WORKING PAPER

AUGUST 2018

Artificial Intelligence: A Rights-Based Blueprint for Business

Paper 3: Implementing Human Rights Due Diligence

BSR
The Business of a Better World

About This Report

This report was written by Dunstan Allison-Hope (Managing Director, BSR) and Mark Hodge, an independent business and human rights expert.

Artificial Intelligence (AI) technologies—and the big data business models underpinning them—are disrupting how we live, interact, work, do business, and govern. The economic, social, and environmental benefits of AI could be significant. But as evidence mounts about potential negative consequences for society and individuals, we urgently need a robust view of what responsible conduct looks like and a vision for how markets and governance mechanisms can guide the right behaviors.

We believe that the speed, complexity, and extensive reach of AI requires an approach to responsible practice that is rights-based. In three papers we draw upon approaches and lessons learned from the field of business and human rights to describe a blueprint for responsible business practice both within and beyond the technology sector.

Deliberate investment in rights-based approaches is urgently needed to avoid two risks: First, that new technologies, capabilities, and business models are unleashed into the world that cause significant harm to the rights to which all human beings are inherently entitled; and second, that a once-in-a-generation opportunity to harness massive advances in technology for the public good is missed.

This is the third of three working papers intended to develop and test new business policies and practices aimed at establishing a sustainable social license to operate for new AI technologies that are capable of creating long-term sustainable value for all stakeholders.

- The first paper outlines 10 beliefs—built on the internationally agreed-upon foundations of the business and human rights field—to govern and guide the use of AI. We draw heavily on the *United Nations Guiding Principles on Business and Human Rights (*UNGPs), the foundational and internationally endorsed road map for addressing business human rights impacts on people.

- The second paper argues for attention to be paid to the AI value chain and demonstrates that the positive and negative human rights impacts associated with AI are directly relevant for companies beyond the technology sector.

- In this third paper we explore what tools, methodologies, and guidance exists or will need to exist to operationalize business respect for human rights in the context of AI development and use.

These three papers have been based on a mixture of desk-based research and direct experience by the authors engaging with business on human rights due diligence. They are positioned as “working papers” to stimulate discussion and influence the ongoing debate. The authors welcome feedback, comment, and dialogue on the papers, and we look forward to working with others to shape the next iteration of these ideas.

Please direct comments or questions to web@bsr.org.

ACKNOWLEDGMENTS

The authors wish to thank Elisabeth Best, Hannah Darnton, Michael Karimian, Michaela Lee, Peter Nestor, Moira Oliver, Jacob Park, and Michael Rohwer for their review, insights, and guidance. Any errors that remain are those of the authors.

DISCLAIMER

BSR publishes occasional papers as a contribution to the understanding of the role of business in society and the trends related to corporate social responsibility and responsible business practices. BSR maintains a policy of not acting as a representative of its membership, nor does it endorse specific policies or standards. The views expressed in this publication are those of its authors and do not necessarily reflect those of BSR members. Working papers contain preliminary research, analysis, findings, and recommendations. They are circulated to stimulate timely discussion and critical feedback and to influence ongoing debate on emerging issues. Most working papers are eventually published in another form, and their content may be revised.

SUGGESTED CITATION

Allison-Hope, Dunstan and Hodge, Mark 2018, “Artificial Intelligence: A Rights-Based Blueprint for Business,” BSR, San Francisco.

Contents

Introduction 4
A description of AI and the need for innovative approaches to human rights due diligence

The Corporate Responsibility to Respect 6
An overview of what implementation of respect for human rights by business entails

Future Testing Human Rights Due Diligence 7
Using methodologies capable of addressing the uncertain impacts of the future, not just the known impacts of today

Addressing Impacts Across the Product and Service Value Chain 9
The role of all actors across the whole value chain in addressing the human rights impacts of AI

Rights-Based Approaches to Opportunities 11
Blending due diligence methodologies with product and service innovation to realize and sustain the positive potential of AI

Human Rights by Design 13
Widening “privacy by design” processes to incorporate a more holistic range of human rights issues

Business Leadership for Collective Action and Public Policy 16
The role of business leaders and multistakeholder collaboration to positively influence the context in which AI is developed and used

Access to Remedy 18
The challenges of providing access to remedy in the context of AI when decisions are difficult to explain

Introduction

Artificial Intelligence (AI) can be defined as intelligence exhibited by machines. It includes both "machine learning" (an approach to achieve AI), which uses algorithms to parse data, learn from it, and then make a determination or prediction, and "deep learning" (a technique for implementing machine learning), which is inspired by understanding the biology of our brains.

