a new or emerging technology on classic or 'baselines' categories of 'vulnerable' individuals and groups, including women, children, older persons, differently abled persons, and minorities, amongst others. These categories might change depending on the technology in question or the local context or conditions, and can thus also include any other readily identifiable group that could be potentially affected by the NET. **Ongoing Impact Assessments (Testing NETs for Disparate Treatment)** Those responsible for developing and deploying NETs must also conduct ongoing impact assessment and monitoring and oversight to ensure there is no disparate treatment, either intentional or unintentional, as a result of the implementation of the NET. **Remediation:** For non-discrimination to be meaningful, when an impact assessment detects bias or discriminatory impacts of a NET, relevant stakeholders must take necessary follow-up action and remediate the shortcomings in good faith.
2B	**Internal Diversity**	To minimize the risk of potential bias or discrimination (or any other negative impact) being baked into a NET, stakeholders must ensure internal diversity and representation within an organization. Such internal diversity can take the form of representation within the teams associated specifically with the development and deployment of NETs, across supply chains (including the diversification of suppliers and vendors) and also within the general workforce at all levels including key leadership and decision-making positions and roles.

3 Safety

Those responsible for the development and deployment of new and emerging digital technologies must ensure the safety of technology at all times. The greater the implications of a NET on human dignity and human rights, the more acute this obligation becomes. For example, if a scientist invents a new and a more efficient lightbulb, the burden to think about the safety of that new technology is relatively minor since it is extremely unlikely the new lightbulb would have profound implications for humanity or society or on the human dignity or rights of anyone who used it.

There are NETs, however, which pose a heightened risk of such impacts. These include a variety of complex and sophisticated AI tools, which are gradually supplanting human decision-making and action at various levels. These powerful AI tools can increasingly rival human capabilities and sometimes pose significant challenges to the human dignity, agency, and self-determination of communities and individuals whose lives stand to be affected by these technologies. Germline genetic engineering can potentially alter the genetic makeup of future generations of humans. Geo-engineering with its potential to manipulate the environment risks, changing medium- to-long term weather patterns, and with it entire ecosystems. Those developing and deploying such technologies should bear a correspondingly high burden to ensure that these technologies are absolutely safe before they are released into the market or the open environment.

There are several components to ensuring safety with respect to such technologies. First, it requires incorporating safeguards or 'guardrails' from the earliest stages of innovation and design throughout the entire lifecycle of a product's development. These guardrails should be updated with every new feature or innovation associated with the technology. Guardrails might necessitate the creation of internal features in the NET that can act as 'emergency brakes' to prevent and avoid harms. They also require ensuring the incorruptibility of the technology itself from external risks, such as hacking and misuse or abuse by bad faith actors. While technologies will always be vulnerable to hacking, misuse, or abuse, the principle of safety necessitates taking a precautionary approach to these risks. This means ensuring that adequate safeguards are integrated directly into the technology or protocols used for a specific network. These safeguards should guarantee its integrity and prevent unintentional exposure of users or consumers to risks beyond what they have agreed to assume. In particular, technologies harnessing personal data including sensitive medical or financial information require higher levels of safety and robust safeguards to ensure protection of privacy and data security.

Such a precautionary approach would also require virtually absolute certainty that a technology is safe, and that human rights implications have been adequately accounted for before a NET can be ethically deployed.

Safety of NETs, especially complex technologies such as machine learning AI, also requires ensuring they perform as intended and are predictable and reliable. Those responsible for the development and deployment of such technologies must have safeguards in place, both technological and procedural, to ensure that such technologies are not used for purposes other than the legitimate intended aims (while of course acknowledging some degree of uncertainty in any such endeavor). This can involve adding design features directly into a technology, as well as purpose specifications for the use of a technology that can be included in contractual provisions between different entities relating to the use or licensing of a particular NET.

Devising such safety features or guardrails may require additional technological innovation. Collaboration amongst various stakeholders, including technologists, academic institutions, think tanks and other actors may be useful in bridging knowledge gaps and ensuring more effective safeguards. Stakeholders may consider sharing or transferring technologies or technical know-how about effective safeguards, and also provide other forms of assistance and capacity building (going beyond the "do no harm" approach) to developing and less developed nations, technologists, civil society organizations and other entities with fewer resources to secure their systems.

Standards are also relevant to ensure the safety of an NET. They help to establish meaningful guardrails and build crucial societal trust in an NET. Stakeholders, typically working through multi-stakeholder initiatives, can establish relevant industry or sectoral standards that provide minimum safety thresholds to be applied by those developing and deploying NETs. This is typically a role reserved for national regulatory bodies or international organizations like the ITU, but can also be facilitated by the important work of civil society organizations. Underwriters Laboratories Solutions, for example, is a US-based independent not-for-profit organization that has developed various general safety standards, including for autonomous systems, that can be tailored to any specific industry or application.[158] Designated standard-setting organizations also play a key role. The Institute of Electrical and Electronic Engineers (IEEE), for example, has developed various standards for the design and development of intelligent systems through its Global Initiative on Ethics of Autonomous and Intelligent Systems.[159] Standard setting organizations should always consider collaborating with relevant international human rights organizations or mechanisms to develop standards for NETs.

Governments must establish appropriate laws, policies and regulatory frameworks to ensure the safety of NETs. This includes drawing "red lines" where applicable, particularly for high-risk technologies, including those that may pose national security threats. When responding

158. UL Standards & Engagement, "Presenting the Standard for Safety for the Evaluation of Autonomous Vehicles and Other Products", https://ulse.org/UL4600; UL Solutions, "Autonomous Vehicle Safety Training and Advisory", https://www.ul.com/services/autono- mous-vehicle-safety-training-and-advisory

159. IEEE SA, "The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems", https://standards.ieee.org/industry-connections/ec/autonomous-systems

to such national security threats, however, governments must nonetheless comply with relevant human rights standards and principles including the customary prohibition on violating non-derogable human rights and limits on restricting any other human rights unless doing so is strictly necessary to tackle the national security challenge.

Regardless of whether such government frameworks are in place, private actors can (and should) also adopt self-regulatory standards, including industry-wide initiatives, to hold themselves to the highest possible standards. This can include self-imposed restrictions by companies or consortia of companies while developing or deploying certain technologies, or limits on research undertaken with respect to certain technologies by educational institutions.

NETs must always have a human-in-the-loop or some other design feature (a 'safety brake') that can prevent the spectre of a runaway autonomous technology. Such human control must be meaningful. In other words, human control or supervision must always be more than merely superficial involvement, for example a human pressing a button once virtually all other decisions have already been made by an automated system. Human operators of NETs must retain the prerogative for human intervention and independent judgements of an action's appropriateness or ethical legitimacy. Meaningful human control is necessary to pull the proverbial emergency brake if a technology spins out of control, to prevent or avoid any harms or unintended and unforeseen consequences, and to deal with nuances and unpredictable exigencies which humans may be better equipped to tackle than machines by virtue of our capacity for subconscious intuition, emotion, empathy, compassion and other quintessentially human subjectivities.

In the case of AI, ensuring meaningful human control over the technology can take a number of forms. First, of course, AI can be used to make mere suggestions to a human actor, who ultimately retains the ability to either accept or reject that automatically generated advice. This model leaves the human actor with the ultimate responsibility for how she ends up using the AI-generated content. Human control can also come in the form of guardrails built into an AI system, for example a large language model that is programed not to generate instructions on bomb-making, for example (as is true in the new ChatGPT system), or one that redirects users to sources of support when they ask, for example, how best to inflict self-harm upon themselves (as is true in Google's search function). Human oversight can also consist of efforts to periodically review an AI system to ensure that it is still operating as intended, and of course by actively monitoring the flow and nature of grievances that may or may not be making their way through a grievance procedure established as part of the accountability and access to remedy principle (see below).

It is also necessary for actors using such automated decision-making systems to be conscious of "automation bias" (the tendency to perceive AI-generated decisions as neutral, objective and accurate, and therefore authoritative) and for them to feel empowered to take active measures to overcome such bias. Such empowerment can only come about in institutional environments where individual (human) actors are not disincentivized to exercise their prerogative,

where there is a certain acceptance that humans may sometimes make an incorrect judgment call but still play a useful oversight role over technology, and where the known risks of a NET are clearly communicated to those who would use those technologies, in terms that they are able to comprehend (see discussion of transparency below). A judge or a public official using the results of an AI system to make risk-assessments about a potential defendant, for example, should be well briefed of the dangers of inherent bias in AI systems before using such a system as part of a sentencing process. Rigorously training these human operators of NETs not just on the technicalities of using an NET, but also the ethics of relying too heavily on that NET, serves as a crucial aspect of ensuring the safety of NETs.

Processes Associated with Safety		
3A	**Guardrail Innovation**	Safety requires adoption of a precautionary approach while developing and deploying new and emerging digital technologies and incorporation of safeguards or "guardrails" from the initial stages of innovation and design all through the technology lifecycle. It includes creation of internal features which act as "emergency brakes" within the technology to prevent and avoid harms. It also includes ensuring incorruptibility of the technology from external risks of hacking and misuse or abuse by bad faith actors.
3B	**Use Safeguards**	To minimize the risk of potential bias or discrimination (or any other negative impact) being baked into a NET, stakeholders must ensure internal diversity and representation within an organization. Such internal diversity can take the form of representation within the teams associated specifically with the development and deployment of NETs, across supply chains (including the diversification of suppliers and vendors) and also within the general workforce at all levels including key leadership and decision-making positions and roles.
3C	**Standardization**	Standards ensure safety, help establish guardrails and build trust in a new and emerging digital technology. Stakeholders including standard setting organizations, either on their own or through multi-stakeholder initiatives, can work towards establishing relevant industry or sectoral standards as minimum safety thresholds to be applied by those developing and deploying new and emerging digital technologies.
3D	**Meaningful Human Control**	Every new and emerging digital technology must have a human-in-the-loop and such human control must necessarily be meaningful. This meaningful human control can be applied through the technology lifecycle including the decision-making, technological and operational stages and can be spread amongst different actors.

4

Accountability and Access to Remedy

Accountability demands for duty-bearers to be identified if they fail to meet their obligations with regard to an NET. It also involves the creation of viable procedures and mechanisms for "consequences" to be levied in cases of non-compliance or negligence. Accountability is a multi-faceted concept that includes the identification of duty- bearers and the delineation of their corresponding responsibilities. As described above in the description of the importance of conducting ongoing impact assessments, in line with the principle of non- discrimination and equality, accountability also requires ongoing monitoring and oversight activities to understand how people's rights are being impacted by a given NET, and of course a procedure designed to provide effective remedies to any negatively affected individuals or communities.

States traditionally have been (and still remain) the primary duty- bearers within the human rights framework. They have the obligation to respect, protect and fulfil human rights within their jurisdictions at all times. Non-state or private actors, including corporations, are increasingly central to the realization of human rights, however, as described in Chapter 1 of this paper. This is especially true with regard to NETs, where private actors often make the most crucial decisions in the development and deployment of those technologies. Private actors are therefore obligated, in line with the provisions of the UNGPs, to respect the laws of the States in which they operate (including laws designed to defend and promote human rights) but also to pursue their own corporate responsibility to respect and promote human rights.

Accountability mechanisms can be both formal (judicial) in nature, but can also include informal or less formal (non-judicial) mechanisms and processes. Part III of the UNGPs contain a description of what constitutes meaningful access to remedies.

The discussion distinguishes state-run judicial mechanisms from informal grievance processes, which can be further subdivided into those administered by the state (for example ombudsman offices or human rights commissions) and those administered by a private company or as part of an industry consortium.

State-based judicial mechanisms typically involve formal judicial institutions and processes. In first order, this refers to the State's formal courts, but can also include regulatory agencies and human rights mechanisms. Seeking accountability through such fora often involves litigating civil or private law claims, in which an individual or a community 'sues' another individual or legal entity over harms caused, and in which remedies might typically include financial compensation for those harms. Alternatively a complainant might petition a relevant regulatory agency to intervene in a certain dispute, and of course it may always involve criminal prosecutions by the States against individual wrongdoers if their pattern of misbehavior rises to a criminal standard. Constitutional law, administrative law, or other regulatory frameworks also act as useful entry points for accountability and access to remedy through the formal justice system.

Non-judicial grievance mechanisms, both run by the State as well as other actors (ex: integrated grievance mechanisms within a corporate governance structure), can also act as important channels for accountability and access to remedy. The UNGPs detail a set of criteria that characterize an 'effective' non-judicial grievance mechanism (Article 31). These so-called "Ruggie Principles" hold that in order for a grievance process to be effective it must be:

a. **Legitimate:**
enabling trust from the stakeholder groups for whose use they are intended, and being accountable for the fair conduct of grievance processes;

b. **Accessible:**
being known to all stakeholder groups for whose use they are intended, and providing adequate assistance for those who may face particular barriers to access;

c. **Predictable:**
providing a clear and known procedure with an indicative time frame for each stage, and clarity on the types of process and outcome available and means of monitoring implementation;

d. **Equitable:**
seeking to ensure that aggrieved parties have reasonable access to sources of information, advice and expertise necessary to engage in a grievance process on fair, informed and respectful terms;

e. **Transparent:**
keeping parties to a grievance informed about its progress, and providing sufficient information about the mechanism's performance to build confidence in its effectiveness and meet any public interest at stake;

f. **Rights-compatible:**
ensuring that outcomes and remedies accord with internationally recognized human rights;

g. **A source of continuous learning:**
drawing on relevant measures to identify lessons for improving the mechanism and preventing future grievances and harms;

And, in the case of operational-level grievance mechanisms, they should also be:

h. **Based on engagement and dialogue:**
consulting the stakeholder groups for whose use they are intended on their design and performance, and focusing on dialogue as the means to address and resolve grievances.

These Ruggie Principles should apply also to any informal grievance processes associated with the deployment of NETs.

Given the transnational reach of many NETs, the international community might also consider the development of an international grievance mechanism, perhaps structured along similar principles, that can handle grievances where jurisdictional complexities might otherwise limit the ability of individual rights holders to access a remedy when they feel their rights have been violated. The international community may also consider leveraging existing institutions, mechanisms or channels by expanding their mandates, jurisdiction, and capabilities to handle grievance related to NETs.

Accountability is a broad concept that can also extend beyond adversarial strategies to 'nudge' stakeholders towards greater compliance with human rights principles. While there is an important and valuable role for adversarial accountability mechanisms, some accountability processes should also be designed to promote positive and value-creating constructive engagement opportunities between various stakeholders, outside of the more adversarial judicial processes described above. Creating these processes will incentivize multi-stakeholder (and multi-disciplinary) collaborations, in line with the overall objectives of the HRBA@Tech model. Such processes might be particularly relevant in the context of independent monitoring and oversight processes for purposes of accountability. Investors and donors can encourage technology startups and publicly traded corporations to prioritize human rights by integrating human rights metrics into key performance indicators and contractual obligations. They can also insist on rigorous due diligence and impact assessments. Multi-stakeholder initiatives like the Global Networking Initiative, which requires member companies to undergo periodic reviews by independent analysts to assess the integrity of their due diligence activities, also serve an important accountability function, primarily by relying on the peer pressure of companies to jointly invest in the integrity of a certain 'certification' standard to legitimate their products in an effort to avoid free-riders undermining that standard. Accountability can also be promoted by means of a mix of mandatory or voluntary measures designed to raise the costs of non-compliance with human rights, while also offering tangible benefits for compliance. Such measures might include designing favorfavorable regulatory environments for certain types of pro-social products or technologies, targeted subsidies or tax concessions for 'deserving' tech projects, the prioritization of certain types of products or technologies in government contracts and public procurement processes, and export controls amongst others.

Of course, private and government actors will also want to promote internal accountability, searching especially for any systemic sources of bias that may produce consistently human rights violative outcomes. A failure to conduct such internal accountability audits would result in an unacceptably large risk of continual legal liability. Machine learning systems, for example, are known to sometimes produce biased outcomes, some of which might hypothetically arise because of latent bias among its programmers. In such a case, accountability measures remedying only the (biased) outputs of such a system would be inadequate and unsustainable, since the bias would never truly be corrected for. Such a situation would require an internal forensic analysis to identify why a system produces consistently biased outcomes, a subsequent effort to correct for those root causes, and finally the institutionalization of processes designed to ensure that such biases no longer taint future technological innovations. Such forensic analysis necessitates either internal or external monitoring and oversight mechanisms, periodic reviews or audits by independent experts, participation in multi- stakeholder initiatives, robust grievance processes, and a good-faith commitment to follow-up on recommended remediation strategies.

Finally, it's crucial to always have a responsible party who can address, justify, or explain actions in response to potential complaints. A technology cannot just exist independently – it must always be owned by some legal entity (a real or legal person). If a technology ever becomes completely autonomous (i.e., if it breaks free from its emergency brake and therefore fails the test of safety (see above, p.53), then the last legal entity to have exercised control over that NET would be the addressee and can be potentially deemed liable for any human rights grievances that may ensue from that technology gone rogue.

Processes Associated with Accountability and Access to Remedy		
4A	**Access to the Formal Justice System**	Accountability and access to remedy, at the outset, requires having formal judicial institutions and processes in place. The domestic legal and justice system should be open and accessible to all stakeholders. Civil, criminal and constitutional law in addition to other regulatory frameworks can act as entry points for seeking accountability and access to remedy for any grievances that may arise as a result of the development and deployment of a NET.
4B	**Development of Accessible Non-Judicial Grievance Mechanisms and Processes**	Governments, corporations, and other stakeholders of the international community should work to establish grievance mechanisms and processes structured around the Ruggie Principles (if they are non-judicial) that can be accessed by rights-holders who believe that they have suffered harm due to an NET.
4C	**Monitoring & Oversight**	Government regulatory bodies, civil society actors, individuals and other relevant stakeholders all share the responsibility to engage in monitoring and oversight of those responsible for the development and deployment of NETs and in cases of potential negative human rights impacts and non-compliance with relevant laws, regulations or ethical guidelines– to use all available means to raise the alarm.
4D	**Constructive Problem Solving**	All stakeholders have a responsibility to engage constructively in the search for viable solutions to promote accountability and ensure that individual rights-holders have access to effective remedies

Processes Associated with Accountability and Access to Remedy		
4E	**Incentivization**	Stakeholders must explore options – in addition to grievance mechanisms and procedures – to raise the costs of non-compliance with human rights standards and simultaneously incentivize pro-human rights behavior by those responsible for the development and deployment of NETs.
4F	**Clearly Identified Responsible Entity**	For every NET, there needs to be a clearly-designated responsible actor, who assumes both legal as well as ethical responsibility to ensure that the technology is in full compliance with applicable human rights standards.

II

Make the World a Better Place

As described above, the HRBA@Tech model requires both the commitment to 'Do No Harm' as well as a more forward-leaning commitment to 'Make the World a Better Place.' This second pillar builds on a number of the aforementioned processes, but goes a step further by transforming them into a strategy to actively improve the world, with a particular focus on the most vulnerable in society. It thus draws on the thinking of those who do not consider technology to be neutral, and who thus see a moral and ethical imperative to nudge new and emerging technologies (in their non-neutrality) towards human rights and pro-social values.

5

Human Rights-Based Empowerment

Human rights-based empowerment means equipping rights-holders to better claim and enjoy their rights and equipping duty-bearers to better meet their obligations and responsibilities. As a process, empowerment entails capacity-building of both rights-holders and duty-bearers and also includes identifying and removing barriers that prevent the realization of empowerment. Empowerment in the context of human rights treats rights-holders as active participants in the processes shaping decisions and policies that affect their lives instead of treating them as mere passive objects or recipients of aid or charity.[160]

Further, empowerment requires paying special attention to particularly vulnerable and marginalized groups, and includes other aspects such as access to relevant information, awareness and understanding of rights, meaningful participation and access to technology. In other words, the principle of empowerment is closely intertwined with the participation, accountability and non- discrimination principles of the HRBA@Tech.

States, as primary duty-bearers, have the obligation to respect, protect, and fulfil human rights. This includes not only negative obligations (i.e., to refrain from violating a human right or protect rights-holders from third-party interference), but also positive obligations (i.e., to secure effective enjoyment of human rights). NETs are increasingly being used by States to improve the efficiency of government systems, access and quality of public services, and in turn, the realization of human rights. The pandemic, in particular, has accelerated digitalization across areas such as education and healthcare, which prevented the spread of COVID and enabled the continuation of vital social, economic and other activities despite physical barriers. Digital technologies were central to COVID response around the world and continue to play a key role in post-COVID recovery of societies and economies.

While private companies do not have direct obligations under international human rights law (as opposed to States), they do have a responsibility to respect human rights, which would fall under the HRBA@Tech 'do no harm' pillar. Companies can also, however, choose to actively "make the world a better place" by contributing to the empowerment of individuals and communities through the provision of products and services that directly advance the realization of human rights or provide rights-holders with the means to claim and enjoy those rights. Doing so would mean that the company has made empowerment a part of its business model, and implemented 'human rights by design' i.e., that it has explicitly prioritized the social benefits of technology. Stakeholders such as governments and international organizations can also actively facilitate and promote such socially beneficial technologies, either directly or by regulatory, financial, promotional or other means of indirect support.

Promoting empowerment under the HRBA@Tech model requires ensuring and expanding access to a particular NET and/or associated services to vulnerable and marginalized communities. These communi-

160. OHCHR,"Empowerment,Inclusion,Equality:Acceleratingsustainabledevelopmentwithhumanrights",https://www.ohchr.org/sites/default/files/Documents/Issues/MDGs/Post2015/EIEPamphlet.pdf

ties can use these technologies to more effectively self- effectuate themselves. Prioritizing such empowerment can, of course, be motivated by a charitable impulse, for example in line with a corporate social responsibility policy. But companies can also find ways to promote empowerment while still generating profits (so- called "social entrepreneurship" businesses models). Even though the profits associated with social enterprises might sometimes be more modest than those associated with strictly profit-driven models, social enterprises have the potential to flourish precisely because they tend to be less vulnerable to boom-and-bust business cycles, and also because they tend to enjoy stronger relationships with customers and affected communities. They are also less likely to fall afoul of changing regulatory environments, especially as policy makers begin to focus on the need to encourage more socially- beneficial business models.

Moreover, businesses with credible social entrepreneurship business models are becoming increasingly attractive to investors, thanks in large part to the growing Environmental, Social, and Corporate Governance (ESG) investment agenda. Governments, businesses, civil society and international organizations can promote a range of financial instruments designed to promote innovation by social entrepreneurs. These include social impact bonds, where a municipal authority or private enterprise premises the repayment of investor funds on the achievement of certain predefined human rights indicators, for example the reduction in poverty rates or a drop in prisoner recidivism rates as a result of the introduction of a certain NET. Social entrepreneurs (especially during their startup phase) might also pursue capital from impact investors (perhaps motivated by the ESG investment agenda), sustainability loans, or innovation funds, which earmark capital specifically for certain socially-beneficial business proposals. The availability of such funding options has grown substantially in recent years due to the high demand for green and socially-beneficial investment opportunities, as well as due to the robust financial returns for many such social enterprises. In 2020, for example, it was found that a majority of European-based ESG funds outperformed the wider market.[161] These initiatives can be supplemented by various other financial incentives favoring social entrepreneurship, for example in public procurement processes and government contracts, or through relevant collaborations and partnerships with international development organizations and financial institutions.

Inherent in the process of empowerment, of course, is the central agency of the rights holders themselves. Rights-holders frequently pursue their empowerment by means of community mobilization. While the capacity for such community mobilization activities, resources, and access to key technologies may vary across communities, NETs themselves can often play a key role in that community mobilization process (as evidenced, for example, by the #BlackLivesMatter or #MeToo movements). Entrepreneurs seeking to develop technological tools that will support such community empowerment must understand how communities would use those tools to mobilize in favor of human rights. Digital technology tools centered on crowdsourcing initiatives and processes, for example, can also be powerful facilitators of community mobilization. A prominent example of such a technology tool is Ushahidi (Swahili for "testimony"), which was created by a non-profit technology company based in Kenya that seeks to empower communities through its crowdsourcing platform. This platform is based on open- source software (which means other communities working can adapt Ushahidi's work and customize it based on the mix of issues they are working on). The software utilizes user-generated reports to collate and map data that can subsequently be used by communities and activists to advance social change.[162] Initially developed in 2008 as a tool to monitor and map post-election violence on the basis of crowdsourced reports of election incidents, the technology has since been adopted in a variety of different contexts, including by the UN in order to geolocate victims during earthquakes and other natural disasters.[163]

A second crucial aspect of empowerment is capacity building. As mentioned already, this includes building the capacity of rights-holders (so that they may better claim and enjoy their rights) as well as duty-bearers (so that they may better satisfy the needs and entitlements of rights-holders). Capacity building is (by necessity) a collaborative process. It often requires the strengthening of institutions and having robust mechanisms in place to facilitate such capacity-building processes (for example strong bilateral institutions to manage international development aid flows). In the context of NETs, capacity building activities include, for example efforts to improve digital or medical infrastructure, and efforts to ensure the more widespread access, availability, and affordability of these services. It might also include equipping rights-holders, individuals or groups with the necessary literacy, skills and training to ensure that they can avail themselves of NETs should they choose to do so. Capacity-building also requires education in order to encourage awareness about relevant rights, responsibilities, and obligations. Educational institutions and civil society organizations can play a key role in this regard by encouraging interdisciplinary engagement to promote the language and logic of human rights. Multi-stakeholder collaborations can also serve vital capacity-building functions, particularly partnerships and networks that span across the Global North-South divide.

Closely related to capacity building is technology sharing and transfer. Empowerment also requires that stakeholders (governments, international organizations, and private corporations) make real efforts to bridge the digital divide at various levels: internationally but also amongst different sections of society within a State. Technical know-how about how to use NETs, as well as NETs themselves, should be shared liberally with developing or less developed States and communities. In addition, private enterprises in the Global North should consider forming partnerships with private enterprises in the Global South, protecting intellectual property while also working to spread the reach of socially-beneficial technologies globally. This is particularly true for technologies with an undeniably positive social benefit (for example vaccines). In such cases technologies should be shared with other actors, even if there is no clear 'market case' to do so. National governments and international organizations in particular have an important role to play in the facilitation of such technology sharing arrangements, and can work to expand technological innovations to areas where it might not be otherwise considered commercially feasible, for example in the development of certain vaccines for diseases impacting primarily poor and rural communities in the Global South, such as the Ebola virus, Polio, or Malaria. Open standards, open data and open-source initiatives and approaches for socially beneficial technologies can serve as crucial enablers of such community empowerment.

Processes Associated with Empowerment		
5A	**Human Rights by Design**	Those responsible for the development and deployment of a NET should explicitly consider prioritizing the social benefits of an NET. Public actors should actively facilitate and promote such socially-beneficial technologies, through regulatory, financial, promotional or other forms of support, and private actors work to create business models that are explicitly built around the socially-beneficial aspects of new and emerging technologies.
5B	**Community Mobilization**	Community mobilization is a process where the community takes the lead in an empowerment process designed to identify and address their human rights needs, and this process informs the empowerment strategy. Despite differences in resources and capacities, a community mobilization effort requires the community itself (or the rights-holders themselves) to retain 'ownership' over key decisions regarding the development and deployment of NETs with the potential to impact them.

161. "Majority of ESG funds outperform wider market over 10 years", Financial Times, 13 June 2020, https://www.ft.com/content/733ee- 6ff-446e-4f8b-86b2-19ef42da3824

162. Ushahidi, https://www.ushahidi.com

163. Ushahidi (2012), "Haiti And The Power Of Crowdsourcing", https://www.ushahidi.com/about/blog/haiti-and-the-power-of-crowdsourcing; Pudasaini, Nirab (2016), "Open source and open data's role in Nepal earthquake relief", https://opensource.com/life/16/6/open-source-open-data-nepal-earthquake

Processes Associated with Empowerment		
5C	**Capacity Building**	Capacity building is a two-pronged process: (1) capacity building of rights-holders to better claim and enjoy their rights and; (2) capacity building of duty-bearers to better meet their responsibilities and obligations. In pursuit of these twin objectives, governments and other stakeholders must work to build the capacity of rights-holders to better advance their concerns, needs and rights claims. They must also build the capacity of duty-bearing entities to better respond to the rights-holders, including paying particular attention to vulnerable groups. This includes strengthening institutional frameworks and having robust mechanisms in place to accommodate capacity-building processes. It also includes education and sensitization about rights as well as responsibilities, and the facilitation of various multi-stakeholder collaborations, including between the Global North and the Global South. It also includes overcoming traditional siloes of thought and learning and to encourage interdisciplinary engagement amongst stakeholders, including educational institutions, to promote the language and logic of human rights.
5D	**Technology Sharing & Transfer**	Empowerment requires efforts to bridge the digital divide at various levels. This includes transfer and sharing of socially beneficial technologies as well as technical know-how with developing and less developed nations, civil society organizations, financially weak private enterprises or any other entities with limited resources in order to enable them to harness the benefits of such NET and secure their operations. Where a certain technology has undeniably positive social externalities (for example, a novel vaccine to cure a pandemic) various stakeholders, notably States and international organizations, can collaborate to facilitate technology sharing and transfer efforts and arrangements and to work towards bringing such technological innovations to areas even when there may be no "viable market" for it. The use of open standards, open data, and open-source initiatives for socially beneficial technologies can also be crucial in this regard.

6

Transparency

By their very nature, scientists, technologists, designers, managers, human rights campaigners, policy makers, and lawyers all have their unique way of speaking, interpreting, and reasoning. These patterns of interpretation and expression can be defined as the 'professional cultures' of these various vocations. In its 2021 report, the Human Rights Council Advisory Committee identified the differences between these various professional cultures as a significant barrier to the development of a more holistic human rights based approach to NETs. Transparency is therefore central to the development of a holistic understanding of NETs, and in turn to a proper and comprehensive assessment of the human rights implications of these technologies. Achieving such transparency entails creating access to relevant information, and ensuring that such information is understandable to a variety of audiences. This can pose significant challenges, especially in light of increasingly complex technologies impacting our societies, such as machine learning, genetic engineering, and some of the more complex ICT products discussed in Chapter 2. Greater transparency gives individuals and communities the information they need to seek redress in situations where they feel an NET has harmed their interests. A lack of transparency, therefore, can act as a major barrier to accountability. This is true not just for potential rights holders but also for the institutions that we look to in order to resolve our grievances. In 2020, for example, a Dutch court deciding a challenge to a predictive algorithm tool called "SyRI", which is used by the Dutch authorities to detect welfare fraud, noted that the failure by the defendants to disclose relevant information and the lack of transparency about the algorithm constituted a barrier for the court to effectively understand the workings of a new technology, and therefore a barrier to effectively address the plaintiff's claim.[164] In this way transparency is also closely tied to the accountability principle under the "do-no-harm" pillar of the HRBA@Tech model.

Transparency is not a concession to be made grudgingly in response to pressure or legal obligation imposed by lawmakers or judges. Rather, it is a voluntary and value-additive responsibility that those with a monopoly over information about how an NET works should exercise as a standard part of their efforts to promote and deploy the technology. We therefore refer to this as "proactive" transparency, to distinguish it from the more reactive type of transparency that might result from a court order, a law, or as the outcome of an embarrassing public advocacy campaign (naming and shaming).

Proactive transparency requires the voluntary disclosure of relevant information regarding the development and deployment of NETs.[165] It can also include retroactive disclosures (i.e., in response to being compelled to disclose some pertinent facts about an NET), but the focus is primarily on anticipatory disclosures, i.e., the periodic or regular disclosure of relevant information throughout the different stages of the technology lifecycle. Proactive transparency can pertain to the nature and working of the technology itself, but can also focus on the human decision-making that led to the development and deployment of the technology. This responsibility can encompass various factors, such as the intended purpose of the NET, the existence of less intrusive or disruptive alternatives, the rationale for selecting one technical solution over another, and other technology-specific considerations that may particularly interest technologists and ethicists familiar with that specific NET. In an AI context, for example, proactive transparency may require the disclosure of training datasets that were used to train an AI system, as well as the specifics of any guardrails that may or may not have been built into the system to prevent human rights violations. Proactive transparency can also create potential feedback loops and facilitate constructive stakeholder at different points of the technology lifecycle, which can help find sustainable solutions to particularly vexing human rights considerations.

Proactive transparency is particularly important in cases of a public- private collaboration in the development and deployment of NETs. An example might be a digital technology developed by a private corporation for use by municipal authorities in order to improve public services. In such a situation, proactive transparency would not be limited to transparency between the company and the government "customer" of that technology, but also the community or individuals who might potentially be affected by that novel technology. Thus, transparency would demand that the details by which the contract was awarded are made public, that the public should know about what data is being shared between the corporation and the municipal authorities, what provisions have been made to keep that data secure, whether and how the government can request data held by a private corporation (or the reverse), and the extent of the private actors' ongoing involvement in the operation of the system, among other relevant questions.

The right to information and corresponding freedom of information laws can be useful tools to ensure transparency related to the use of NETs, but usually only in the context of retroactive transparency. The HRBA@Tech model, in contrast, requires stakeholders to assume a more proactive posture towards transparency. To incentivize such an approach, governments can draft laws, policies, and regulatory frameworks designed to encourage proactive transparency, along the lines envisaged by the UNGPs. Such provisions might include disclosure requirements with respect to human rights due diligence processes and obligations to publicize the impact assessments for NETs and corresponding risk mitigation measures put in place to avoid negative human rights outcomes as a result of a newly- developed NET. Such transparency requirements, of course, must be limited by considerations of national security, intellectual property protections, and the right to protect trade secrets. And yet even in such situations, governments can come up with creative institutional solutions to facilitate proactive transparency while also allowing businesses to profit from their innovations. Doing so, far from being a regulatory burden, will inure directly to greater societal trust of NETs, and increase the likelihood that NETs will be biased in favor of human rights.

Open standards, open data and open-source initiatives can also promote transparency.[166] Open standards refer to publicly available standards. Open data

164. Rb. Den Haag - C/09/550982/HA ZA 18/388, https://gdprhub.eu/index.php?title=Rb._Den_Haag_-_C/09/550982/HA_ZA_18/388

165. Naver Agenda Research and Seoul National University AI Policy Initiative, "NAVER, SAPI AI Report" (Nov. xx, 2022). Original Korean material published Nov. 29, 2021.

166. Principles for digital development, "Use Open Standards, Open Data, Open Source, and Open Innovation", https://digitalprinciples.org/principle/use-open-standards-open-data-open-source-and-open-innovation/

refers to information that can be freely accessed with an open license (while securing necessary privacy protections). Open source is software with source code accessible by anyone and is based on the idea of collective ownership.[167] Technologists, academic institutions, civil society organizations and other relevant stakeholders can all join forces to develop socially-beneficial open- source and open-data tools, training sets, and benchmarks (such as impact or bias assessment tools), that can be shared with and used by other actors to evaluate and test their technology tools. Governments can also adopt an "open government" approach by leveraging such open data and open-source initiatives and digital platforms to ensure greater transparency for citizens and access to relevant government information.

A major barrier to meaningful transparency is also the inherent complexity of many NETs, and the almost inevitable "black box aura" that tends to surround many such technologies for most lay people, especially during their early days. Most lay persons, even those who often jump to adopt new technologies when they first emerge, often lack the requisite technical competence, scientific literacy, or awareness to fully understand them. The complexity barrier cannot be overcome merely by adding more disclosure requirements to technology corporations. For transparency to be truly meaningful in such a context it must also be accompanied by intelligibility i.e., ensuring that the information provided is understandable to diverse audiences, irrespective of their technical or legal knowledge, and is accessible also in terms of format or language. At its core, transparency is also about translating technical standards and language into intelligible and comprehensible information that can lend itself to evaluation, also by those without technical insider expertise (for example by affected communities, policy makers, and human rights activists).[168] Translation ensures a better understanding of the logic and functioning of a new and emerging digital technology, which is necessary to effectively and comprehensively scrutinize the human rights impacts of such tools and also establish responsibilities and hold actors accountable.[169]

The process of translation would also, at a minimum, require technologists and other stakeholders to make themselves accessible to questions regarding the technologies they may develop or deploy. While it may be relatively straightforward to demand answers, explanations and justifications from human programmers of traditional digital technology tools, this is much less the case for complex NETs, such as machine learning, where the distance between the human in the loop and the actual decision making process is becoming more and more attenuated.[170] In such cases, the process of translation would also include being accessible to questions and available to provide explanations or justifications when needed regarding the workings and processes associated with a NET.[171]

Technologists, therefore, must prioritize the development of simplified and comprehensible translation strategies that enable meaningful transparency. In so doing, they should collaborate with non-technologists, for example with academics, policy makers, and human rights activists. An example of such an initiative is the creation of simple and standardized documentation describing datasets and algorithmic models that inform an AI system, which provides consumer with relevant information about the technology. These have been known by various names such as model cards,[172] factsheets,[173] datasheets,[174] or even data nutrition labels.[175] Translation in the case of complex algorithms and machine learning tools can also come in the form of providing counterfactual explanations, for example descriptions of what changes in the input data might have resulted in different or desirable outcomes. Such information could work to create greater transparency about the AI without necessarily needing to pierce the 'black box' of the AI system itself.

167. Ibid.

168. Jessica Fjeld et al., "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI," Berkman Klein Center Research Publication 2020-1 (2020)

169. https://www.torontodeclaration.org/declaration-text/english/

170. Atty. Ma "AI Transparency, Governance and Legislative Challenges" (Korean original language), 2 DAIG (2021.9) Credible AI.

171. See generally University of Montreal, "Montreal Declaration for a Responsible Development of Artificial Intelligence," 2018, https://www.montrealdeclaration-responsibleai.com/the-declaration, p. 12.

172. https://arxiv.org/abs/1810.03993; Calderon, Ania et al., "The AI Blindspots cards," Berkman Klein Center and MIT Media Lab, 2019, https://aiblindspot.media.mit.edu/

173. IBM Research, AI Factsheets 360, https://aifs360.mybluemix.net/

174. Cornell University, Arxiv, https://arxiv.org/abs/1803.09010

175. The Data Nutrition Project, https://datanutrition.org/

Google Cloud's AI Explanations, for example, help understand how a machine learning model reaches its conclusions and why it made the decisions it did by providing summaries quantifying each input data factor's contribution to the output decision.[176] Such simplified translations can, ensure meaningful transparency and also build trust in the technology.

To be effective, other stakeholders must collaborate with technologists to create such translations. Consumers of NETs, government regulators, international organizations and standard- setting agencies, individual rights holders and their civil society representatives all have important roles to play in this process. First, they should of course progressively work to increase their own technical fluency with NETs, especially in the case of larger institutions that can afford to hire specialized resource persons. In dialogue with technologists and scientists, these outsiders can also work to promote the intelligibility of NETs to interested laypersons and affected communities. This is directly related to capacity building process described earlier under the empowerment principle of the HRBA@Tech model.

Processes Associated the Proactive Transparency		
6A	**Disclosure**	Transparency requires proactive disclosure of relevant information regarding the development and deployment of an NET, including information about its intended use and any direct or indirect effects of such use. Transparency applies both retroactively as well as proactively, requiring disclosures throughout the various stages of the technology life cycle. Governments should consider appropriate legal and regulatory frameworks for meaningful mandatory disclosure requirements as appropriate. In situations where public disclosures may not be possible (for instance, in case of national security), relevant information may nonetheless be made available to some centralized independent authority or entity, and governments should consider devising appropriate legal or regulatory frameworks for that purpose.
6B	**Translation**	For transparency to be meaningful it must be intelligible to diverse audiences, regardless of their technological and legal knowledge. Technologists must translate relevant information regarding the development and deployment of NETs to ensure it is understandable and also accessible in terms of format and language. At a minimum, the process of translation requires technologists and other stakeholders to make themselves accessible to questions about technologies that they may develop or deploy and devise ways for such simplified translations. Stakeholders, by themselves or through collaboration with other stakeholders, can also engage in capacity building to improve technical literacy and skills amongst various actors to ensure intelligibility and meaningful transparency.