AI is advancing rapidly, thanks to ever-more-powerful computing, massive growth in the availability of digital data, and increasingly sophisticated algorithms. These advances bring enormous opportunities to address big social challenges, such as improved health diagnostics, self-driving vehicles that improve road safety, and enhanced fraud prevention, to name just three. AI also brings social risks, including new forms of discrimination arising from algorithmic bias, labor impacts arising from the displacement of workers by machines, increasing the potential of surveillance by employers and the State using tracking devices and facial recognition tools, and new risks to child rights as the volume of data collected about children increases substantially.

In the first paper of this series, we made the case for why a human rights approach offers a strong basis for addressing the social license to operate of these new technologies and put forward 10 beliefs about the relevance of the UN Guiding Principles on Business and Human Rights (UNGPs). In the second paper, we set out the ways in which addressing the potential societal risks of AI are already manifesting themselves as concerns for companies in non-technology industries.

This paper turns to the practical implementation of human rights due diligence in the context of AI. A significant body of best practice has emerged for human rights due diligence of business operations, such as mining sites, manufacturing facilities, and employment relationships. By contrast, an equivalent body of best practice is generally lacking for human rights due diligence of product design, development, and sales—and yet this knowledge and best practice will be essential when undertaking human rights due diligence of AI.

We propose five elements of human rights due diligence implementation that are especially important in the context of AI.

1. **Future testing human rights due diligence**: The UNGPs state that companies should address "actual and potential human rights impacts" and that human rights due diligence should be carried out on an ongoing basis to allow for changes in operating context. Because of the rapid and unpredictable innovations taking place in the field of AI, conducting human rights due diligence can be particularly challenging. For this reason, we propose new human rights due diligence methodologies that can much more effectively address the uncertain impacts of the future, as well as the known impacts of today.

2. **Addressing impacts across the product and service value chain:** In our second paper, we made the case that the responsible use of AI is something that enterprises in diverse sectors along all parts of the value chain need to pay attention to. In this paper we present how the UNGPs offer a clear approach for differentiating the roles and responsibilities of diverse actors along the AI value chain, and emphasize the importance of "know your customer" approaches capable of ongoing due diligence, rather than limited to single moment-in-time transactions.

3. **Rights-based approaches to opportunities:** The UNGPs state that companies should identify, prevent, mitigate, and account for how they address the adverse human rights impacts of their activities, operations, products, and services. This is too often simplified into a "do no harm" ethos, and we propose the deployment of rights-based approaches to AI opportunities, alongside the more conventional identification of risks and adverse impacts.

4. **Human rights by design:** Companies today deploy increasingly sophisticated "privacy by design" processes that integrate privacy considerations during key milestones in product development. We believe there are opportunities to integrate a broader range of human rights considerations—such as nondiscrimination, freedom of expression, and labor rights—into existing processes. These efforts should also draw upon learning from the practice of human rights due diligence in other settings, such as cross-functional collaboration, the integration of rights-holder perspectives, and a focus on vulnerable groups.

5. **Business leadership in collective action and public policy:** As emphasized in our first paper, we need both governance and technical solutions to ensure responsible development and use of AI. As we have seen in many other contexts, such as labor abuses in global supply chains, filling governance gaps is in the interest of the private sector, which otherwise becomes burdened with initiatives that should be the purview of the State. We propose that proactive private sector engagement in collective action and public policy development, including regulatory efforts, is a central part of companies operating responsibly.

By calling out these five aspects of human rights due diligence, our aim is to offer a launch point for companies to develop approaches the are fit for purpose in the context of AI. We believe this can happen through individual companies innovating and communicating openly about their successes and shortfalls; in the context of peer learning exchanges; through industry collaboration; and in existing or new multi-stakeholder partnerships.

In addition to human rights due diligence, we also believe that the rapid development of AI raises three new challenges for securing access to remedy: (1) guaranteeing remedy when violations result from decisions made by machines and algorithms, rather than humans, and as a result are challenging to explain; (2) providing operational grievance mechanisms when there are hundreds of millions of rights-holders and billions of decisions; (3) safeguarding access to remedy when dozens of companies, rather than a single corporate actor, are linked to a human rights violation via the interaction of different AI-based products and services. We propose the need for further research in this important area.