7 Participation

The HRBA@Tech model requires the owners, developers and deployers of NETs to solicit the active participation of those stakeholders who may be directly or indirectly affected by a NET and encourage their direct participation in crucial decisions about that technology. All people have the right to participate in any decision- making process affecting their human rights, and a human rights-based approach requires that such participation must be active, free and meaningful.[177] At the outset of a technology development process, participation can take the form of consultations with relevant rights- holders and stakeholders. As it may be virtually impossible to consult with all possible stakeholders in such a process, consultations should focus

176. Frey, Tracy (2019), "Google Cloud: Increasing transparency with Google Cloud Explainable AI", https://cloud.google.com/blog/products/ai-machine-learning/google-cloud-ai-explanations-to-increase-fairness-responsibility-and-trust

177. Care about rights, "What is a human rights based approach?", https://careaboutrights.scottishhumanrights.com/whatisahumanri- ghtsbasedapproach.html

on the involvement of bona fide representatives and advocates of the relevant stakeholder groups, and stakeholders must make efforts to ensure that such representation reflects the true interests of the groups or communities being represented. For such participation and consultation processes to be meaningful, they must be structured in a way that they effectively facilitate bottom-up claims and adopt a collaborative approach to identifying and assessing needs. Participation is also truly meaningful when the inputs received in the process are meaningfully acted upon. Therefore, meaningful participation ensures for relevant rights-holders and stakeholders an active role and co-ownership over the processes that affect them. Participation, which in this sense is also directly related with the empowerment process, is crucial to achieving all the other principles of the HRBA@Tech.

Fostering participation necessitates recognizing that, if left unchecked, the power dynamics between various stakeholders in a typical technological development process are likely to be inherently unequal. In order to foster true participation, therefore, the balance of power must be actively pushed in favor of those stakeholders that would otherwise be less empowered absent a HRBA@Tech model approach.[178] To promote participation, duty-bearers or those with responsibilities towards the promotion of human rights (i.e., governments and private companies), must proactively seek out multi-stakeholder participation, both from external stakeholders (e.g., rights holders, communities) and also internally within the duty- bearing entity itself. We therefore again describe this as a "proactive" principle, not a reactive one. Proactive Participation requires that a range of stakeholders, including vulnerable and marginalized individuals and communities, be brought into decision making and consultation processes with a real potential to impact how an NET will be developed and deployed, not as a result of having been forced to do so but rather as a general ethical imperative.

Participation under the HRBA@Tech requires engagement with a range of stakeholders (for example the public sector, the private sector, civil society, academics, technical experts, industry interest groups, and international organizations and standard setting agencies). Moreover, this participation should span the entire technology life cycle, not just the most high-visibility points along that cycle (for example the moment when a technology company wishes to 'go public' and become a publicly traded company). Earlier, in our discussion of non-discrimination and equality (under the 'do-no- harm' pillar of the HRBA@Tech Model), we discussed the importance of harnessing the power of internal diversity in order to avoid bias. Proactive participation supplements such internal diversity, allowing for the fact that no organization, no matter how diverse its staffing patterns, can ever bring 'in house' the full range of stakeholder perspectives. It is particularly important in this context to seek the participation of vulnerable and marginalized communities in order to integrate cultural sensitivities and awareness into products and services, a process which has sometimes been referred to as 'diversity by design.'

Participation is crucial during the early stages of a NET's design phase (for example during an initial human rights due diligence audit or impact assessment). It is also important during later stages of a technology's lifecycle, where it can take the form of independent or third-party audits, internal or external reviews, and other forms of accountability mechanisms. All of these processes provide avenues for the genuine participation of relevant stakeholders and communities that stand to be affected by an NET. Consultants, auditors, regulators, lawyers, and specialists tasked with conducting such reviews should make conscious efforts to incorporate participatory elements and processes throughout the NET lifecycle.

States should also encourage participation in the formation of their own regulatory approaches to NETs. Many states already rely on participatory process to craft policies, for example in environmental law. Applied to the HRBA@Tech model, participation requires public notifications of proposed new laws or policies pertaining to NETs, proactive solicitation of feedback on these policies from affected stakeholder groups, surveys, studies and interviews, consensus building, and other such related activities. In recent years, so called "regulatory sandboxes" (in which regulatory authorities partner with industries to co-craft meaningful regulatory frameworks that do not stymie the dynamism or creativity associated with NETs) have also been used to promote the participation of industry stakeholders in efforts to craft meaningful and regulatory frameworks.

Meaningful participation almost always requires expending significant efforts to overcome hurdles that otherwise prevent or chill effective participation. These include both "1.0 barriers" (such as language barriers, distrust of outsiders in many traditional communities, power differentials, and resentment about past instances of exploitation) as well as "2.0 barriers," mostly having to do with a lack of access to modern communication technologies or technical expertise. Overcoming 1.0 barriers requires a great deal of proactive effort, and dovetails with the strategies that have been honed by human rights and development experts over decades of hard-won experience. Here one things of "old-school" community engagement strategies, usually based on skilled experts actually consulting in person with affected communities. Overcoming the more modern 2.0 barriers tracks closely to our discussion of capacity building efforts associated with the principle of empowerment (see above). States, international organizations, civil society, educational institutions, and also the private sector can all join forces in efforts to overcome both 1.0 as well as 2.0 barriers to participation.

The private sector can also make efforts to engage with local stakeholders and communities that stand to be affected by their operations. Companies should do so from both a risk-management perspective (in order to avoid potentially costly or embarrassing lawsuits or public protests), but also in line with a commitment to "make the world a better place." This is particularly true in the case of tech companies behind transnational operations. Such companies have a particular obligation to ensure participation, even if that requires substantial efforts to engage in local languages or ensure that information about a NET is made available in formats that are understandable in the local context.

Achieving participation requires the formation of trust-based and collaborative partnerships across traditional stakeholder divides. Such multi-stakeholder (and multi-disciplinary) collaborations and partnerships, focused on dialogue, ensure that a plurality of perspectives are considered with regard to the development and deployment of NETs, and is vital for the HRBA@Tech Model to be truly holistic.

Processes Associated with Participation		
		Participation requires consultations with rights-holders who may be potentially or actually impacted by an NET as well as other relevant stakeholders. Participatory consultation efforts can be initiated by those responsible for the development and deployment of NETs or the affected communities themselves. Regardless of who initiates the consultation process, they should be structured to facilitate bottom-up needs assessments. Furthermore, all who are involved in such consultation efforts should remain open to collaborative problem-solving efforts. While such consultation processes must be open to traditionally vulnerable populations and their representatives, it can also include other individuals or groups based on local contexts and conditions. Participation should also include other forms of collaboration such as partnerships by public and private actors with educational institutions for participatory research and industry-wide or multi-stakeholder initiatives.
7A	**Consultation and Collaborative Needs Assessment**	
7B	**Representative Advocacy**	Since it is usually impossible to involve all potential stakeholders in such participatory processes, usually there must be representatives and advocates who speak on behalf of affected stakeholder groups. Such representation can be ensured through the presence of advocates engaged in a participatory consultation effort (for example civil society representatives). It can also be ensured by increasing the diversity of the teams developing and deploying NETs within technology companies or scientific research teams. Regardless, all stakeholders have an ethical (and possibly also a legal or fiduciary) obligation to ensure that any such representatives remain accountable to the groups they represent and not the entities sponsoring or convening the consultation effort.
7C	**Multi-Stakeholder Dialogue**	Multi-stakeholder dialogues are vital tools to ensure the holistic nature of the HRBA@Tech model. They must be designed to ensure the representation of a diverse set of perspectives in decisions pertaining to the development and deployment of NETs.

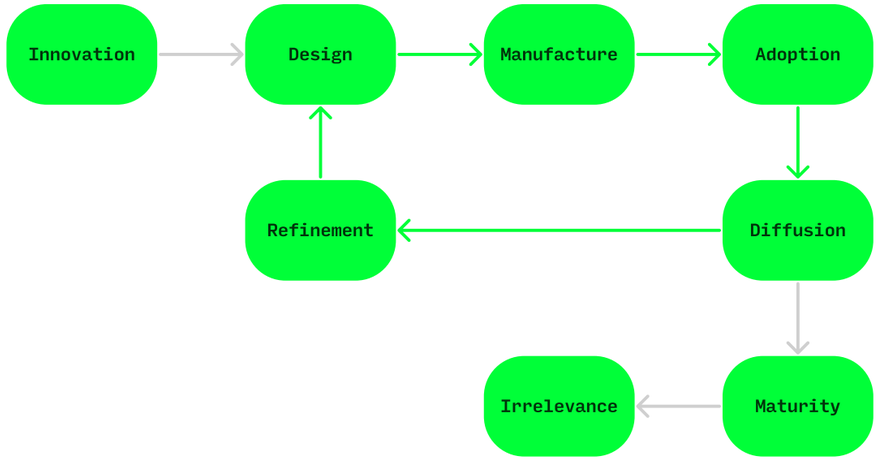
178. Falkenburg, Naomi (2021), "An introduction to participatory monitoring and evaluation: the missing link between inquiry and impact ", https://www.activityinfo.org/blog/posts/2021-03-15-an-introduction-to-participatory-monitoring-and-evaluation-the-missing-link-between-inquiry-and-impact.html

CHAPTER 4: "THE HOW" OF THE HRBA@TECH MODEL

Chapter Summary:

This Chapter analyzes "How" the HRBA can be applied throughout the lifecycle of an NET. Speaking in abstractions, we describe the lifecycle of an NET as being composed of eight phases: (1) innovation, (2) design, (3) manufacture, (4) adoption, (5) diffusion, (6) refinement, (7) maturity, and (8) irrelevance. The discussion in this Chapter highlights how the key principles and processes of the HRBA@Tech model can be actualized in the process of developing and deploying new and emerging technologies.

Lifecycle for New & Emerging Technologies: From Innovation to Product Maturity



In analyzing the 'How' of deploying our HRBA@Tech model across the entire lifecycle of an NET, we consider that there are eight phases to any NET lifecycle: (1) innovation, (2) design, (3) manufacture, (4) adoption or implementation, (5) diffusion or deployment, (6) refinement, (7) maturity, and (8) irrelevance. This model draws on traditional Technology Life Cycle (TLC) models, but adopts them from the perspective of the HRBA@Tech model, where the primary objective is to identify meaningful intervention points from which to 'nudge a technology into the direction of being more prone to support the realization of human rights.'

Not every technology passes through all eight phases of the TLC. Indeed, some (the majority, perhaps) go directly from innovation to irrelevance, either because they fail to attract the necessary start- up capital, or simply because they fail to convince those who would have to adopt the technology for it to become viable. Perhaps a technology is simply before its time, or – perhaps – it is simply not a very persuasive technology. Other technologies flourish, get adopted by numerous consumers, and then suddenly die off, either as a result of a changed market or simply because the initial psychological 'hype' that may have driven its take-off simply faded away again, for reasons that often have as much to do with the advent of a new technology as with the insatiable need of many to pursue the 'next big thing.' Finally, some technologies flourish but then they also evolve and mature into new cycles of innovation and technological development, perhaps fueled by the initial success and the re-investment prowess of an innovator who understands that no technology – no matter how innovative– will last forever.

In this oversimplified model, the technology lifecycle typically begins with a flash of genius, an insight, or an innovation. This may, of course, be an oversimplification: in reality many such insights derive from or build on previous technologies (a process we describe as 'refinement' in the sixth phase of this model). Nonetheless there may be some technologies that literally start from zero. From this innovation phase, so-called 'founders' (the innovators, technologists, and entrepreneurs who originally own the idea) must first bring that product across the 'proof-of-concept' threshold. To do so typically requires a great deal of ingenuity, research, and marketing, which in turn requires resources. Hence, this period also usually requires a great deal of investment, since most new technologies are hardly economically viable. The classical TLC model describes this as the Research and Development Phase (R&D), but our model breaks it into three successive phases: (1) innovation, (2) design, and (3) manufacture. Particularly when it comes to the innovation and design phase of an NET, it matters a great deal if the founders' motivation is to make profits (a market-driven motive) or to make the world a better place (a socially-beneficial orientation). Depending on that baseline intention, different principles and different processes may apply to the development and deployment of a particular NET.

At this point our model reconnects with the classic S-curve model of the TLC, in that if and when a tech-

nology crosses the proof-of-concept threshold, it may enter the next phase of its lifecycle, which would be that of (4) adoption, where the demand for a technology skyrockets. Protected either by a temporary technological edge or by intellectual property (IP) protections (or both) the founders of profit- driven technologies generally tend to profit handsomely during this phase, and much of those profits go to grow the market. Founders of socially-beneficial technologies, on the other hand, begin to see the desired impacts of their NET during this adoption phase, when increasing numbers of users take up the technology. Eventually, however, competitors typically begin to offer credible alternatives to an NET, at which point the founders might need to focus either on diffusing the original technology into new areas of application or reducing the costs associated with bringing that original technology to new markets. We are referring to this phase as the (5) diffusion phase, since this is where some technology owners begin to explore joint ventures with other entities, expand the scope of what a technology can achieve, or focus on reducing the costs of the process used to generate that technology rather than the specifications of the technology itself.

At this point, our model (unlike many other TLC models) posits a fork in the cycle. Some founders are able to use the momentum, reputation, and potential profits of their earlier success to initiate a phase of (6) refinement, during which the company reflects on the presumed ageing and decline of its initial technology and identifies potential avenues for rejuvenated growth, perhaps premised on another new technological innovation. If successful, this refinement process then leads to a renewed TLC, leading back into the phases of (re)design, manufacture and so on.

Absent such a refinement phase, even the most exciting technologies tend eventually to (7) mature, at which point the technology returns to the initial innovation investment begin to flatten and taper, eventually leading to abandonment and finally (8) the complete irrelevance of a former technology.

The discussion in this Chapter highlights how the key principles and their associated processes in the HRBA@Tech model can be operationalized across each of these eight phases of the TLC.

Innovation

Every new technology begins with innovation - the sparks of scientific insight that come together after years of scientific research, the flash of ingenuity that strikes an entrepreneur, or the fruits of intense teamwork. Without innovation there can be no technological progress.

Innovation is a fundamentally creative process. It relies on a determination that change can (and perhaps should) happen, and a self-assuredness that the inventor (or the technologist, entrepreneur, visionary, etc.) can solve a confounding problem. This creativity can rarely be forced from above. There are, or course, examples of the State driving technological innovation, for example the State-run 'Manhattan Project' to develop the atom bomb during World War II, the more recent 'Project Lightspeed' to develop an effective vaccine against the COVID-19 virus, and state-owned grant-giving institutions that exist in many countries to fund academic and scientific innovation. But those efforts the exception rather than the rule. A closer look at many of those large-scale state-driven efforts also reveals that they relied on a permissive approach to innovation, recognizing that success often depends on throwing exploring numerous ideas at once and seeing which ones ultimately bear fruit. This approach embraces non-conformist thinking as a matter of necessity, letting even 'crazy' ideas benefit from funding and support. For every vaccine created with the support of taxpayer funds, there were many others that never came to fruition. Wars, pandemics, and other crisis situations can legitimate such an approach, but this logic typically falls apart if the goal is merely to spur "mundane" scientific progress.

Most advances in civilian technology is not driven by the government alone. Much more common is the approach of letting private innovators shoulder the risk of failure, perhaps with some facilitation by government grant-giving agencies. This model incentivizes innovation by offering entrepreneurs and innovators the prospect of substantial rewards should their ideas yield fruit. Today's innovation is increasingly taking place in 'innovation hotspots'; usually densely populated metropolitan areas with dense and self-reinforcing concentrations of innovators. Innovation hotspots usually depend on being open and interconnected with the rest of the world, especially other innovation hotspots. Hence the direct flights between Hyderabad in India and San Francisco in the Silicon Valley. Collaboration and professionalization are core the process of innovation, and these hotspots evolve to facilitate such synergies.[179]

WIPO reports that patents for new inventions and scientific papers are increasingly being co-authored by teams of international collaborators, stating that '[b]y 2017, lone wolf scientists had become half as important as they were 20 years before.'[180] The stereotyped image we may have in our minds of unkempt tinkerers, college dropouts, and non-conformist visionaries surviving on a diet of sugary drinks and cold pizza to create the next market-dominating technology is quickly becoming the stuff of legend and not reality. For the technology sector to sustain the same pace of innovation that has characterized the industry since the 1980s, it now needs to hire teams of researchers multiple times as large as would have been the case in previous decades.[181] At the center of these networks are 'skilled individuals and innovative companies,'[182] buttressed by supportive policy environments and modern infrastructures, such as telecommunications networks and more traditional transportation infrastructures, that allow for easy access to global markets. They typically revolve around high-quality educational institutions churning out qualified workforces, and usually promise a decently high quality of life and open visa regimes to appeal to global workforces seeking jobs in the tech sector. This is not to say that 'lone wolf' entrepreneurs no longer exist. They do, but they too are now increasingly drawn to some of those same innovation hotspots where they can benefit from a critical mass of services that can help them realize their vision.

These innovations centers typically also draw a very significant portion of available investment. Venture capitalists (VCs) concentrate the investing power of high-value investors and organizations into long-term support for entrepreneurs with a vision, usually on terms that would not be available from standard banks. VCs commit themselves to entrepreneurs ('founders') during the earliest stages of their path—when their ideas are still untested—and help them bring their ideas across the 'proof of concept' threshold. Individual founders benefit from the intellectual diversity and synergies that come about whenever a community of entrepreneurs conglomerate in one spot, thus, driving an ecosystem of innovation and invention. Regardless of how dynamic and innovation-oriented such a hotspot may be, success is far from guaranteed for individual founders (or their investors). Overall, even after receiving grant funding, survival rates can be as low as eight per cent.[183] Done properly, VC investing requires a strong relationship between founders and investors, and a willingness by both to engage in open dialogue about the proper balance between management considerations and technological innovation.[184] The nature of this challenge requires VCs to carefully select the founders they work with, concentrating not only on the founders' ideas and business plans, but also the kind of person they are working with.[185] VCs tend to spread their investments judiciously, prioritizing quality over quantity. This same approach is also true for those VCs operating outside of the 'innovation hotspots' mentioned above.[186]

179. World Intellectual Property Organization (2019). World Intellectual Property Report 2019: The geography of innovation: Local hots- pots, global networks. Geneva: World Intellectual Property Organization, 113-120

180. Ibid., at 41

181. One study found that for innovation to keep pace with what is now known as "Moore's Law" that microchips will continue to double in capacity every two years, 18 times the researchers are now required to sustain that pace of innovation than would have been the case in the 1970s. Nicholas Bloom, Charles Jones, John Van Reenen, and Michael Webb (2020). Are Ideas Getting Harder to Find? 110:4) American Economic Review 1104-1144

182. Ibid., at 9

183. https://corporatefinanceinstitute.com/resources/valuation/how-vcs-look-at-startups-and-founders/

184. Anonymous interviews with VC Investors / Senior Partners, (Nov. 3, 2022) Seoul, RoK

185. Ibid.

186. Racine, Jean-Louis (2017) "To foster innovation, let a hundred blossoms bloom?" (May 15) World Bank Blogs: https://blogs.worldbank.org/psd/foster-innovation-let-hundred-flowers-bloom.

Intervention Points:

Given the daunting nature of what it takes to bring scientific insight from the moment of innovation to 'proof of concept,' one might ask how reasonable it is for the human rights community to demand that human rights considerations be addressed already at this earliest phase of an NET's lifecycle. The answer depends on which of five scenarios apply to the particular innovation process. To determine which scenario applies, one must ask two fundamental questions.

1. Who bears the risk of a particular technological innovation? The government, a well-financed corporation, or individual founders?
2. What kind of technology is it?

Category 1: All government-sponsored technological innovations	
Who bears the risk?	A government.
What kind of technology is it?	Any technology.
Corresponding obligations:	Standard institutional human rights safeguards that apply also to all other government activities

In some situations, a government sponsors or serves as the primary driver of tech innovations (for example, when a government contracts the development of a particular technological product to a private company). Governments can and frequently do turn to new and emerging technologies as part of their operations. Sometimes they do so to improve citizen-facing government services. At other times, governments deploy new and emerging technologies in order to protect national security or keep the peace, for example when government agencies set up surveillance systems around sensitive infrastructure.

Regardless of the circumstances, governments – as the ultimate guarantors of human rights – must always be guided in whatever they do by human rights. Whenever a government sponsors a new and emerging technology, it is subject to rigorous human rights safeguards, with a strong presumption in favor of protecting human rights except in narrowly defined 'derogable' situations justified by public safety or national security.

The same logic also applies to non-profit organizations and academic institutions, both of which can also be reasonably expected to work towards the public good. Individual scientists, academics or students in such institutions are innovating in large part to further their personal career advancement process, incentivized to do so either because it is their job to do so, or (in a university environment) driven by brighter career prospects upon graduation or the promise of professional accolades as a long-term academic. In such efforts, technological developments should typically not be driven by the need to maximize profits, but rather by the desire to maximize socially beneficial outcomes. The rigorous and mandatory 'human subject review' processes familiar to any academic researcher exemplify this idea.

In other situations, private entrepreneurs or 'founders' launch a new technology. Frequently, these founders are individuals working in groups or alone. One might imagine a group of recently graduated students deciding to create a startup based on an innovative idea. In other situations, a major corporate entity sponsors an entrepreneurial push to develop a new technology, for example when a major existing company creates a new sub-unit (or acquires a startup) to promote a particular NET. In that second situation, the creative startup may still be quite small, but it will enjoy the financial, logistical, and intellectual support of the sponsoring corporation.

Category 2: Any private effort to develop a product that will "make the world a better place"	
Who bears the risk?	Any private entity (regardless of size)
What kind of technology is it?	A technology specifically intended to 'make the world a better place'
Corresponding obligations:	Careful vetting by investors and/or corporate or government sponsors of proposed business plans to ensure that they clearly articulate how the NET will work to 'make the world a better place'

The second category pertains to any private effort to develop a technology that will 'make the world a better place'. For this category of products, it is irrelevant whether the risk attaches to a group of individual founders or a major corporation. For such technologies, all innovators must demonstrate during the innovation phase how their innovation will contribute to the protection and promotion of human rights or increased welfare. If the entire 'business case' justifying a NET is that it will serve to improve social welfare in some way, founders will want to make that logic clear at the earliest possible moment. This is not only an ethical statement, but also a deeply practical one: entrepreneurs need support for their innovations—especially if they are not driven exclusively by a profit motive—and only the most convincing technological solutions will receive the support they need from the kinds of charitable donors and foundations who might typically be interested in supporting such ideas. Any innovators pursuing such technologies would be well-advised to present their case convincingly at the earliest possible moment. Human rights considerations should be central to the innovation phase for such technologies, by for example, seeking inputs from the intended beneficiaries either through direct engagement with members of the target population or indirectly through their representatives such as civil society or any other relevant stakeholder.

The next two scenarios are ones where a technological innovation is intended only to be commercially profitable. There is nothing illegal or ethically suspect with a strictly profit-motivated technological innovation. Indeed, many of the modern-day technological conveniences we use, including the machines we are likely using to read or write this document, were driven by corporations pursuing a profit motive. Nonetheless, the HRBA@Tech model requires even of profit-driven entrepreneurs that they think about human rights. The extent of what is reasonable to expect of those innovators, however, depends on the capacity of the private entity that ultimately bears the risk of an innovation.

Category 3: A private effort by individual innovators or an independent startup to develop a profitable technological innovation	
Who bears the risk?	Private founders (independent startups)
What kind of technology is it?	A technology intended primarily to generate profits
Corresponding obligations:	Voluntary risk management brainstorming sessions as part of the investment cycle (risk mitigation as the objective)

Some innovators bear the risk themselves, or perhaps divided amongst a limited number of co-founders. For those innovators, human rights considerations are likely to be incentivized primarily as part of an overall risk-minimization strategy. The urge to reduce risks might be inherent to a startups survival strategy, as any negative press associated with a NET might be enough to kill off even the most promising of startups in a hyper-competitive market. Profit-conscious investors are likely to reinforce such risk minimization strategies, especially in markets that are sensitive to strong ethical preferences by consumers.

Category 4: A private effort by innovators or startups backed by a well-resourced corporation to develop a profitable technological innovation	
Who bears the risk?	Well-resourced corporate sponsors
What kind of technology is it?	A technology intended primarily to generate profits
Corresponding obligations:	Due diligence and mandatory risk management brainstorming sessions as part of the innovation cycle (in compliance with ESG-compliant business practices)

Other innovators are backed by major corporations. For these startup innovators, they benefit from the dynamism and creativity of a startup while also enjoying the resources and security that comes with having a major corporate backer. With these added resources and greater sense of organizational capacity comes a heightened responsibility to take the potential human rights impacts of an innovative new technology seriously. Here too, the market should drive this focus. The basic logic of risk management still holds, as it did for the group of individually liable founders, except this time the risk extends beyond the founder(s) themselves and extends also to the supporting corporate sponsor. Thus, not as a matter of charity or ethics, but rather out of self-preservation, the corporate sponsoring entity would want to ensure that all possible human rights implications of a new technology be explored and mitigated to the maximum extent possible.

The distinction between categories 3 and 4 is justified by the substantially reduced risk that individual entrepreneurs or innovators face when they are sponsored by a well-resourced corporation. The risk facing individual technologists, scientists or entrepreneurs is often minimal, since typically they might be pursuing their work subject to a regular employment contract. Those employment contracts shift liability for any technological innovations towards the sponsoring corporation. On the one hand, this means that the risk of any one scientific investment failing can be absorbed within a larger portfolio of similar investments. On the other hand, however, any one misguided or unethical investment decision can also tarnish the reputation of the sponsoring entity as a whole. Given the greater resources of established entities, it is also reasonable for society to have higher expectations of their willingness to spend resources to safeguard and promote human rights as part of their ongoing development of NETs.

Innovators and startups that originally fall into category 3 may gradually evolve to more closely fit into category for, such as when a startup grows rapidly and eventually has the resources (and risk sensitivity) to match the capacity of a well-established corporation to ensure the absolute safety and reliability of a NET.

Category 5: A private effort to a technological innovation meant (potentially) to harm people	
Who bears the risk?	[Irrelevant question]
What kind of technology is it?	A technology intended (potentially) to harm people
Corresponding obligations:	Due diligence of government sponsors and potential use of a NET

Finally, there are some technologies that are in fact created with the intent of harming people, or at least technologies that have a clear and high risk of consistently doing so. This may sound terrible, but in fact many such technologies exist. A weapons producer may, for example, develop a less-than-lethal weapon designed for crowd control, such as a taser. While tasers do, in fact, harm people, that does not mean that no private innovator should ever invent such a technology, nor does it necessarily mean that the technology cannot be used to improve human rights (imagine, for example, a police department being forced to choose between a less-than-lethal or lethal crowd-control strategy). Thus, private innovators, sometimes bearing the risk themselves, and sometimes backed by corporate sponsors can – and frequently do – produce such technological innovations.

Such innovations can be justified only if done in service of a government sponsor or a publicly delegated entity that can be trusted with the use of such technologies. To return to our earlier example: if an entrepreneur produces tasers for a police department known for its scrupulous adherence to human rights policies, they can confidently produce the technology for the benefit of that client. Once they developed the technology (or transferred its control) to that government entity, the innovators can rest assured that the liability for the use of that technology will rest exclusively with the government. If the maker of the taser technology sells the product to a police department known for its strict adherence to human rights, and then later a rogue police officer unexpectedly decides to misuse that taser to violate the human rights of arrested individuals, the producer of that taser cannot be held liable for that misuse. The liability for that misuse would remain with the police department.

Drawing sharp distinctions between technologies that can be used by the military and civilian-use technologies is simply not possible. Technologies initially devised for civilian use may eventually trickle to military use and vice versa. Even today, many scientists and technologists receive irreplaceable funding and support from governments. Governments provide such funding as part of their efforts to secure their national security, attracting stakeholders from all sectors of the economy, including scientists and technologists who might not otherwise consider themselves to be members of the national security sector. In the United States, for example, the Defense Advanced Research Projects Agency (DARPA) reports that as a result of a 'half century' of investments into efforts to develop artificial intelligence, the US military has developed 'automated or machine-assisted surveillance and text understanding' capabilities.[187] The same publication also points out, however, that those same innovations are currently 'being operationalized by all major information technology companies (Google, Microsoft, Facebook, etc.) as the basis for new services and efficiencies, contributing significantly to their work with image recognition, speech translation, robotics, and facial recognition technologies.'[188] Google, Microsoft, and Facebook, as well as the individual scientists and technologists who later went on to develop those civilian-use AI technologies, are not typically thought to be part of the US national security apparatus, and yet their innovations are still inextricably intertwined with the State's national security initiatives.

Key in such discussions about the innovation of so-called 'dual use' technologies is whether the technologist knows or should have known that the technologies they are developing are being specifically to commit human rights violations (as distinct from merely 'normal' defensive or military applications for such technology). Under the HRBA@Tech model, individuals and corporations in such situations have a duty to decide for themselves if the government entities they are working to support respect human rights. This entails a basic level of due diligence, even at the earliest innovation phase. If an individual or corporate technology innovator 'know[s] that its actions will substantially assist the perpetrator in the commission of a crime or tort in violation of the law of nations,'[189] they can be found guilty of aiding and abetting in the violation of human rights.

If, to revert to our previous example, the manufacturer of the taser had reason to believe that the police department in question was interested in purchasing the technology specifically to violate human rights, for example if the procurement officer from the police department had specifically inquired with the manufacturer whether a taser could be used to torture detainees without leaving any visible marks, and if the police department were known nationally as a particularly vicious abuser of human rights, the company would have an obligation to refrain from selling that product to that police department. Any innovator who knowingly transfers the technology anyway to such a notoriously rights-abusing government entity, would not be able to shield themselves from liability for its eventual misuse. This doctrine is derived from human rights law, where corporations are required to conduct basic due diligence to ensure that they are not inadvertently facilitating the perpetration of human rights abuses by a government sponsor. The bar for such aiding and abetting is fairly stringent, and would require showing not only that the technology is ultimately used in the commission of human rights violations, but rather that the individual

187. DARPA (date unknown), Deep Learning (accessed Nov. 5, 2022), https://www.darpa.mil/attachments/DEEPLEARNING_Layout_Final. pdf

188. Ibid.

189. Herdegen, Matthias, (2013), "Principles of International Economic Law", Oxford University Press, p. 132

innovator knew (or should have reasonably known) that the State sponsor would use it in such a way.

These categories are not mutually exclusive. Firms in the market are driven by profit making, and thus there may be instances where an innovator professes fidelity to the goal of making the world a better place while also pursuing a profit. Our core contention is that the logic of these categories can be derived from a combination of a risk-minimizing corporate strategy, coupled with universally applicable human rights norms. The key point of this categorization exercise is that not all tech entrepreneurs can reasonably be held to the same standards. The applicable standards depend on where along a technology's life-cycle a technology may be, and correspondingly how much capacity a given startup to focus on human rights.

From the perspective of lone wolf technologists and entrepreneurs (as well as the investors who support them), the innovation phase is a very fragile stage of a NET's lifecycle. Paramount during this phase for such 'founders' is the challenge, against great odds, of bringing an idea to viability. While human rights considerations in such situations do not fall away entirely, it can be much less reasonable to expect founders to implement the full slate of human rights processes proposed in the HRBA@Tech model. Nonetheless, simple brainstorming processes around potential negative human rights impacts, designed to minimize the risk of a potentially disastrous public relations crisis, can make a tremendous difference in the ultimate viability of a technological innovation.

The following example, taken from the public history of one very prominent tech company, illustrates the difference between a private innovator in Category 3 as opposed to one (a decade later) that at that point falls clearly under Category 4:

Case Study:

Meta (Facebook) is currently in the midst of a major corporate transition, renaming itself and reorienting its focus towards the Metaverse (an online social space designed to be more immersive than previous social media platforms). The company hopes this investment will ultimately pay major dividends, and by 2021 invested over 10 billion USD to develop the Metaverse. This period of 'innovation' could hardly be more starkly in contrast with the development of Facebook itself, which was started in 2004 by five college students working from their dormitories.

Even assuming that those five college students in 2004 might have dreamed that their invention might one day be used by 2.91 billion people, it would have been unreasonable to ask those initial innovators to conduct a full human rights impact assessment, consulting with (for example) millions of people from around the world to survey their needs and how social media might help support them. At most, one might have asked them how (for example) they might design a social media platform that would not obviously discriminate against or objectify its female users, or fall afoul of basic copyright or privacy protection laws.

Indeed, had Mark Zuckerberg and his fellow classmates had such conversations, they might have decided against an early version of Facebook based on a website called 'Hot or Not,' in which Harvard students were able to rank two randomly paired profile photos of their classmates against one another, generating cumulative 'scores' of who was most attractive in the class. This website, which was live for only a few hours but nonetheless caused major outrage among his classmates, also led to disciplinary proceedings that could have resulted in Mark Zuckerberg's expulsion from college.

Such early risk management conversations, however cursory, might also have dissuaded Facebook's founders from using data gathered by the website to hack into the private email accounts of journalists who had written critically about a business dispute Mark Zuckerberg had been involved in.

Such basic conversations, focusing on very predictable human rights vulnerabilities and privacy safeguards, could have prevented Facebook from committing those early blunders, and might also have set Facebook, as a whole, on a more socially-responsible trajectory in later decades, when it had grown into a global site linking billions of users.

Now, in 2022, Facebook (Meta) has grown to employ almost 78,000 employees. With Meta again investing in a new innovation period (this time into the Metaverse), the expectations of what Meta can reasonably do to ensure that its investments are in line with human rights standards are vastly greater than those that may have applied to Mark Zuckerberg and his college classmates in 2004. Today, Meta is leading discussions among civil society activists, academics, technologists, and policy makers around the world, asking them what they believe to be the biggest threats to Meta's investment plans and soliciting their input on how best to address those vulnerabilities. Meta presumably has small armies of lawyers checking compliance with every conceivable regulatory and legal standard that may apply to this NET, as well as specialized staff spanning the globe who can research local standards and interact with local stakeholders.

This example serves as a vivid illustration of how the size and capacity of the entity bearing the risk of a certain innovation matters in terms of how much can be expected of it within the HRBA@Tech model during the innovation phase.

Design

During the design phase, technologists and entrepreneurs need to refine and bring the idea of a technology into fruition, such that it can realistically meet its objectives within the real-world environment where that technology is meant to operate (inter alia, the market, the natural environment, the human body, the sum-total of publicly available data on the internet). The design process seeks to maximize the chances that a product built on an NET will have sufficient appeal to the target audience such that they would be motivated to adopt it (in the case of process-based technological innovations) or buy it (in the case of new products built on an NET).

Intervention Points:

The design phase is one of the most important intervention points for injecting human rights considerations into an NET. At this point, after the kernel of a new technology has been developed but before the final form that a new technology will take has been ossified by virtue of a manufacturing process, human rights considerations can still be mainstreamed with relative ease into a new technology. The integration of human rights considerations directly into the design phase should happen for all types of technology (categories 1-5 in the chart above), regardless of the objective of a technology and regardless of who bears the ultimate risk for the development of that innovation. At a minimum, this design process should be focused on ensuring that the NET lives up to the 'do no harm' principle. As much as possible, designers should carefully consider how issues of security (see above, p.53) and non-discrimination and equality (see above, p.51) are mainstreamed directly into the design process.

Designers can do so by means of a futures thinking methodology, whereby they posit as a thought experiment that their technology has met all of its aspirational goals in the distant future (a 'what if' statement), and then reverse engineer the hypothetical progression of that technology from innovation to maturity (writing the 'history of the future').[190] Doing so enables the futurist designers to also posit what else would have to be true for a technology to meet those aspirational goals, and—in the process of doing so—identify potential human rights risks and hazards that might inadvertently or inherently be lurking in the technology's lifecycle, as well as potential human rights opportunities to leverage, especially for technologies seeking to "make the world a better place".

To sharpen their analysis from a human rights perspective, designers might start their analysis by thinking about the 'standard' categories of vulnerable populations, including individuals or communities who are vulnerable due to their tenuous economic status or geographic factors that make it difficult for them to access a certain new technology (for example poverty or extremely rural communities), age, poor health, or individuals who are physically or intellectually differently abled, individuals who face communication barriers with mainstream institutions they depend upon, and of course individuals and communities who face social or cultural discrimination based on, inter alia, their race, ethnicity, sex, gender preferences, political viewpoints, or health status, etc. This list is not exhaustive, and would need to be updated reasonably based on the particular dynamics of a community for which the technology is being designed.

For some technologies, especially those that promise to alter the way fundamental social, biological, cultural, or natural processes work, this futures methodology may also need to assess the impact of a technology on future generations of human beings. For example, in considering a project to develop general-level AI that exceeds human intellect and therefore would eventually eliminate the need for humans to become educated (since no matter how long we learn we will never outcompete the AI that will, from that moment forward, do all the thinking for all of humanity) a futures thinker might consider the intangible value of learning and discovery that future generations might no longer enjoy, and then (as a necessary second step) design safeguards into the technology itself to prevent those future generations from involuntarily being deprived

190. Pendleton-Jullian, Ann M. & John Seely Brown (2018), Design Unbound: Designing for Emergence in a White Water World - Ecologies of Change (Vol.2), p.19

of those fundamental aspects of what it means to be 'human.'[191] These are, of course, highly subjective discussions, which can greatly benefit from considerations of universal human rights norms.

If, for example, technologists decide to create a ride-sharing application in Abu Dhabi – the city with the highest foreign-born immigrant population in the world[192] – with the aim to eliminate all traditional taxi companies (a primarily profit-oriented technology falling into either categories 2 or 3 above), then a futures-thinking exercise might identify that travelers from foreign countries who have not previously installed the applications on their phones, those without online bank accounts or access to stable internet, and those who do not have or wish to have a mobile telephone would be virtually excluded from the market for taxi services in an entire urban area. This type of futures thinking (from a 'do-no-harm' perspective) would necessitate some consideration of how to address the rights of vulnerable populations (such as, immigrants, the poor, the elderly, and differently-abled persons who might be prevented from using cell phones) to continue accessing the services they depend upon to live in an urban environment.

This vulnerability analysis would of course also benefit from a proactive consultation process (and other forms of feedback loops), where those so-called 'vulnerable' communities are specifically consulted, either directly or through civil society representatives, to see how they might be impacted by a future technology. Assuming the technologists wish to not only 'do no harm,' but also 'make the world a better place', the designers can also consult openly with those vulnerable stakeholders on how the technology might even help them in the future, thereby increasing their level of well-being, not just leaving it the same. Of course, it is necessary that inputs from such vulnerability assessments and consultations are meaningfully considered and incorporated into the design and development of the technology. Having cross-functional and inter-disciplinary teams, as well as ensuring diversity of representation amongst the members of the team designing and developing the technology, can help integrate considerations of how the technology might impact various groups at an early stage.