We invite comments and critiques about the content of this paper, and we look forward to exploring how approaches to human rights due diligence can be improved.

The Corporate Responsibility to Respect

In 2011, the United Nations Human Rights Council unanimously endorsed the UNGPs as guidelines for States and companies to prevent and address human rights abuses committed in business operations. The UNGPs contain three pillars—protect, respect, and remedy—and define steps for governments and companies to meet their respective duties and responsibilities to prevent human rights abuses in company operations and provide remedies if such abuses take place.

The corporate responsibility to respect pillar clarifies what is expected of companies and outlines the process for companies to identify their actual and potential adverse human rights impacts and demonstrate that their policies and procedures are adequate to address them. The UNGPs state that companies should prevent, mitigate, and remedy human rights abuses that they cause or contribute to. Companies must also seek to prevent or mitigate any adverse impacts related to their operations, products, or services, even if these impacts have been carried out by business partners.

The responsibility to respect applies to all internationally recognized human rights expressed in the International Bill of Human Rights and the International Labor Organization Declaration on Fundamental Principles and Rights at Work. While the actions businesses need to take to meet the responsibility to respect will depend on their severity, likelihood, and complexity, the responsibility itself applies to all businesses regardless of size, sector, or location.

To meet the responsibility to respect, companies must have the necessary policies and processes in place. The UNGPs identify three components of this responsibility.

- Make a policy commitment to meet the responsibility to respect human rights.
- Undertake ongoing human rights due diligence to identify, prevent, mitigate, and account for human rights impacts.
- Enable remediation for any adverse human rights impacts they cause or contribute to.

Human rights due diligence refers to the process of identifying and addressing the human rights impacts of a company across its operations and products, and throughout its value chain. Human rights due diligence should include assessments of procedures and systems, as well as external engagement with groups potentially affected rights-holders and stakeholders. The UNGPs state that companies should integrate the findings of their human rights due diligence processes into policies and procedures at the appropriate level, with resources and authority assigned accordingly. Companies should evaluate their efforts over time and communicate how they address their human rights impacts.

Since 2011, many industry associations, governments, and civil society organizations have developed resources that provide more detail about how to implement the corporate responsibility to respect, including for the technology industry. Rather than repeat this guidance here, we encourage business leaders and others who are not familiar with the UNGPs to review these. The remainder of this paper proposes five elements of human rights due diligence implementation that are intended to build upon, rather than duplicate or replace, existing guidance. We also consider access to remedy.

Future Testing Human Rights Due Diligence

Futures thinking, also known as strategic foresight, can provide structured ways to explore multiple possible futures of AI deployment, use, and regulation. This can help chart a path forward on human rights that considers various possible outcomes that might occur.

The potential impact of AI on human rights over time is highly uncertain, yet decisions made about AI today can have long-term consequences. For this reason, human rights due diligence methods will need to be capable of addressing rapid change, uncertainty, and complexity. We can't know exactly what path the development and deployment of AI will take, so we should be prepared for different versions of the future and think through the possible long-term implications of today's decisions. This requires human rights due diligence methods capable of informing human rights identification and mitigation strategies that are resilient to a range of different plausible scenarios and that consider potential cascading impacts.

In addition to identifying actual and potential human rights impacts, approaches to human rights due diligence often focus on three impact prioritization criteria centered on the severity of the adverse impact for the rights-holder:

» **Scope**: How many people could be affected by the adverse impact?

» **Scale:** How serious are the adverse impacts for the victim?

» **Remediability:** Will a remedy restore the victim to the same or equivalent position before the harm?

These are sometimes then combined with three further criteria to inform human rights mitigation strategies that companies can deploy:

» **Likelihood:** The level of possibility that the impact will take place in the coming years.

» **Attribution:** Whether the company caused or contributed to the impact through its own activities, or was directly linked to the impact by its operations, products, services, or business relationships.

» **Leverage:** The ability to affect change in the wrongful practices of an entity that causes a harm.

These criteria have proven to be extremely valuable when developing conclusions and mitigation strategies in human rights impact assessments across diverse industries. However, other than considering whether an adverse human rights impact is theoretically possible, these criteria do not necessarily provide enough insight into the uncertain and multiple different versions of the future that may unfold. When deployed in isolation—absent any future testing—these criteria may miss key impacts that occur in the medium and long-term future. Being shortsighted will not work when seeking to implement human rights due diligence for AI and other disruptive technologies.