The design phase is also when technologists can ensure that all aspects of security are considered including building relevant safeguards, to the extent possible, and in line with the potential risks of a new technology in the case of a lapse in security. At this stage, technologists must also be mindful of any relevant general or industry safety standards and ensure that the design of the technology complies with such standards. In case of automated systems such as AI tools, technologists should think about ensuring there is a "human-in-the-loop" and devise ways in which such human control can be meaningful. Similarly, designers can think of how to craft meaningful grievance processes that will ensure that individual rights holders can be assured of their right to a remedy if a new technology inadvertently does jeopardize their rights. Designers can also craft specific strategies (and proactive policies) detailing how and when their technology will be brought to underserved markets (technology transfer), either by virtue of licensing arrangements or proactive efforts by the company to extend those services even to areas where the market may not immediately make such an investment profitable.

HRBA@Tech Intervention Vectors in the Design Phase (For Technology Categories 1-5)	
Futures Thinking + Vulnerability assessment	How will a technology – if and when it reaches full maturity and adoption by all relevant audiences – impact the most vulnerable in society, and can those vulnerabilities be somehow lessened by means of concrete design innovations in the technology itself before it goes to manufacture? (Corresponds with 2A on p. 68: Impact Assessment)
Security Research	Careful stress-testing of the technology to ensure that there are no unintended consequences – even inadvertent or accidental ones – that could jeopardize the rights or well-being of impacted stakeholders. (Corresponds with 3A-D on p. 68: Guardrail Innovation, Use Safeguards, Standardization, and Meaningful human control)
Grievance Processes	Design of grievance processes that can accompany the technology and that can serve as 'hazard indicators' in cases when the technology does not meet its stated purposes or when the technology inadvertently leads to negative consequences for rights-holders. (Corresponds with 4B-D & F on p. 68: Development of Accessible Non-Judicial Grievance Mechanisms and Processes, Monitoring & Oversight, Constructive Problem Solving, & Clearly Identified Responsible Entity).
Technology Transfer	Technologists should already be thinking during the design phase how their technologies will be shared with underserved markets as part of the technology diffusion phase. This may be part of the company's future ESG strategy, but should be planned by design. (Corresponds with 5D on p. 68: Technology Sharing & Transfer).
Transparency	If the design team hopes to use the technology to actively empower vulnerable populations, it would need to first 'translate' the projected impact of that technology into terms that a vulnerable population can understand. (Corresponds with 6A & 6B on p. 68: Disclosure & Translation).
Consultation	Designers of NET should consult with vulnerable communities and any other potentially impacted groups to not only ensure that their technology 'does no harm' but that it also 'makes the world a better place.' (Corresponds with 7A-C on p. 68: Consultation & Collaborative Needs Assessment, Representative Advocacy & Multi-stakeholder dialogue).

Manufacture (& Regulatory Approval)

The third phase of the TLC is the manufacturing phase. Each technology is different, and some technologies do not require physical manufacturing at all. Social media platforms, for example, might give rise to companies worth billions, but not manufacture a single product (except, perhaps, advertising t-shirts or other such merchandise with logos emblazoned on them). Other technologies are highly dependent on manufacturing, notably physical 'things' that we use to drive, fly, get healthy, play games, watch movies, listen to music, wear, or fight wars. For such physical objects, the logistics of the manufacturing process is a crucial element of bringing a product across the proof-of-concept stage. For still other technologies, notably the biotech industries, the 'manufacturing' stage may also entail significant regulatory compliance and approval processes, which are often far more complicated than figuring out the logistics of manufacture.

Intervention Points:

The manufacturing stage is rife with numerous human rights intervention points, most of which are not unique to NETs at all. At the manufacturing stage, technologists must comply with the relevant laws and policies regulating various aspects of the manufacturing process including compliance with relevant quality control standards to avert potential product liability downstream, fair labor practices, responsible and ethical supply chain management, consumer protection regulations and supply chain human rights due diligence amongst others. Technologists must also ensure that there are internal policies in place to give effect to these regulatory safeguards, including incorporating requirements associated with ESG policies and procedures.

At this stage, companies can also make active efforts to ensure representation across supply-chains by diversifying suppliers and vendors. It is also necessary to ensure relevant monitoring and oversight mechanisms at the manufacturing stage to enable inspection of the manufacturing processes and avert any potential human rights abuses. Such a monitoring and oversight function can also be performed by civil society as independent observers, and technologists should therefore provide adequate channels of access to civil society. Companies must also establish internal grievance processes to enable employees working on such new and emerging technologies to file complaints or claims pertaining to any harm arising out of the process of developing such technologies and ensure access to meaningful remedy.

191. Risse, Matthias, (2021) "The Fourth Generation of Human Rights: Epistemic Rights in Digital Lifeworlds," 8(2) Moral Philosophy and Politics 351-378

192. ¿?

HRBA@Tech Intervention Vectors in the Manufacture Phase	
Environmental, Social, and Governance Policies	ESG policies and procedures should be built into the manufacturing phase of any NET. (Corresponds with 1C on p. 68: Mainstreaming Human Rights).
Supply Chain Management Policies	Responsible and ethical supply chain management policies need to be enacted in order to ensure that there are no negative human rights impacts at any stage of an NET's supply chain. (Corresponds with 1C, 2B, 4C-D, 5A-C & 7A-C on p. 68: Mainstreaming Human Rights, Internal Diversity, Monitoring & Oversight, Constructive Problem Solving, Human Rights by Design, Community Mobilization, Capacity Building, Consultation and Collaborative Needs Assessment, Representative Advocacy, & Multi-Stakeholder Dialogue).
Consumer Safety and Protection Policies	Policies protecting consumers are necessary for all NETs. (Corresponds with 3A-D on p. 68: Guardrail Innovation, Use Safeguards, Standardization, & Meaningful Human Control).
Fair Labor Practices	Companies developing NETs need to ensure that they have fair labor policies and that they are not violating the human rights of their employees. (Corresponds with 5A-C & 7A-C on p. 68: Human Rights by Design, Community Mobilization, Capacity Building, Consultation & Collaborative Needs Assessment, Representative Advocacy, and Multi- Stakeholder Dialogue).
Establishment of Internal Grievance Processes	Internal grievance processes for employees are necessary for any company working on an NET, so that its employees can file internal claims for issues such as discrimination or unlawful practices. (Corresponds with 4B & 4D on p. 68: Development of Accessible Non-Judicial Grievance Mechanisms and Processes, & Constructive Problem Solving).
Independent Monitoring	There should be independent monitoring of the manufacturing stage of all NETs, both internally and externally. (Corresponds with 4C on p. 68: Monitoring & Oversight).
Civil Society Involvement	Civil society should be consulted and involved in the manufacturing stage, whether as independent observers or other types of participants. (Corresponds with 5B & 7A-C on p. 68: Community Mobilization, Consultation and Collaborative Needs Assessment, Representative Advocacy, and Multi-Stakeholder Dialogue).

Adoption & Marketing

Assuming the founders successfully clear the first three phases of the TLC process, they will have crossed the proof-of-concept threshold. At that point, assuming their planning was successful, their technology (or product(s) built on that technology) will be ready to be released into the real-world target environment. For many such technologies, this will be the market, where the founders hope to earn profits that will allow them to pay off their creditors and grow their business. For other technologies – those with a social mission for example – this may not be an open market but rather an environment where the technology will hopefully 'solve' a particular social or environmental problem.

Initially, only innovators and early adopters will use a new technology. At this point, the adoption of that new technology may be relatively sluggish, with only modest growth. Eventually, however, these earlier influencers and their (hopefully positive) experience with the new technology will then spur a much larger number of users (the 'majority') to adopt the new technology. At this point, the rate of adoption will skyrocket, with accelerating (exponential) growth. That phase is unlikely to last forever, however, since eventually competitors will crop up, new entrants will dwindle, and the technological adoption process will again level off and slow, with only a few remaining 'laggards' still adopting the technology. This start 'slow > accelerate > decelerate > stagnate' cycle is often described as the S-Curve of the TLC.

Due diligence of government sponsors and potential use of a NET. The adoption and marketing phase of the TLC also opens up numerous fairly obvious opportunities to intervene from a human rights standpoint. Some of these mechanisms can still be driven internally by the company or agency promoting the new technologies. These processes involve continuous due diligence and impact assessments, adopting ethical marketing strategies, and participating in outreach and consultation efforts. This also includes actively monitoring and overseeing internal grievance procedures. Additionally, technologies should be adapted, as much as possible, to prevent misuse by unscrupulous third parties. This can be achieved by incorporating procedural or other safeguards for use.

Since the technology will be widely accessible, internal safeguards should be complemented by external human rights measures. These external measures include monitoring by civil society activists, regulatory oversight, and judicial accountability in instances of illegal activity or negligence.

HRBA@Tech Intervention Vectors in the Implementation Phase	
Due Diligence	Developers of NETs should conduct ongoing due diligence to protect from (and correct) negative human rights of their product(s). (Corresponds with 2A on p. 68: Impact Assessment).
Ethical Marketing	Ethical marketing requires a focus on not only how the NET benefits customers, but also how it 'makes the world a better place,' by, for example, benefiting socially or environmentally responsible causes. It includes avoiding false or misleading claims or representations of the product. (Corresponds with 1C, 3B, & 6A-B on p. 68: Mainstreaming Human Rights, Use Safeguards, Disclosure & Translation).
Outreach & Maintenance of Grievance Procedures	Companies should conduct outreach to potentially affected communities and groups and maintain grievance procedures for anyone negatively affected by their products. (Corresponds with 6A-B, 4B, 4D & 4F on p. 68: Disclosure, Translation, Development of Accessible Non-Judicial Grievance Mechanisms and Processes, Constructive Problem Solving & Clearly Identified Responsible Entity).
Adjustments to Prevent Distortion	The necessary adjustments need to be made by those creating and marketing technologies to prevent distortion. (Corresponds with 3B on p. 68: Use Safeguards).
Naming & Shaming	Stakeholders such as civil society have the responsibility to disclose negative human rights impacts of NETs. (Corresponds to 5B-C, 7A-C, 4A-B & 4F on p. 68: Community Mobilization, Capacity Building, Consultation & Collaborative Needs Assessment, Representative Advocacy, Multi-Stakeholder Dialogue, Access to the Formal Justice System, Development of Accessible Non-Judicial Grievance Mechanisms and Processes & Clearly Identified Responsible Entity).
Regulatory Enforcement	Regulators must create appropriate frameworks for scrutiny of adoption and marketing processes and ensure adequate enforcement mechanisms and processes are in place. (Corresponds with 1A-C, 2A, 4C, 4E-F & 6A-B on p. 68: Implementation of International Human Rights Norms, Policy Coordination, Mainstreaming Human Rights, Impact Assessment, Monitoring & Oversight, Incentivization, Clearly Identified Responsible Entity, Disclosure & Translation).
Judicial Accountability	There must be judicial accountability for victims who suffer human rights violations at the hands of NETs. (Corresponds with 1A, 1C, 4A & 4F on p. 68: Implementation of International Human Rights Norms, Mainstreaming Human Rights, Access to the Formal Justice System, & Clearly Identified Responsible Entity).

Diffusion

Numerous TLC models posit that at the point when a technology is about to reach maturity – in other words when the exponential growth of new users adopting a technology begins to level off – companies might have stronger incentives to enter into novel licensing agreements whereby they allow others to extend the technologies into new markets, perhaps adapting those technologies to serve new corporate agendas. For profit-driven technologies (categories 2 & 3 above), this process makes business sense, in that it allows the technology owners to enter into new and lucrative partnerships at the precise moment when their own profits from independently marketing the technology may be beginning to wane. For social entrepreneurs, on the other hand, such subcontracting and licensing arrangements might always make sense if it means that the impact of the new technology will be multiplied.

Intervention Points:

The diffusion phase also presents several opportunities for interventions from a human rights standpoint. In continuation of the previous stages, it is necessary for companies to conduct ongoing due diligence and impact assessments of the new and emerging technology. At this stage, it is also relevant for companies to carefully vet potential licensees, including governments, and take all other precautionary and reasonable safeguard measures prior to entering into a licensing agreement to ensure the technologies are put to uses as intended and not used to harm or commit human rights violations.

HRBA@Tech Intervention Vectors in the Diffusion Phase	
Vetting of Potential Licensees	Potential licensees of NETs must be thoroughly vetted to ensure that they will not use the licensed technology to harm others or commit human rights violations. While an 'owner' of any particular NET cannot ensure with complete certainty that a licensee will not use the NET in a harmful manner, licensors should engage in practices such as reviewing the human rights records of potential licensees (e.g., in the example of a government licensee, examine the government's human rights record) and the licensee's stated desired use of the NET. (Corresponds with 2A, 3B, 4E, & 7A-C on p. 68: Impact Assessment, Use Safeguards, Incentivization, Consultation and Collaborative Needs Assessment, Representative Advocacy, & Multi-Stakeholder Dialogue).
Due Diligence	After licensing or otherwise diffusing an NET, the owner (as well as the licensee) should conduct ongoing due diligence to ensure that the technology is not having any potential negative human rights impacts. (Corresponds with 4C-E on p. 68: Monitoring & Oversight, Constructive Problem Solving & Incentivization).
Installation of Technological Safeguards to Prevent Abuse	As detailed in Chapter 3, guardrails should be installed within any NET to prevent negative human rights impacts. (Corresponds with 3A on p. 68: Guardrail Innovation).

Refinement

After a new technology has matured, the TLC, like all lifecycles, predicts the eventual demise of even the most exciting and revolutionary novel technologies. However, while the demise of one specific technology may be inevitable, a technology company (or an organization making use of a particular technology) can stave off its own corresponding fading out by engaging in a process of reflexive self-analysis and refinement. For instance, the company formerly known as Facebook underwent a significant transformation to address its declining business model. It rebranded itself as Meta, launched a new line of virtual reality headsets, and invested billions of dollars into creating worlds within the Metaverse, aiming for a company rejuvenation.

Intervention Points:

Such refinement processes again offer up some unique opportunities for human-rights focused intervention points. Quite unlike the earlier Implementation and Diffusion phases, when technology companies (certainly those in category 4) tend to be focused primarily on the active promotion and marketing of their new technologies, during a refinement process companies tend to be very receptive to input, consultation and critical reflection. This is a time when technology companies tend to be openly welcoming of 'constructive feedback' from the human rights community, especially if that feedback suggests ways they might re-energize their product line and thereby also re-establish themselves as market innovators.

Thus, two primary vectors of influence present themselves, the first focused on internal efforts to solicit information, and the second focused on vetting those new ideas for their consistency with those human rights objectives, similarly to what one might have asked of a social entrepreneur promoting a category 1 technology during the innovation phase (with the crucial difference that as a well-resourced corporations one can expect much higher levels of sophistication during this phase of innovation than one might of a start-up).

HRBA@Tech Intervention Vectors in the Diffusion Phase	
Vetting of Potential Licensees	Potential licensees of NETs must be thoroughly vetted to ensure that they will not use the licensed technology to harm others or commit human rights violations. While an 'owner' of any particular NET cannot ensure with complete certainty that a licensee will not use the NET in a harmful manner, licensors should engage in practices such as reviewing the human rights records of potential licensees (e.g., in the example of a government licensee, examine the government's human rights record) and the licensee's stated desired use of the NET. (Corresponds with 2A, 3B, 4E, & 7A-C on p. 68: Impact Assessment, Use Safeguards, Incentivization, Consultation and Collaborative Needs Assessment, Representative Advocacy, & Multi-Stakeholder Dialogue).
Due Diligence	After licensing or otherwise diffusing an NET, the owner (as well as the licensee) should conduct ongoing due diligence to ensure that the technology is not having any potential negative human rights impacts. (Corresponds with 4C-E on p. 68: Monitoring & Oversight, Constructive Problem Solving & Incentivization).

Maturity

Assuming that a technology or the company or institution promoting that technology cannot reinvent itself by refining its technology, the adoption of a new technology will eventually peak and the product or technology will reach its maturity. If the technology is being promoted by a corporation, at this point the company will stop growing, and therefore begin to lose its appeal to investors and shareholders. Such companies might often seek to merge with other companies (to capture economies of scale or reduce inefficiencies), shrink their workforces, or otherwise try to cut costs in order to keep their profits growing. Eventually, however, the TLC model would predict that the technology's adoption rates will continue to shrink and eventually drop off altogether, oftentimes just as precipitously as they once grew.

In such situations, all an ethical technology company can do to remain consistent with the HRBA@Tech model is to plan for that downsizing in ways that do minimal harm to the human rights of those who stand to be impacted.

Intervention Points:

As always, companies in such situations should continue to engage in due diligence monitoring. Are there certain communities that stand to be disproportionately impacted by the gradual erosion of a technology, and strategies should be developed that will minimize these anticipated harms. Can the disruptions to the business model be mitigated somehow, perhaps by virtue of transparent communication strategies, gradual workforce reductions, and ethical re-investment strategies to ensure that the negatively impacted communities can invest in reskilling activities? These are some of the difficult questions that managers of such technology companies should ask themselves as they reach the maturity phase of a technology.

HRBA@Tech Intervention Vectors in the Maturity Phase	
Due diligence	As with other phases, due diligence must continue to be carried out, and also adapted to the Maturity Phase. (Corresponds with 2A & 7A-C on p. 68: Impact Assessment, Consultation & Collaborative Needs Assessment, Representative Advocacy & Multi-Stakeholder Dialogue).
Ethical re-investment strategies	When considering ways to reinvest re-invest resources (whether financial or other), companies or other 'owners' of NETs must evaluate the ethical implications of that reinvestment and carry out ethical investment strategies. (Corresponds with 1C, 4E, 5C, 5D on p. 68: Mainstreaming Human Rights, Incentivization, Capacity Building & Technology Sharing & Transfer).
Transparent communication	Companies and other 'owners' of NETs must engage in a transparent manner with stakeholders and the public in general. (Corresponds with 6A-B on p. 68: Disclosure & Translation).

Irrelevance

Finally, at the very end of the TLC, a given technology will eventually become irrelevant. At this phase, the company will have either reduced significantly or gone out of business entirely, and most of its workforce will have likely already moved on to other industries.

Intervention Points:

At this point, the HRBA@Tech model suggests only that a company can still engage in planned obsolescence and redundancy planning.

Managers should ask themselves what they can do to help transition departing workers from a gradually decaying company to other industries. What can be done to help sustainably recycle any old and no-longer-relevant physical products that relied on the increasingly irrelevant technologies? Is there data or other irretrievable goods that need to be salvaged or be lost as one underlying technology fades into the past? From a human rights perspective, these considerations must not be neglected as part of the final wrapping up process of an enterprise associated with a dying technology.

HRBA@Tech Intervention Vectors in the Irrelevance Phase	
Planned Obsolescence & Redundancy Planning	Consider any potential negative human rights impacts from the obsolescence of an NET, as well as how to sustainably recycle or salve any goods or technologies. (Corresponds to 5A, 1C & 7A-C on p. 68: Human Rights by Design, Mainstreaming Human Rights, Consultation and Collaborative Needs Assessment, Representative Advocacy, & Multi-Stakeholder Dialogue).

CHAPTER 5: "THE WHO" OF THE HRBA@TECH MODEL

Chapter Summary:

This Chapter highlights the role that different stakeholders play within the HRBA@Tech model. It explores the rights, obligations and responsibilities of these actors, and also discusses the sources of influence or leverage that they can wield or enjoy as they apply the model through the TLC described in Chapter 4.

This Chapter describes the HRBA@Tech model through the lens of stakeholders. It describes the rights and obligations, as well as the roles and responsibilities that each of these categories have across the TLC, as well as the ways that multi-stakeholder coalitions can form to ensure that NETs serve to respect, protect and fulfil human rights.

The Chapter is framed at an abstract level. A technologist might, for example, see a major distinction between the CEO of a corporation and an individual engineer employed by that corporation, and yet both of those individuals would be described in this chapter as falling under the umbrella term "private sector." Similarly, all of us are individual rights holders, not only the CEO and the engineer, but also a worker at a technology factory, a consumer of an innovative technology- based product, and also an indigenous person in a community that barely has access to NETs. Moreover, individual rights holders often also have some notion of 'community rights,' for example related to a community's right to development or the right of a marginalized community not to be harassed or exposed to representational harm by unchecked hate speech on social media or biased AI systems.