For this reason, we propose a future-testing methodology based on a structured approach to test human rights mitigation strategies against a range of high-level future scenarios. The key objectives of this methodology would be twofold: first, to test how the severity (i.e., scope, scale, remediability, and

likelihood) of an adverse impact may change over time; second, to test whether the proposed human rights mitigation strategy is resilient against different plausible futures.

These objectives can be met by using high-level future scenarios that broaden the imagination, open our minds to a wider range of future possibilities, challenge assumptions, and identify blind spots. These high-level future scenarios would include uncertainties relating to both developments in AI itself (such as how rapidly AI is deployed) and developments that are exogenous to AI (such as whether we live in times and places of peace or war, or in a world with open or closed societies), but which may impact the way we think about human rights and AI. These scenarios would not be intended to be the human rights due diligence themselves; rather, they would be tools to inform human rights due diligence and to assist in the development of human rights strategies fit for a highly uncertain future.

Futures thinking, strategic foresight, and scenario planning methodologies have been in existence for many years, as have human rights due diligence methodologies. However, to our knowledge, the two methodologies have never been combined since the publication of the UNGPs, and our proposal represents a novel methodology.

Indeed, when piloting this methodology in BSR we have noted several challenges. For example, in a typical scenario planning approach, it is considered important that none of the scenarios be purely utopian or dystopian, and that each scenario be roughly equally plausible and desirable; however, when creating scenarios in a human rights context, it has been very challenging to maintain this discipline. Indeed, in a human rights context, a more normative approach—with some scenarios being preferable to others—might be an appropriate approach to take. The company could deploy a human rights strategy capable of addressing all four scenarios, but which also seeks to use the company's leverage to move the world in the direction of one scenario rather than another.

There are other future-thinking methodologies that could be useful for human rights due diligence. For example, "futures wheels" are a simple but useful tool to map out the possible cascading impacts of an event or development. Starting with a plausible future development (e.g., "driverless cars have become ubiquitous in large cities"), the futures wheel is used to consider potential first order, second order, and third order social, technological, economic, environmental, and political implications of that development. For example, the ubiquitous deployment of autonomous vehicles could lead to the emergence of a "right to drive" movement among older adults who grew up driving and feel it is part of their identity, which could in turn create a rift between older adults and young adults who have never needed a car and blame their elders for many current sustainability challenges, and so forth. Implications of clear relevance to human rights could be introduced, such as arrests being made by driverless cars, the disappearance of locational privacy, and the emergence of zones where driverless cars are unable or not allowed to travel. Although the use of foresight tools such as futures wheels does not enable us to predict the future, it can be a helpful way to anticipate plausible, important, and nonobvious future developments of relevance to human rights.

In the annex we provide incomplete and high-level illustrations of these scenario-based approaches. It should be noted that in real life these scenarios are the subject of extensive development, with each scenario described in much greater detail than possible here. BSR is experimenting with both approaches in real-life human rights due diligence at the time of writing, and will share lessons learned at a later date.

Addressing Impacts Across the Product and Service Value Chain

There is a need for product- and technology-focused human rights impact assessments and a deeper understanding of the policies, processes, and practices that different actors across the value chain can use to mitigate and remediate adverse impacts.

The UNGPs establish two important aspects of business responsibility that are relevant to human rights due diligence in the context of AI. First, companies should identify, assess, and mitigate the actual and potential adverse human rights impacts of their *products and services*, not just sites, factories, farms, and corporate offices. This means that data-sets, algorithms, insights, intelligence, and applications should be subject to proactive human rights due diligence. Second, different actors across the value chain of a given product—such as suppliers, subcontractors, manufacturers, brands, licensees, franchises, retailers, traders, and customers—all have a responsibility to address adverse impacts.

The UNGPs clarify what companies should do based on the degree of their involvement in the impact identified. Specifically, the UNGPs state that where a company *causes or contributes* to an adverse impact, it should cease doing so and use its leverage to seek to mitigate any remaining impact. The UNGPs further state that where a company is *directly linked* to an adverse impact (for example, where its product is being misused or abused by a third party), it should use its leverage to seek to mitigate any remaining impact. Either way, the business should be proactive, including in establishing leverage where it may not normally exist.

In the context of human rights due diligence for AI, we believe that there is a need to develop, share, and spread practice in three areas:

- **Methodologies to assess the actual and potential impacts of AI solutions and products:** Some technology companies are already carrying out impact assessments for emerging products, such as self-driving vehicles or facial recognition technologies. Lessons can also be drawn from product stewardship experience in diverse industries such as health care, pharmaceuticals, chemicals, agriculture, and defense. It will be important to ensure that such impact assessments address far more than the technicalities of a given AI solution and how it is used, but also include the risks to people and society of the business models that underpin them—for example, identifying and engaging with vulnerable populations impacted by the AI solution will be especially important.

- **Integration of human rights due diligence into processes and functions concerned with selecting, managing, ending, and renewing business relationships:** Companies developing and selling AI solutions need to be confident that their private or government customers will not knowingly use the solution being sold to them to violate human rights. While existing “know your customer” processes tend to take place at the moment of the transaction and focus on legal compliance (such as export control restrictions), we believe that developments in AI increase the significance of human rights factors being considered, both at the moment of the sale and

throughout the use phase. Indeed, the UNGPs themselves state that human rights due diligence "*should be ongoing*, recognizing that the human rights risks may change over time as the business enterprise's operations and operating context evolve" (our emphasis).

Illustrations of this approach are emerging. For example, Microsoft's AETHER (AI and Ethics in Engineering and Research) committee[1] has the power to give up new sales and specify what customers can use the company's AI solutions for, while Google's new AI principles specify that the company will not design or deploy AI in several areas, including where product use would "contravene widely accepted principles of international law and human rights."[2] Recent debates about the relationship between AI companies and the defense, military, and national security sectors reinforce the need for clarity on actual and potential adverse human rights impacts arising during the product use phase.

At the same time, the development of "responsible procurement of AI" by companies will become important, especially as part of implementing digital transformation strategies. It is now widely accepted that procurement functions and processes have a critical role to play in addressing the labor and human rights issues in supply chains, but less so in the context of technology. Are my suppliers respecting the privacy of individuals whose data they may be using? Are the data-sets that my suppliers use to train their AI solutions free from bias? How are human rights integrating into (public and private) tenders for AI solutions?

» **Increasing transparency about the AI value chain to involve diverse actors in addressing impacts:** Many industries are taking steps to demystify the complexity of their value chains. Examples include fashion retailers publishing a list of the factories that they source from; food and beverage companies publishing a map of the human rights risks along value chains all the way from agricultural products through to manufacturing to marketing and consumption; electronics companies setting out the length and complexity of the supply chain between their products and so-called conflict minerals; or banks setting out how their own corporate lending activities interface with diverse stages of bringing a commodity such as gold to the marketplace. All these transparency efforts educate the public about the nature and complexity of issues, but they also focus company, industry, multistakeholder, investor, and regulatory efforts on nodes in the system that can drive the most impact in terms of responsible behaviors. Because AI is a relatively young and little understood industry, we believe that part of the journey to address adverse impacts will be to demystify and educate about the structure of the value chain. What is the role of small developers? How does research at universities feed into private sector solutions? What is the role of data brokers? How do investors affect innovation and business models?

[1] https://news.microsoft.com/2018/03/29/satya-nadella-email-to-employees-embracing-our-future-intelligent-cloud-and-intelligent-edge/

[2] https://ai.google/principles

Rights-Based Approaches to Opportunities

Corporate respect for human rights should underpin the positive economic, social, and environmental contribution of AI. This will ensure that the positive potential of new technologies is fully realized, scaled, and sustained.

AI technologies and the big data business models underpinning them are disrupting how we live, interact, work, do business, and govern, and the resulting economic, social, and environmental benefits could be significant. Promises include increased road safety and lower emissions due to an increase in self-driving vehicles; improved in food security and reduction in the use of herbicides thanks to so-called smart agriculture solutions; increased access to education by using AI-based tutoring and learning tools; advances in medicine such as reducing drug costs using AI to model clinical trials and research and development; and improved poverty alleviation strategies based on new data and insights.

Some are beginning to frame the opportunities that AI brings in relation to the achievement of the UN Sustainable Development Goals (SDGs). For example, at a recent UN meeting between governments, technology firms, and innovators convened to discuss "Sustainable Development in the Age of Rapid Technological Change," one speaker listed the ways in which AI can contribute to each of the SDGs. These include:

- **SDG 1 (no poverty):** AI will provide real-time resource allocation through satellite mapping and data analysis of poverty.
- **SDG 2 (zero hunger):** Agriculture productivity will be increased through predictive analysis from imaging with automated drones and from satellites.
- **SDG 3 (good health and well-being):** Preventative health-care programs and diagnostics are improved through AI, leading to new scientific breakthroughs. For example, eight billion mobile devices with smartphone cameras are being used to diagnose heart, eye, and blood disorders.