The chapter is organized with the State and the individual arrayed at the edges of a spectrum. This model is taken from introductory political philosophy, which theorizes the reciprocal social contract between States and individuals as the basis for political society. Thomas Hobbes (1588-1679) theorized States (Sovereigns) and individuals entering into a mutual contract – the individuals ceding their unfettered freedom and agreeing to respect the Sovereign's authority in exchange for the State (Sovereign) providing a sense of safety to all of its subjects (individuals). Thus—so the theory goes—States work to protect and promote the rights of its citizens in exchange for their allegiance and taxes. In a healthy political system, the individual's expectation of safety (having his or her individual human rights protected) is matched by a corresponding State obligation to safeguard those rights. Similarly, the individual owes the State a duty to respect laws and regulations and generally behave in a more 'civilized' manner than would have been customary in the State of Nature. In the more modern language of our current era, this reciprocal relationship between a State and an individual also requires individuals to assume a personal duty of care (ethics) towards the plight of others with whom they share a community. Thus, the language and logic of rights and State obligations, which exists primarily between a State and an individual, is supplemented by a parallel and complementary language of responsibilities, which flow like a supportive ether between private individuals, groups of individuals (communities), and other entities and institutions.

In societies, a number of institutions have come to intermediate these relationships, most at the behest of those original two stakeholders. States, for example, created international organizations like the United Nations, thus delegating some of their State functions to these international institutions. Similarly, individual rights holders also sometimes transfer some of their authority to civil society organizations in an effort to better claim their rights, have their interests represented or take action on their sense of responsibility towards others. Individuals, in order primarily to earn a sustainable living, cede personal freedoms to private employers, who – in an echo of the original Hobbesian contract between a State and an individual – again owe their employees a certain duty of care (and a salary). Educational institutions educate a future generation of scholars and professionals, but also feel a sense of responsibility to inculcate pro-social and sustainable values into its students. Each of those institutions "in the middle" have some mix of rights, responsibilities, and obligations flowing towards other stakeholders in this matrix.

In this chapter, we highlight only four such institutions:

1. The UN and other International Organizations;
2. Civil society;
3. The Private Sector, including Technology Companies;
4. Educational institutions.

Surely, more categories could be imagined, and these existing ones could be subdivided into a near-infinite kaleidoscope of increasing complexity. We will leave such exercises to analysts with a more specific focus on certain technologies, access to certain points within the Technology LifeCycle (TLC), and a better idea of potential resources they might mobilize in favor of the HRBA@Tech model. Any more detailed stakeholder analysis will likely need to push beyond these rudimentary six categories, breaking some down further according to the functions played by different sub-groupings. This analysis, by necessity, would be more context specific than the one contained in this chapter.

STATES

States are traditionally at the center of the international human rights framework. They are the primary duty-bearers who are vested with the obligation to respect, protect and fulfil human rights. This obligation entails that States must not only refrain from directly interfering with the enjoyment of human rights (i.e., do no harm), but also ensure protection from interference in the enjoyment of human rights by third parties. States act as the indirect enforcers of private actors' moral and ethical responsibilities towards one another. However, a State's commitment to human rights also entails the important obligation to proactively facilitate the enjoyment of human rights. A State cannot simply rest idly in the knowledge that its citizens' enjoyment of human rights is not getting worse; it must actually work proactively to make that situation better. This constant striving to protect and promote human rights is not a sign of a dysfunctional state, but rather the sign of a healthy one that functions as it should. As we have repeatedly stressed throughout this report, this means States also have a duty to harness the enormous potential of NETs to improve human well-being. Unlike private companies, States have always had the duty to "make the world a better place" through the steady and determined promotion of human rights.

With regard to NETs, States must ensure adequate and effective protection of human rights by devising relevant legal and policy frameworks that reflect the changing realities of ongoing technological developments. This requires States to constantly build their own capacities as duty-bearers in order to better meet their obligations to respect, protect and fulfil human rights in the context of NETs. They can do so by fully implementing relevant recommendations from international human rights mechanisms, establishing National Mechanisms for Implementation, Reporting and Follow-up (NMIRFs), and actively engaging in dialogue and cooperation at international level to share best practices in addressing the human rights implications of NETs.

States must establish and adopt appropriate frameworks that translate international human rights norms into locally-relevant laws, policies and practices for the promotion and protection of human rights during the development and deployment of NETs. This includes devising strategies to regulate or otherwise incentivize private companies to prioritize a focus on human rights throughout all stages of a TLC. Regulatory approaches can take the form of mandating human rights due diligence or impact assessment requirements, including requirements for such due diligence to be ongoing throughout the TLC as well as instituting transparency and disclosure requirements and corresponding mechanisms to track compliance. It also includes developing robust and accessible accountability mechanisms to redress grievances related to the development and deployment of NETs.

To do this, some States may wish to create new institutions. Others may prefer to strengthen existing institutions and regulatory bodies (for example data protection authorities or national human rights institutions). Ensuring accountability requires States to have in place accessible formal judicial mechanisms as well as informal or quasi- judicial dispute resolution, monitoring and oversight mechanisms and processes. States can also ensure accountability through a range of additional regulatory and governance efforts, including a mix of mandatory and voluntary measures. These might include the creation of incentive structures designed to raise the costs of non-compliance with human rights principles while offering tangible benefits for those actors who attempt to hard-wire human rights priorities into NETs (for example by embracing an HRBA@Tech model). States can create such incentives, by providing financial support to explicitly pro-social entrepreneurs or by easing the administrative burdens on NETs with a promise to 'make the world a better place'. Ensuring effective accountability also requires taking proactive steps to capacitate civil society actors to perform their crucial monitoring and accountability function.

States increasingly recognize their responsibility to protect their populations from influential corporate actors, especially those responding only to market-based incentives. To craft truly effective regulatory frameworks based on the HRBA@Tech model, States must proactively engage and build the capacity of all stakeholders (including civil society) to promote knowledge of, and buy-in to, the HRBA@Tech model, and to ensure that there is sufficient technical expertise at all levels of society to promote and protect human rights in the context of NETs. For this to happen, all stakeholders (not just States) will need to overcome a trust deficit that tends to posit the interests of one stakeholder category against those of the others. States, however, are in a unique position to convene multi- stakeholder dialogues and coalitions that offer the best chance for building that trust and creating a more collaborative and multi- stakeholder approach to NETs.

Additionally, States themselves frequently make use of NETs for various purposes, for example to improve the functioning of government systems, improve the quality and accuracy of public services, advance national security, and advance the realization human rights. In this capacity, States themselves can play an important precedent- setting role when they embrace the HRBA@Tech model whenever they develop or deploy NETs as many States already do. These are welcome developments that can greatly improve State capac-

ities to improve the well-being of their populations. The HRBA@Tech model provides a viable roadmap for States to hold themselves to the same high standards they should also expect of private actors operating in their territories.

In order to implement not just the "do no harm" pillar of the HRBA@ Tech model but also the "make the world a better place" pillar, States must actively facilitate and promote socially-beneficial technologies. They can do so by providing various forms of assistance, including regulatory, financial, promotional, or other forms of support. States should promote socially beneficial technologies after carefully vetting how such NETs would work to "make the world a better place," and only upon ensuring that such technologies do not, even inadvertently, harm people or violate human rights standards. One promising way to do so can be through the creation of "regulatory sandboxes," which promote innovation while enabling entrepreneurs to test their potential technologies for any issues including safety or bias, thus preventing human rights harms while also specifically promoting technological and regulatory dynamism.

States must also build their own capacity to embrace a holistic understanding of technology (embracing both the promise of NETs while simultaneously guarding against unintended human rights downsides) as well as a holistic approach to human rights (embracing both classical human rights concepts, language and institutional strategies while also engaging with other communities with their own established ethical standards). Whenever possible, states should adopt a learning mindset (as per the recommendation of the UNSG's High-Level Panel on Digital Cooperation in 2019). Such a commitment requires that States engage in multi-stakeholder cooperation, as was also recommended by the Advisory Committee. This can mean leveraging technical expertise internally by facilitating coordination between various departments or ministries within the government, and bridging gaps between centers of technological and human rights expertise. This can also mean engaging externally with other stakeholders, particularly the tech community, international organizations, civil society, and educational institutions. Governments can incentivize collaborations, partnerships or multi-stakeholder dialogues. They can (and should) hire experts to close any potential knowledge gaps which may prevent them from effectively crafting appropriate policies to 'nudge' NETs towards the direction of human rights. Finally, they should open their inward-focused capacity- building efforts also to other stakeholders to promote awareness about the language and logic of human rights as well as the specifics of the technology sector.

States must also cooperate to develop, at the international and regional levels, binding as well as non-binding normative frameworks for the protection and promotion of human rights in the context of NETs. They can do so by creating new instruments or expanding and clarifying the scope and application of existing ones to NETs. The Human Rights Council provides an ideal forum for such human rights normative development, and States should fully leverage this space to consider the human rights implications of NETs in a manner that minimizes selectivity and politicization.

One of the recurring themes throughout this policy paper has been the inherent potential of some NETs to have disparate impacts on differently situated populations, often by entrenching and accentuating existing social inequalities. States have the primary obligation to address such inequalities. In the context of NETs, States should assess and systematically mitigate these direct and indirect discriminatory effects. This requires, as an initial matter, having laws in place that prohibit discrimination on the basis of traditionally protected characteristics or any other characteristic as per local contexts or conditions. It might also require adding additional protected classes of individuals to be shielded from discrimination, such as whether one has access to certain technologies or whether one wishes to exempt oneself from using a particular NET. States must ensure equitable access to the benefits that NETs provide by bridging the digital divide at various levels. Governments must ensure that vulnerable population groups and entities such as financially weak private actors or small enterprises with limited resources or civil society organizations nonetheless have access to opportunities that NETs provide, and that they can harness the benefits that flow from them. Governments can do so by engaging in capacity-building, securing equitable access to digital infrastructure, and by promoting literacy and technical skill-building for rights-holders to better claim and enjoy their rights in relation to NETs. Governments must also work to deploy NETs in order to improve their own functioning as well as the quality and access of public service delivery and dispensation.

In doing so, governments must ensure that they embrace the HRBA@ Tech model into its own development and deployment processes for those NETs.

In an interstate context, States, especially developed, middle- income or upper-income States should consider actively supporting technology transfers and engage in genuine efforts to share the benefits of scientific progress internationally. This can be done in the context of bilateral or multilateral development aid, and might involve capacity building, technology transfer, as well as a reformed approach to Intellectual Property protections that often serve to prevent such knowledge transfer. As the world painfully witnessed during the global COVID-19 pandemic, failures to equitably share the fruits of scientific knowledge (i.e., vaccine nationalism) ultimately made everyone worse off.

States also have an obligation to ensure the general safety of NETs. States can frame relevant laws and policies for that purpose, drawing appropriate ethical "red lines" where applicable, for example by strictly restricting or banning the development and deployment of certain technologies that may be especially high-risk. This obligation requires states to create institutions designed to promote transparency and engagement. States must also encourage standard-setting efforts at the national, regional and international levels to guide the design, development and deployment of such technologies. Such standards should be crafted in a form that technologists can use to evaluate their NETs, complementing abstract standards with concrete processes and quantifiable metrics. States must establish relevant regulatory authorities to monitor and oversee the development and manufacturing of such NETs in adherence with relevant standards and ensure quality control. States especially can ensure that the discussions over how to craft these regulatory standards embrace a holistic approach to technology, human rights, and governance.

THE UNITED NATIONS AND OTHER INTERNATIONAL ORGANIZATIONS

The UN and other international organizations play a key role in advancing the global development of human rights norms and standards in relation to NETs. Such institutions convene the world's expertise and diplomatic voice and facilitate the gradual emergence of authoritative and widely embraced norms, policies and better practices. While such normative development processes tend to be slow, sometimes frustrating, and subject to the interests of States, the importance of international organizations lies in their convening power. They can take up an important issue like this one and consistently bring together various stakeholders to systematically identify consensus positions on how best to address global challenges. The UN and other international organizations provide key platforms for multi-stakeholder engagement and cooperation, which are crucial for devising and informing holistic strategies on how to balance competing interests in the field of NETs and human rights. International organizations have a particular obligation to ensure that traditionally vulnerable or marginalized communities are represented in these multi-stakeholder discussions, either directly or indirectly through civil society actors. International organizations should actively build the capacity of such vulnerable communities, either by themselves through their development activities or in collaboration with relevant local stakeholders such as governments and civil society. Finally, as with all other operations, international organizations should be mindful of geographic representation in this field, drawing expertise not only from the developed world but also the Global South.

International organizations like the UN and others are often noted for their complex bureaucratic structures, many of which seemingly duplicate functions in several different organs. This can be confusing in some situations, but in others it can also be an asset, as different organs competitively jostle for relevance and gradually drive forward a cross-sectoral discussion on important issues. The labyrinthine institutional structures of the UN can drive forward a range of discussions on the human rights implications of NETs. This should include the international human rights mechanisms and in particular, the Human Rights Council, which is tasked with coordinating the UN's human rights activities. The Special Procedures falling under the mandate of the HRC and the various Treaty Bodies are key drivers of human rights normative development. To-date, however, the international human rights mechanisms have not always been able to effectively translate the human rights normative framework into clear policy prescriptions to address the implications of NETs (though significant progress has been made). While this may require the normative development of human rights law (a question far beyond the scope of this report), it certainly implies strengthening the capacity of existing human rights mechanisms to clarify the application of human rights law to various technological contexts.

Efforts must also be made to mainstream human rights, its language and logic, within international organizations and agencies dealing with aspects of NETs. Certainly, several Treaty Bodies and some Special Procedures have addressed the human rights implications of technological developments. Nonetheless, these mechanisms, along with others like the Universal Periodic Review, could begin to work to develop more actionable recommendations related with regard to human rights and technology. International organizations must also integrate the expertise of international human rights mechanisms, including the OHCHR and other technical bodies, to ensure greater policy coordination of activities amongst the various organs and agencies. Doing so will close potential knowledge gaps

(or prevent them from opening) and ensure that human rights and NETs are handled holistically at the international level. Standard setting organizations, in particular, can play a key role in developing relevant safety and other standards to guide the design, development and deployment of new and emerging technologies and to ensure they are directed towards the protection and promotion of human rights. In doing so, they should draw on the expertise and insights of international human rights mechanisms and bodies.

The international community should consider the establishment of a new Special Procedure, such as Special Rapporteur or a Working Group, with the mandate to explore the paradox of technology and human rights in greater detail and further delineate concrete recommendations for the better protection and promotion of human rights in the context of NETs. While various Special Procedures, Treaty Bodies, and other human rights mechanisms have discussed aspects of NETs, they have done so within their usually narrower thematic focus, and have rarely engaged with the human rights implications of NETs in a more general and overarching way. We argue that this can only be accomplished through a designated mandate specifically for that purpose. Such a mandate for NETs would give a qualified expert or set of experts access to channelized resources, logistical and diplomatic support to undertake research and consultative activities, as well as the opportunity to undertake extensive fact-finding, consultation and consensus-building processes with various stakeholders. The outcome of such a mandate – at least in the short term – could be to elaborate a human-rights based approach (perhaps building upon the HRBA@Tech model presented in this report) to the development and deployment of NETs.

The UN and other international organizations can also act as vital sources of information for states and other stakeholders. Relatively small teams of substantive experts, supported by a small but competent secretariat, can stay abreast of the rapid developments and complex science that usually characterizes technological innovation, and dedicate themselves to the sharing and promotion of knowledge as well as any best practices of how best to deal with associated human rights challenges (or capture human rights opportunities). They can disseminate information to provide guidance to private actors/companies, especially the tech community, to better incorporate and operationalize human rights in the development and deployment of new and emerging technologies. The work of the OHCHR's B-Tech Project is very relevant in this regard, as it already provides guidance on the application and implementation of the UNGPs in the technology context.

International organizations can also engage in capacity building by means of directly providing technologies or supporting digital infrastructure, technical know-how, expertise, and other forms of resources and assistance to various actors for them to better leverage and harness the benefits of NETs while also mitigating associated human rights risks. International organizations can facilitate technology sharing and transfer arrangements amongst countries to promote equitable access to socially-beneficial technologies. This is especially true less-developed and low-income countries, where such activities would actively promote progress with regard to the Sustainable Development Goals (SDGs).

Lastly, international organizations, including the UN, can also better leverage NETs to improve their operations in the achievement of their mandates and "make the world a better place." In doing so, international organizations must themselves also embrace the HRBA@Tech model.

CIVIL SOCIETY

Civil society includes a range of organizations and entities at various levels of governance, including general human rights organizations as well as specialized non-profit groups focused on technology and human rights. It also, of course, includes less explicitly human-rights or technology-oriented groups that nonetheless engage with these themes based on their particular institutional orientation. Civil society plays a key role in the advancement of human rights, particularly as a bridge between rights-holders and duty-bearers. Civil Society plays a crucial role in representing the interests of vulnerable communities and ensuring that the voices of their constituencies, who may face countless barriers to effectively articulate their needs or participate in decision-making processes, are nonetheless heard by decision makers. In the context of new and emerging technologies, this can include engaging with both public (State) actors, but also private actors, such as tech companies and various other stakeholders.

Civil society must actively participate in multi-stakeholder efforts to govern and regulate new and emerging technologies for the better protection and promotion of human rights. Civil society often already does (and should continue to) actively forge networks, partnerships and collaborations bridging the Global North and Global South divide. Civil society actors already do (and should continue to) engage with all possible mechanisms where the development and deployment of NETs are discussed in order to better represent their constituencies and interests. In doing so, civil society must ensure they advocate for the interests of the constituencies they represent, and that they work towards the realization of those communities' goals. This requires civil society to make itself accountable to its constituencies. It also requires that they embrace a community-centric approach where communities are involved in the co-design of advocacy goals, empowerment strategies and intervention approaches.

Civil society must also actively engage with all other stakeholders through relevant partnerships and collaborations. This requires, of course, that other stakeholders (such as governments, companies and international organizations) must ensure the open access of civil society into the institutions, mechanisms, processes, and fora where decisions with respect to the development and deployment of new and emerging technologies are made. This requires the creation of participatory platforms and processes, including the frequent use of open-ended consultations, the inclusion of civil society representatives into policy making processes or directly into regulatory and governance frameworks, opportunities for monitoring, oversight, and active engagement in human rights due diligence and impact assessment processes, amongst other initiatives.

Finally, Civil Society organizations should seek out the support of other stakeholders in order to themselves benefit from NETs so that they too may better secure their operations and harness the power and efficiency of NETs, putting them in the service of their constituencies and their interests. Civil society, subject to their resource constraints, can leverage NETs to improve their operations while ensuring adequate safeguards, such as robust privacy and data protections for the communities they represent.

PRIVATE SECTOR INCLUDING TECHNOLOGY COMPANIES

The private sector is the primary driver of technological and scientific innovation today, and is therefore central to the development and deployment of NETs. This includes "tech-giants"—major corporations often with annual profits rivalling GDPs of mid-sized developed economies—but also other tech companies of various sizes including start-ups. The ecosystem of private actors associated with NETs also includes other entities, including commercial banks, wealth managers, venture capital firms and international development organizations. As digital technologies have transformed virtually all sectors and areas, NETs are increasingly an integral component of the business models of a range of companies, not just tech companies. Private actors are compelled by market pressures to 'out-innovate' their competitors and keep churning out 'the next big thing', as failure to do so can lead to the decline and eventual obsolescence of even the most innovative of businesses.

Private actors typically exist to generate profits. The HRBA@Tech model in this report is built with this reality in mind, and attempts to balance these competing interests while ensuring that the development and deployment of new and emerging technologies is nonetheless better positioned to protect and promote human rights. While private actors do not have direct obligations under international human rights law, over the years there has been growing recognition of the crucial role they play in the advancement and realization of human rights and the need for corresponding responsibilities leading to efforts to accommodate private actors within the international human rights framework. In this regard, the UNGPs provide an authoritative framework for the corporate responsibility to respect human rights and a reference point for companies involved in the development and deployment of new and emerging technologies.

In light of the truly transformative potential that new and emerging technologies hold, this report suggests that a "do no harm" approach is no longer sufficient. The HRBA@Tech model described in this report suggests moving beyond the UNGPs to embrace the possibility of actively crafting NETs to put them in service of human rights, or – to put it simply – to "make the world a better place." The HRBA@ Tech model recognizes the interests and constraints faced by private actors, and acknowledges that the promotion of human rights must always be weighed against the prerogative to continue generating profit. The HRBA@tech model need not be antithetical to the interests of private actors and companies or incompatible with most existing business models.

Private enterprises can incorporate various processes within the HRBA@Tech model directly into the TLCs of NETs, either on their own or jointly with other stakeholders. At the outset, private actors must comply with the processes structured on the UNGPs and must have human rights policies in place making formal and explicit commitments to relevant human rights standards. For those enterprises intending to develop socially-beneficial technologies, human rights policies should also highlight the intended human rights objectives of such a technology and the proposed roadmap for achieving them. Entrepreneurs promoting such technologies should embrace a "human rights by design" process. The HRBA@Tech model recommends conducting human rights due diligence and impact

assessments to ensure that the new and emerging technologies do not, even inadvertently, harm people. It also requires paying special attention to those constituencies that may be particularly vulnerable to the impacts of NETs, including those who may wish to opt out of its use. Private actors must conduct human rights due diligence right from the initial stages of the innovation and design of NETs. Impact assessments are important at the early stages of an NET's development in order to anticipate potential human rights impacts of the technology, but also essential later on its TLC to assess the actual impact of that technologies once it has been embraced more widely. Such assessments should focus in particular on whether there has been any disparate impact of an NET on vulnerable or marginalized groups and communities.

It is necessary for companies to adopt a futures thinking mindset throughout the development and deployment of new and emerging technologies. This is particularly important during the earlier phases of a TLC, when the impacts of a technology are still relatively unknown. Private actors at this phase can devise and adopt appropriate risk mitigation measures for upstream prevention of human rights harms. This requires that private actors ensure the safety of proposed NETs by ensuring the incorporation of safeguards or guardrails including "emergency brakes" during the design stage.

Private actors must also be cognizant of relevant general or industry standards, including voluntary codes of conduct, and ensure that any NETs they promote have been designed and developed in adherence to such standards. In case of technologies involving automated systems such as AI systems, private actors must also ensure that human control over the technology remains meaningful, even while embracing the considerable upsides of such technologies. Subsequently, companies should also ensure that the technologies they develop are also used by others (subcontractors, clients, consumers, and licensees) in ways that are consistent with their intended use. Private actors can achieve this through both legal (contractual) as well as technological means. Companies should carefully vet potential licensees, including government clients, and take reasonable measures to ensure that such users are not obviously seeking to use the technologies to harm or commit human rights violations. While no actor could ever guarantee that an NET would not be misused by another user, it is reasonable to expect that they conduct basic due diligence to that effect.

Private actors must also make proactive efforts to be transparent about NETs (and their expected impacts on other stakeholders) throughout the various stages of the TLC. Private actors must be willing to openly share their knowledge with all relevant stakeholders seeking to gain a more holistic understanding of the potential human rights impacts of NETs, including those who might potentially be impacted by such technologies. Private actors should proactively seek to close knowledge gaps with respect to NETs and regularly disclose information (to the maximum extent possible under applicable laws and standard business practices) about technologies they develop and deploy.

Private actors must work towards making transparency meaningful by translating relevant information to ensure it is understandable and that it is particularly accessible by those potentially or actually vulnerable to the impacts of a technology. In so doing, they should work with other stakeholders, such as educational institutions and civil society organizations. Transparency efforts are most meaningful when private actors have put in place feedback loops to solicit inputs from relevant stakeholders. Such processes are especially relevant at the earlier and refinement stages of a TLC. In addition to technical and operational aspects, such transparency should also extend to the trail of human-decision making surrounding the development and deployment of a new and emerging technology. Companies must always ensure that someone is designated to answer questions and handle potential complaints regarding the development or deployment of an NET.

Companies must also establish grievance mechanisms that align with the principles outlined in the UNGPs, in addition to other monitoring and oversight measures. This will help companies track the processes associated with the development and deployment of NETs. Such monitoring must be ongoing throughout the TLC, and should always be connected to mechanisms that can alter or (if necessary) completely suspend the function of an NET if monitoring suggests that serious human rights impacts are occurring as a result of an NET. Private entities should be open to the role of other stakeholders, including the State, civil society organizations, educational institutions and other right-holders that also play important roles in those monitoring, oversight and accountability functions.

EDUCATIONAL INSTITUTIONS

Few other institutions in society are structurally as well-positioned to engage with a variety of stakeholders in a range of collaborations and partnerships as are educational institutions. Educational institutions tend to fall between the cracks of the other stakeholder categories. They can be state-run or private, but are generally spaces where independence of thought is prioritized. Educational institutions provide a service (education) that serves the public interest, and yet they cannot typically be classified as civil society. Education institutions are, thus, distinct stakeholders. Moreover, they play a key role under the HRBA@Tech model.

Academic institutions—including but not limited to institutions of higher education—provide optimal environments for multi- disciplinary engagement. It is here where disciplinary gaps between professional cultures are most likely to be bridged. Educational institutions can tackle structural issues related to systemic knowledge gaps. These gaps often occur when different professions lack the specialized terminology and technical understanding to comprehend the expectations set by other stakeholders. If engineering students, for example, never take a human rights class during university, it becomes much more difficult for them to later think about how the technologies they are working on may one day contribute to (or undermine) human rights. Similarly, if law students never get an opportunity to learn about technologies or engineering during university, it can lead them also to lack a holistic understanding of the technologies and issues involved in this discussion.

Educational institutions should make active efforts to break free from disciplinary silos and encourage cross-disciplinary engagement. This can help narrow the knowledge gap between the human rights community and the tech community. In addition, universities can act as social connectors, bringing together students (decision-makers of the future), the private sector, and policy makers. They can also cultivate international networks with peer institutions, engage with government actors and civil society representatives. Educational institutions are optimally suited to facilitate knowledge sharing and fruitful discussions, especially across various stakeholder groups. In service of "learning" and education, such venues also constitute an important catalyst for consensus building. Educational institutions with diverse student populations can ensure that various perspectives are represented in these discussions about human rights and new and emerging technologies.

Interdisciplinary opportunities for research, collaboration, and partnerships are crucial to ensure that ethical and human rights concerns migrate "upstream" into the research and innovation process that gives birth to NETs. As one group of medical ethicists described in the context of 3D bioprinting (promising the 3D printing of human organs), "the way in which science policy and research funding are stimulating biofabrication in general, and specific lines of research within bioprinting in particular is fundamental."[193] These authors described a trend towards Responsible Research Innovation (RRI),[194] which consisted of the alignment of "research and innovation to the values, needs and expectations of society, which requires that all actors, including civil society are responsive to each other and take shared responsibility for the processes and outcomes of research and innovation."[195] Specifically, this "means that scientists and their social sciences and humanities counterparts are working together in research projects or centers to continuously interact and influence each other's thinking and the framing of the new technologies and their applications, while also connecting to societal actors and different users and publics."[196] Educational institutions must adopt such a responsible research approach with respect to NETs and think about how the research might be potentially used (or misused) by various other actors. Educational institutions should consider adopting a precautionary approach vis-à-vis particularly controversial technologies (as many have done with regard to geoengineering), and drawing red-lines around others (for example imposing strict limits on certain types of research, for example genetic engineering on human embryos).

Education institutions can contribute to both the "do no harm" and "make the world a better place" pillars of the HRBA@Tech model through their research initiatives and agendas. Educational institutions produce vital research to inform the development and deployment of new and emerging digital technologies, right from the initial stages of the TLC. Universities are also frequently interdisciplinary places, allowing a built-in mechanisms for multi- stakeholder and multi-disciplinary discussion that is usually absent in other institutions.

193. Vermeulen N, Haddow G, Seymour T, et al. (2017) "3D bioprint me: a socioethical view of bioprinting human organs and tissues." 43(9) Journal of Medical Ethics 618-624

194. Ibid.: Vermeulen et. al describe RRI as having "developed mainly in the context of nanotechnology [but gaining] momentum within synthetic biology."

195. Ibid.

196. Ibid.

Universities are also optimally suited to generate socially beneficial technologies, since many students are highly idealistic and not yet driven by a profit-motive. Faculty and researchers at such educational institutions can provide skillful guidance and research advice to promote a 'human rights by design' mindset among students and young entrepreneurs.

Other stakeholders can partner with educational institutions to design and develop NETs, taking advantage of these dynamics in a mutually beneficial exchange between young students hoping to gain experience and more established institutions hoping to benefit from their energy and passion. This process can lead to technological innovations, but it can also inform new legislative proposals, policies, and processes with respect to these technologies. States and technology corporations can include educational institutions in crucial monitoring and oversight mechanisms, drawing on their independent expertise as a crucial safety guardrail. They can structure co-design strategies where technologists collaborate or partner with education institutions at the very outset to design and develop an NET.

Educational institutions can also play a key role in devising relevant technical standards, benchmarks, metrics, indicators or other tools of measurement to properly assess the human rights impacts of new and emerging technologies. They are also particularly well-situated to contribute to capacity-building in terms of providing education and training to improve digital literacy and skills and can directly engage in capacity-building activities themselves or support various other stakeholders in their capacity-building efforts.

INDIVIDUAL(S)

The final category deals with individuals. In classic political theory, "individuals" are typically equated either to citizens or potential human rights claimants (as "rights-holders").

Under the HRBA@Tech model, a rights-based understanding of an individual's role would only describe half of an individual's role. The other half has to do with an examination of the reasonable responsibilities that individuals also play in the context of new and emerging technologies.

A person, as a user of a particular technology, may be impacted by it (perhaps in ways unknown to them). In such cases they would qualify as a "rights-holder," entitled to seek remedy. That same person may also, however, be a compliance officer at the local technology company, in which case she also has concrete responsibilities, either within the corporate structure as a fiduciary of the company, or as an ethical citizen. Therefore, individuals are endowed with both rights and responsibilities, often at the same time. Moreover, some situations (most situations, in fact) pit individuals with certain inalienable rights in opposition to other individuals with certain responsibilities.

A "responsibility" is notably not the same as a "duty" or an "obligation." Responsibilities are accountability relationships in which one party can and should be held accountable for his or her actions. This idea is related to but distinct from a duty (or an obligation) in which there is a specific legal or moral code that compels certain pre- defined behavior. There are, of course, many duties that flow from responsibilities, and many responsibilities that flow from duties, and yet there can be (and often are) responsibilities that exist without a corresponding legal duty. Responsibilities may merely suggest (but not dictate) appropriate patterns of behavior. Duties, on the other hand, are more specific and can be enforceable.

While some human rights scholars are skeptical of the responsibilities discourse, concerned that it may be deployed as an effort to weaken the international human rights framework,[197] other scholars see the responsibilities discourse as a way to expand and operationalize the toolbox towards the realization and advancement of human rights and improve human dignity and wellbeing.[198]

At the outset, the HRBA@Tech model requires an individual to be conscious of the rights and/or responsibilities they might wield in the context of new and emerging technologies and their impacts. Additionally, it requires individuals as rights-holders to actively assert their rights and claim their entitlements while also exercising responsible behavior. Incentivizing such responsible behavior is particularly vital for individuals who wield decision-making or effective control or power in processes associated with the development and deployment of new and emerging technologies. While this includes a "do no harm" approach to ensure individuals do not contribute to human rights harms, the "make the world a better place" approach would also require individuals play a more positive and proactive role towards taking the necessary actions to live up to that objective.

Human rights actors are more familiar—and perhaps more comfortable—using an exclusively rights-based frame when thinking about NETs. For the HRBA@tech model to succeed, they also crucially need to develop greater fluency with responsibility-based narratives, modes of advocacy, and mobilization strategies. This will be a challenge for all other stakeholders, but will likely prove to be transformative for individuals operating in this space as both rights holders and responsible members of a shared community.

197. Küng, Hans (2005), "Global Ethic and Human Responsibilities," Markkula Center for Applied Ethics, Santa Clara University (Submitted to the High-level Expert Group Meeting on "Human Rights and Human Responsibilities in the Age of Terrorism" 1-2, April 2005, Santa Clara University); and Del Valle, Fernando Berdion and Katheryn Sikkink (2017), "(Re)discovering Duties: Individual Responsibilities in the Age of Rights," 26 Minn. J. Int'l L. 189

198. For a selection of Buddhist writings on this topic, see e.g., Inada, Kenneth (1990) 'A Buddhist Response to the Nature of Human Rights', in Claude Welch Jr. & Virginia Leary (eds), Asian Perspectives on Human Rights, Boulder: Westview Press; Keown, Damien Keown (1995) 'Are There "Human Rights" in Buddhism?' in Damien Keown, Charles S. Prebish, Wayne R. Husted (eds.) 1998 Budd- hism and Human Rights Richmond, U.K.: Curzon Press; and Hershock, Peter (2000), 'Dramatic Intervention: Human Rights from a Buddhist Perspective', 50(1) Philosophy East and West 9

CHAPTER 6: THE HRBA@TECH MODEL AND AI: AN EXAMPLE

Chapter Summary:

Chapters 3 – 5 explored various aspects of the HRBA@tech model. In this final chapter of Section II, we illustrate the HRBA@tech model in the case of a realistic example of an NET. The scenario is that of a government that decides to deploy an AI system to help address the occurrence of a particularly insidious human rights problem. In our case, the human rights problem under discussion is child abuse, but it might just as well be another social issue (such as sexual and gender-based violence (SGBV), human trafficking, elder abuse, etc.). We describe this scenario as though it were a backward- looking narrative of how the government went about designing and deploying that technology. The scenario is fictitious, and yet is also based on numerous real-life examples of precisely such initiatives taking place in a number of countries, and informed by policy makers, entrepreneurs, AI specialists and technologists who assure us that these are realistic and implementable strategies, not unrealistic demands by human rights dreamers. Through this description we show how various elements of the HRBA@tech model fit together to 'nudge' this technology in the direction of human rights.

In 2021, *333 ("Lelandia's" designated emergency hotline) received a call from a mother who claimed that her child, Kate, fell and suddenly "stopped breathing." When the police arrived they found Kate and, after checking her vital signs, pronounced her dead. The police forensic examiners later found multiple bruises on the child's body that had occurred at different times, whereupon the police subsequently arrested the parents. Kate was eight years old when she died. Her height was 110cm and weight 13kg. An average 8-year-old girl's height and weight should be 125cm and 26kg. In addition to her multiple wounds and bruises, Kate had been deprived of food for multiple days prior to her death. Police later discovered that her parents had physically abused her for 'lying' or for having urinary incontinence.

Looking back, there were a few opportunities when external involvement or interventions might have saved Kate's life. In Lelandia, schools are obligated to confirm the safety of children who have unexcused absences lasting three days or more. If the school cannot confirm the safety of the child or his/her whereabouts, they are required to immediately report this to the police. In Kate's case, the parents had notified the school that their son Kurt had a lung problem, and that their daughter Kate had a bone tumor and had therefore requested for their absences to be excused. The parents never provided any medical evidence to support their claims. The parents also declined home visits from school officials who were sent to check on the children's well-being. They offered a range of excuses, such as "the children have gone to their grandparents' house," or "they are too sick to see anyone," among other reasons. Teachers in such situations lack the authority to force their way into a home to verify such claims. As a result, the schoolteachers did not see Kate for an entire year in 2020.

Every time a tragic child abuse fatality comes to light, there is a huge public outcry demanding that we as a society need to do more to protect our children. The public tends to be outraged that children are left unprotected from such heinous criminal abuse. Lelandia's Child Protective Services (CPS) social workers in such situations also face enormous scrutiny and criticism for not having been able to prevent such a tragedy from occurring. This is especially true in cases (like Kate's) where CPS had already previously had contact with the victim, and where popular outrage soon turns into calls for criminal prosecution – a dynamic that is not lost on the already embattled social workers. Such moments often prompt governments to respond by promising institutional reforms and other policy changes. In this case, the government responded to the popular outcry over Kate's death by proposing "Kate's law." This law announced a fast-track program to develop a machine-learning AI system that would identify child protection "blind spots." This AI system, it was proposed, could be used to target resources towards those households most likely (statistically speaking) to be unsafe for children.

The Lelandian government already has experience in developing AI-based systems that address social issues. Specifically, they have implemented a program designed to identify households that may be eligible for social welfare support services, of which they might not be aware. This kind of programing—even though it relies on a potentially invasive analysis of massive amounts of data—has generally been popular with the Lelandian public since it provided at least partial solutions to otherwise very difficult-to-solve social problems.

At the heart of such programs is an analytic procedure for identifying such households called Predictive Risk Modeling (PRM). This process requires the development of models that generate a risk score for households based on the data that has been utilized for the analysis. Individuals are categorized based on their level of risk for specific negative outcomes, such as extreme poverty. Those with the highest risk scores are then chosen for further counseling and intervention, typically conducted by a qualified human case worker. In an ideal situation, those who are the most likely to experience an adverse life situation will be identified before this experience even occurs, or at least during the very early stages of that adverse situation and will be provided with the resources and services necessary to help them resolve or ameliorate the situation.

In Lelandia so far, the performance of such PRM models to identify blind spots has been controversial. Proponents claim that it has been useful for discovering new welfare cases and helping the underserved who may otherwise not have known about applicable welfare policies. Detractors have claimed that there are too many false positives (i.e., system errors where an individual or household is flagged as being "high risk" when in fact they are not). Regardless of such ongoing debates, the Lelandian government is pleased with the progress it has made establishing and developing administrative processes through such e-government initiatives.

Kate's law promised to utilize big data to identify at-risk children and their families. According to the politicians who spearheaded the initiative, such a new system would draw on similar sources of data as existing PRM models. Relevant information would include a family's welfare status, its internal structure (i.e., marital status, divorces, etc.), the employment status of individuals in a household, information about any economic hardships they may have suffered (for example a delayed payment for an insurance premium or a disconnection from the municipal water, electricity or gas system), and public insurance payments. This existing data set can then be supplemented with additional data drawn from the children's school records, childcare records, medical and disability records, and of course any relevant CPS records. Information such as whether a child's immunizations are up-to-date, whether the parents took them to their recommended pediatric development check-ups, whether they have large numbers of unexcused absences from school or childcare, whether the family has applied for a child allowance, and whether CPS was ever called upon to intervene on behalf of a given child will all be analyzed by this proposed system. If a family has ever been reported to the CPS for any reason, information collected during the ensuing CPS investigation, as well as any final decision and interventions made (e.g., separation, criminal charges, case management), would also be included in this system's algorithmic analysis. Combining all this data, it is claimed, would provide relatively comprehensive information about the child and his or her family.

The PRM system also incorporates machine learning processes to improve its accuracy. Feedback data, specifically concerning whether children flagged as "high risk" for resubstantiation actually experience resubstantiation later on, is used to refine the system's analytical criteria over time. As the AI system was being trained on this dataset, certain risk factors began to emerge as playing a particular role in its parameters, including whether a child was living with disability and/or suffered from behavioral issues, whether there had been a history of family violence or alcohol abuse in the household, whether the alleged abuser was living with disability, whether the child had previously lived in a residential facility (orphanage), and whether the alleged perpetrator is unmarried but cohabiting with a significant other. Human rights activists and scholars from the field of social work pointed out that some of these indicators also constituted hallmarks of vulnerability, and warned that the AI system may be inadvertently perpetuating harmful stereotypes against, for example, orphans, or persons living with disability.

What follows is a brief description of how "Kate's law" and the associated PRM was designed and deployed in this hypothetical context, with a particular emphasis on the processes and principles highlighted in this paper to "annotate" that description. It is a hypothetical description, specifically drafted to illustrate some of the core principles and processes of the HRBA@Tech model.

Like any technology, the PRM in this scenario is appropriate for some, but not all, of the 24 processes described in this paper. This is a non-exhaustive description of what we argue can and should be done in such technology development projects.

By way of background, it should be noted that this scenario is taking place in Lelandia, a country that has signed and ratified all nine-core international human rights instruments along with several of the associated optional protocols.[199] Lelandia engages actively with the international community and engages domestically to ensure that human rights discussions are mainstreamed throughout society. In 2021, for example, the Lelandian Government, amended a provision in its Civil Code that made it unambiguous that parents could not claim corporal punishment as a legitimate disciplinary "parenting" method. This reform was the direct result of a large-scale public advocacy campaign spearheaded by Lelandian civil society organizations. The Lelandian government, civil society, and the international community had, in other words, done significant advance legwork to ensure that human rights thinking was mainstreamed throughout society. The concept of 'human rights' is not novel in Lelandia, nor would it come as a surprise to government agencies and private actors that the new PRM should be consistent with international and domestic human rights standards.

1A

1B

1C

The Lelandian government also has a number of standards in place that govern any AI system. Those standards apply not just to private actors but also internally to any government agencies intending to deploy AI systems. These regulations provided important safeguards, many of which are described below and many of which flowed directly from the need to satisfy those regulations and standards.

3C

Turning now to the development of the PRM itself, the government announced the launch of the new PRM on November 20th, which is celebrated as World Children's Day. In a televised speech on this occasion, Lelandia's Prime Minister recalled Kate's tragic passing and announced that the Lelandian government would be launching this innovative system. The Prime Minister clearly stated that the purpose of this new system would not be

1C

6A

199. UN Treaty Body Database, https://tbinternet.ohchr.org/

to cut costs or social workers at CPS, but rather to liberate them to do the parts of their jobs that only humans could do: counselling, consoling, empathizing, and caring.

Paperwork and document management, the Prime Minister argued, should be safely left to machines to do. Much more importantly, the Prime Minister proclaimed, this initiative should help protect the lives of innocent children and also empower socially and economically vulnerable families to be better able to care for their children.

Since AI systems are inherently difficult to understand, the Prime Minister's office also announced a major nationwide effort to explain how AI works, what data would be used, what protections would be in place to prevent data misuse or breaches, and most innovatively a portal where users could raise their concerns about this system and have them answered within 72 hours by a (human) case manager. This hotline, which had multiple access points, also had experts on staff who could explain AI in simple and intuitive ways, trained as both social workers and technology experts.

In subsequent weeks, the Prime Minister's Office created a special task force to study this issue and make recommendations about the PRM's roll-out. This task force was staffed with a well-known industry expert on AI systems, a professor of social work from one of Lelandia's premier universities, two experienced social workers (one from a rural area and one from the capital city), two representatives of civil society (a parent's rights group as well as a child rights non- profit organization), and two representatives from Lelandia's Ministry of Health and Welfare, which employ many of the country's social workers and also oversees the National Health Service (as well as the datasets that the proposed PRM system would draw upon).

The civil society organizations that played a role in this process began an intensive nationwide consultation process, each spearheading the development of so-called "dialogue committees" tasked with soliciting input from their respective constituencies using a variety of methodologies to do so. The child rights non-profit partnered with a prominent local university to help it develop this consultation process, designed specifically to compile a comprehensive assessment of the needs of children in economically and socially vulnerable households.

The task force held regular open-door meetings and sought input from a wide range of stakeholders. They also traveled to various parts of the country to ensure that diverse viewpoints were represented, including those from rural and urban areas as well as minority communities. The task force, in partnership with some of its university contacts, began to develop a range of educational materials about artificial intelligence as well as a "know your rights" briefing tool for families impacted by this new system. They did this even before the system began to operate in order to articulate clearly the rights that families would have, working either through the formal judicial system or informal grievance processes, and the process they could use to correct incorrect (allocative) harms that may or may not have been caused by the AI system. These materials helped to alleviate the concerns of some on the outside of this process who might otherwise have worried that the AI system leaves impacted families with no means of redress.

After an initial round of consultations, the task force concluded that the benefits of such a system outweighed the potential risks and presented a report to the Prime Minister's office of for approval.

Next, the task force created a subcommittee of private sector engineers with expertise in the development of AI systems as well as a group of experienced social workers. The task force was diverse, comprising both men and women, young and old, as well as experts who have experience working with marginalized migrant laborers in Lelandia's rural areas. The rationale behind this composition was to foster collaboration between technologists and social workers. By understanding the social workers' professional reality, the aim was to identify strategies through which smart technology could support their efforts.

Since one of the primary concerns associated with the project was the concern that families would be targeted by a non-transparent and potentially flawed AI system, a separate subcommittee, composed of civil rights lawyers, technologies, and dispute systems designers began to brainstorm strategies to mitigate those harms. They began not with the assumption that a perfect system could be designed, but rather with the assumption that any false-positive identifications of a household as "prone to child abuse" could be perfectly rectified. This subcommittee, therefore, set about designing a remediation process for any failings of the AI system. Further, to ensure that such a remediation system would respond to real-time input, they also designed a series of ongoing impact assessments that would need to be funded by the Ministry of Health and Welfare to accompany the system.

A third and final subcommittee was convened to study the technological safety of such a proposed system. This subcommittee was composed of technologists and data security activists. Their task was to design systems that would account for the known risks of bias in the AI and privacy breaches due to insecure handling of the data. This committee recommended, and later received approval for, a limited beta-test of an advance version of the PRM, where social workers in two similarly situated provinces would "stress-test" the PRM. To test for bias, one of those provinces convened a task force of specially trained (human) social workers to analyze caseloads from that province using only their human skills. The other province beta-tested an advance version of the PRM. The results of the AI PRM and the human specialists were then compared according to a pre- defined list of criteria measuring for bias. Social workers in those provinces were also carefully tracked to see what they were able to accomplish with their extra time (having been 'liberated' from mountains of paperwork by the new AI system), and the impacts of those additional services were carefully monitored. These findings gave rise to new models of how to deploy the considerable talents of social workers given their reduced administrative workload, allowing for new expansions of services in the 'beta' provinces.

Another test involved hiring a crack-crew of professional data security experts whose sole job it was to try to "crack" the data in some way.

With both "stress tests" passed, the team had invaluable inputs on how to strengthen the overall security of the AI system, as well as strategies to make its use more intuitive to non-technologically inclined social workers. These strategies also included so-called "PICNIC" (Problem In Chair, Not In Computer) safeguards. Such PICNIC safeguards may have seemed laugh-

6A 4B 6B 7C 4D 7B 7A 7A 6A 2B 7A 4C 2A 4B 4C 2A 3A 5A 3A 3B

able to tech-savvy engineers, but the stress tests showed how without such safeguards, the PRM risked succumbing to preventable safety breaches.

5A

After the three subcommittees had reported back to the task force on their progress, the Prime Minister's office commissioned the design of the PRM to a private technology contractor. Re-iterating the social welfare objectives of the project, the Prime Minister's office demanded to see evidence that the system would be designed specifically to facilitate the empowerment of socially and economically vulnerable communities, and to be designed to minimize the potential for potential harms flowing from the use of an AI system.

The designers returned with a proposal for an AI system that included not only mechanisms to check the machine-learning parameters themselves, but also robust process recommendations to prevent any representational harm from being done. For example, the designers proposed a model where specialized data teams would be embedded within the Ministry of Health and Welfare who would compile the lists of "households of concern," but that these lists would not be marked specifically as "AI-generated" suggestions when they were forwarded to the local social workers' offices. Individual social workers would thus receive only a notice to visit a certain household, but would not know which cases came to her based on an AI system (as opposed to some other more traditional entry point, including reports by neighbors, a call to a hotline, school reports, etc.). This was intended as a safeguard to prevent social workers from inadvertently placing too much faith in the "inviolability" of an AI-generated suggestion, as opposed to other forms of less-automated leads.

5A

3D

The families, of course, would continue to enjoy the right to know how their household was identified, and would also be given clear and accessible opportunities to contest any determinations they considered to be unjustifiable.

Furthermore, the AI system was programed not to produce a single "score" (as was originally proposed), but rather a "heat map" for each family with specific issues highlighted for the case worker to pay particular attention to. Thus, for example, a case worker would not receive a notation to visit a household because they received a "36" or a "49" on some AI-generated risk assessment, but rather to visit that household because of "financial concerns that may be relevant to look into," all of which would be color coded instead of numerically scored to give a more holistic picture to the individual caseworker. These mechanisms also served to leave the human in control of any decisions relating to individual households, services provided to those households, or any potential further remedies, even while it also greatly facilitated processes that used to consume large percentages of an average social worker's workday.

The designers, informed by comparative best practices and literature reviews of similar models from other countries, discovered that a key human rights consideration was not just how the system functioned, but also the broader implications it had. Whether an alert would trigger a visit by a police officer in tactical riot gear, versus a more casual visit by a community social worker makes a difference, not in the way the AI system is designed, but certainly in the way the AI system is received in the community. Similarly, the way the AI system is configured can significantly influence public perception and impact. For example, an AI model geared towards enforcement and punitive measures may be less well-received compared to another version designed to help families apply for available welfare benefits. The latter approach could foster a more positive community response towards the AI system. The design committee recommended a policy that would prioritize visits by social workers, equipped with a full toolbox of social support services that have been statistically shown to reduce instances of child abuse without breaking apart socially or economically marginalized families.

After another round of public commentary, which attracted a good deal of attention due to the active efforts of government and civil society actors to promote discussions about the system, the Prime Minister's Office announced the gradual roll-out of the PRM, first in the nation's rural areas and then moving into the urban centers. The logic was that in small rural areas the social workers would be far more connected with their communities and could therefore better see any problems that may be associated with the new PRM and raise the alert to correct it.

7A

3A

Civil society groups played an active role in the launch of various initiatives aimed at mobilizing communities, including child rights and parental rights activists. The focus of this mobilization was twofold. First, it aimed to help families harness the new technology to gain access to much-needed social welfare support for which they qualified. Second, it explored ways to use the system's AI-generated metrics to advocate for the restoration of parental rights. For instance, parents who had lost custody of their children due to alcoholism could use the metrics to make a case for regaining custody after successfully completing a specialized sobriety program.

5B

Five years after the successful rollout of the PRM across all of Lelandia, the Lelandian International Cooperation Agency (LICA) began to offer capacity building support to countries in the global south, especially in Southeast Asia and the Pacific Islands, wishing to implement a similar program according to the "Lelandian model."

5D

The above description is fictional, and should certainly not be described as a "best practice." It includes all but two of the processes highlighted in Chapter 3 of this paper.

- **4F: Clearly Identified Responsible Entity,** since it is assumed that the "Lelandian" government (in this scenario) would remain responsible for a PRM that it develops, deploys, and implements, and

- **4E: Incentivization,** since it is assumed that no government will need to be "incentivized" to work towards the protection and promotion of human rights. Furthermore, in our hypothetical, the passage of "Kate's law" was also precipitated by a popular outcry that presumably incentivized the government to take action in light of its own desire to seek democratic legitimacy.

The example shows how intuitive it can be to craft a relatively efficient strategy, drawing on all 24 processes described in this paper, that will collectively serve to 'nudge' a technology – in this case a technology that shares attributes with some of the most heavily critiqued AI systems in the literature (predictive policing models, etc.) – in the direction of being a force for the protection and amplified enjoyment of human rights.

Part III

Conclusion & Recommendations

Recommendation 1:

Moving beyond the 'do no harm' paradigm to guide efforts towards 'making the world a better place'

The international human rights community should work collaboratively with the business and technology communities to ensure that new and emerging technologies do no harm, and moreover that they are also actively hard-wired to 'make the world a better place' (human rights by design).

As central actors in the field of new and emerging technologies, greater attention should be paid to the role of corporate actors in the promotion and protection of human rights. The UN Guiding Principles on Business and Human Rights provide an authoritative and increasingly authoritative framework that must nonetheless be better implemented by both States and business enterprises to prevent and remedy corporate human rights abuses. Nonetheless, it is also important to develop a more positive framing that encourages private actors to wield their positions of influence to improve the world we live in and not simply avoid doing harm.

This is not a call to return to the early days of corporate social responsibility. It is, however, a recognition that we all have a responsibility to make the world a better place, especially when the externalities of our actions can and do make a difference in the world. This is especially true in the case of NETs being developed by States, which are bound by human rights law, or entrepreneurs who make explicit their intention to develop a technology that they claim will 'make the world a better place.'

Many technologists genuinely believe that they are at the vanguard of efforts to 'make the world a better place', and in many instances, can be shown to have done precisely that. Accordingly, the human rights community should refocus some of its energies towards guiding such efforts to meaningfully improve the realization of human rights for all.

Recommendation 2:

Developing a more nuanced analysis of technological contexts to design more holistic intervention strategies

In order to improve the effectiveness of their strategies through targeted and realistic interventions, stakeholders seeking to promote and protect human rights in the context of new and emerging technologies should take a more nuanced approach that considers the stage of maturity of a given technology, as well as the nature (i.e., size and stakeholder type) and underlying objective of the actor developing or deploying that technology.

Any efforts to incite or compel greater respect for human rights should take into consideration the particular characteristics of the stage of the lifecycle of a given technology, as well as the nature of the actor being held accountable. For example, a small or medium enterprise cannot be held to the same standards as a multinational corporation, and the human rights considerations in the design of a new technology may not be the same as those during its manufacturing. Such an approach can facilitate the identification of practical and effective intervention strategies that can result in the intended positive impact on the enjoyment of human rights, without stifling innovation or creating undue market asymmetries.

Recommendation 3:

Developing an ethic of mutual trust and joint learning to deliver on the promise of multi-stakeholder dialogue and action

Stakeholders should recognize the importance of multi-stakeholder dialogue and engage in a constructive process of co-learning and expertise-bridging, anchored in mutual respect for each other's complementary roles and responsibilities.

NETs are extremely complex, but so too is society. The inherent complexity of the societal and technical contexts in which new and emerging technologies are developed and deployed requires a refined understanding of the complementary and mutually-reinforcing roles played by different stakeholders as they jointly address the impacts of NETs on individuals, communities and their rights. Any credible efforts to grapple with these complex interrelationships must be grounded in multi-disciplinary and multi-stakeholder dialogue. To be truly effective, such dialogue must embrace an ethic of collaboration, and be premised on the non-hierarchical and non-exclusionary nature of that discussion. In other words, no one stakeholder in this constellation can "own" the discourse, but also none should be excluded.

To start, such multi-stakeholder dialogue must overcome the trust deficit between States, the private sector, and human rights actors. Old narratives of abusive regulators, uncompromising bureaucrats, amoral technologists, profit-obsessed or opportunistic managers, and never-satisfied human rights activists must be set aside as unhelpful caricatures, even if in some instances there remains some truth to them. To succeed in the HRBA@Tech model, all stakeholders must learn to cultivate an ethic of constructive engagement and consensus-building in addition to performing their traditional roles.

Technologists, policy makers and human rights experts need to build more bridges towards one another and between their respective disciplines. Doing so requires translating the fundamental and universal principles underlying human rights into a language that can be readily understood and applied in the context of business operations, while conversely rendering intelligible the technicalities of law and new and emerging technologies.

By acting as neutral conveners, well-respected 'norm-translators', and specially mandated capacity-builders, international organizations and intergovernmental fora, such as the Human Rights Council, can play an essential role in fostering such an ethic of mutual trust, understanding and collaboration. Educational institutions are equally well-placed to drive multi-disciplinary thinking and understanding of the complex inter-relationships between new and emerging technologies and their impacts on communities, individuals, and human rights.

Recommendation 4:

Learning new justice terminologies

Human rights actors must learn to embrace the discourse of responsibilities and technological ethics, as well as the metrics of scientific inquiry and business development, as part of any holistic strategy to motivate action on new and emerging technologies and human rights. At the same time, there is a need to better explain and convince of the practical added- value of using human rights norms to guide the development and deployment of NETs.

Speaking in broad generalizations, many human rights actors are familiar with justice terminologies steeped in rights language, whereby rights holders demand to have their rights respected. Technologists and business ethicists speak in terms of science and quantifiable outcomes, as well as some operational and ethical principles that govern how certain technologies are managed. Politicians speak in terms of aggregate social welfare, and their duty to live up to popular expectations. Finally, some philosophers (including faith-based philosophers) think in terms of human responsibilities towards one another and perhaps towards fellow sentient beings.

Regardless of one's perspective, it must also be acknowledged that none of these discourses command a monopoly over the 'how' of thinking about the impact of new and emerging technologies. To truly motivate all stakeholders into meaningful action, a combination of all four discourses will need to emerge, mixing an unshakable commitment to human rights principles with new and energizing narratives designed to elicit action.

Hybridizing these various justice languages is complex, and will require openness towards different ways of thinking, and multi- stakeholder collaborations to agree on terms and concepts.

Recommendation 5:

Establishing a new special procedure tasked with the development of a Human Rights–Based Approach to New and Emerging Technologies

The international community should strengthen the capacity of the international human rights system to address the human rights implications of new and emerging technologies and consider the establishment of a newly-minted Special Procedure, either in the form of a thematic Special Rapporteur or, given the breadth of the topic, a better resourced and capacitated Working Group, mandated to address the human rights implications of all new and emerging technologies.

The United Nations, especially its human rights machinery, is uniquely positioned to establish norms in this field. It serves as a hub that brings together diplomatic representatives from around the world, technical experts from diverse disciplines, as well as social activists and civil society actors. This convergence of expertise and perspectives makes the UN an ideal platform for developing a balanced and effective governance framework for emerging technologies.

To date, efforts to develop a unified human rights-based approach to NETs, while comprising an important part of discussions on human rights and NETs, have been rather fragmented and technology-specific.

This may be the inevitable by-product of the thematic mandates of existing human rights mechanisms.

Recommendation 6:

Moving beyond human rights principles to address their real–world procedural applications

The international community, and in particular its human rights system, should not only clarify the application of established human rights norms to new and emerging technologies, but also focus on specific processes that – in the aggregate – will 'nudge' technologies in the direction of the human rights agenda and thereby improve the enjoyment of human rights by everyone, everywhere.

While human rights norms are universally recognized, there is an increasing tendency towards the politicization of the language and logic of human rights. This reality is magnified in the inherently novel space of new and emerging technologies, in which the application of human rights norms is often viewed with skepticism by actors who fear that the language and logic of human rights will be used as a proxy argument to stifle innovation, development and competition.

That being said, NETs still threaten to disrupt established livelihoods, established economies, established ecosystems, established biological systems – in both welcome and unwelcome ways – and almost always in ways where differently situated stakeholders will inevitably disagree about how to weigh those benefits and risks. Moreover, it is an undeniable reality that NETs are being increasingly deployed to repress, censor, harass or surveil.

This dualism of technology is a true universal reality that faces us all, regardless of where we live and where we are situated in life.

There is therefore an urgent need for the development of authoritative guidance on the application of human rights standards to new and emerging technologies. This guidance must not only be articulated in the classic language of human rights, but also harness the spirit and intent of the human rights corpus, oriented (in its essence) towards promoting and protecting both individual rights and human dignity.

To succeed, the authors of this report recommend an exploration by human rights experts of concrete processes designed to "nudge" new and emerging technologies in the direction of the human rights agenda, rather than a renewed effort to identify universal norms that will bridge the inescapable dualism of new technologies. The focus of any potential future Special Procedure, following in the footsteps of the former Special Rapporteur on Business and Human Rights, should be on distilling the hallmarks of "legitimate" processes from less-legitimate or illegitimate "box-checking" exercises.

PERMANENT MISSION OF THE
REPUBLIC OF KOREA IN GENEVA

SNU AI POLICY INITIATIVE

UNIVERSAL RIGHTS GROUP

Not for sale

ISBN 979-11-985703-0-7