However, while the SDGs provide an excellent framework for AI innovation, we believe that a human rights lens can also be used to generate similarly positive innovations. The UNGPs state that companies should identify, prevent, mitigate, and account for how they address ***adverse*** human rights impacts (our emphasis), but say nothing of opportunities to promote the realization of human rights. Given the high potential of AI to address major human rights priorities, we propose the deployment of rights-based approaches to AI opportunities, alongside the more conventional identification of risks and adverse impacts.

We believe that the discipline of a human rights due diligence process—especially its inclusive approach and engagement with a wide range of rights-holders and stakeholders—opens the possibility of spotting new opportunities to deploy AI in ways that benefit human rights. Examples of this emerging today include the use of AI to enhance freedom of expression, protect users from hate speech, and address human trafficking. We believe this focus on opportunities can be more deliberately included in a human rights due diligence process. Just as the universe of internationally recognized human rights creates a

"long list" from which companies can systematically identify actual and potential adverse impacts, so the same "long list" and rights-based approach can be used to identify actual and potential positive impacts.

However, many of the first applications of big data and "AI for good" are also being challenged for unintended adverse impacts, such as the use of AI-based recruiting tools preferencing the traits, and therefore the genders and ethnicities, of the existing workforce. In addition, there are inherent privacy risks in business models and AI solutions that depend on the amassing and analyzing of huge amounts of personal data.

We believe that companies should not only address the adverse impacts of business models, products, and services across the value chain, but they should also be sure to conduct human rights due diligence on product and service innovations focused explicitly on doing social and environmental good, even when these are deployed by the public sector or offered as part of philanthropic investments.

As a starting point to this way of working, we believe that organizations—not just companies, but also universities, nonprofits, development agencies, and governments—developing, deploying, and using AI "for good" could begin by exploring some basic questions, such as:

» Have we considered how the product or service could be misused to do harm to people?

» Will the innovation in some way reinforce existing discrimination?

» Are the intended benefits of the innovation going to be accessible to all segments of society?

» If the innovation requires collecting, analyzing, and using large volumes of personal data, what are the privacy risks involved?

» If unintended harms do occur, is there a mechanism for those adversely affected to express grievances and seek remedy for that harm?

» How do we openly communicate about the limits and risks of what the technology can achieve?

Human Rights by Design

Companies today deploy increasingly sophisticated "privacy by design" processes that integrate privacy considerations during key milestones in product development and deployment. We believe there are opportunities to integrate a broader range of human rights impacts into these existing processes.

Privacy by design was established as a standard in systems engineering in the mid-1990s and has since evolved as common industry practice. Principles at the foundation of the practice include: being proactive rather than reactive by anticipating and preventing privacy invasive events before they happen; being embedded into the design and architecture of IT systems and business practices, not bolted on as an add-on, after the fact; and requiring architects and operators to keep the interests of the individual uppermost.

Many of the principles of privacy by design align strongly with the spirit and intent of human rights due diligence, meaning the industry norm is one clear way to mainstream human rights in the development of AI. Our proposition is that the industry should adopt a "human rights by design" standard and set of practices which address a wider set of rights than privacy such as non-discrimination, impacts on mental health, safety of especially vulnerable groups, freedom of expression, and labor rights.

BSR has tested the theory of this approach with several member companies, and it shows promise. Some questions that could be considered during a "human rights by design" process include:

- Were users consulted or involved during the design and testing of the product?
- What human rights risks and opportunities (for example, nondiscrimination; privacy and data security; freedom of expression, association, and assembly; hate speech; access to public services; access to culture; child rights) are potentially relevant for this product?
- How severe are these risks, such as the number of users impacted, the seriousness of impact for the potential victim, and whether a victim could be restored to the same or equivalent position before the harm?
- Do certain markets or user segments present higher risks than others, such as women, children, ethnic minorities, and different nationalities?
- What existing mitigation measures (e.g., policies, controls) are already in place, and what mitigation measures could be added?
- During the use of the product or service, what unintended product use cases may occur?
- Have other companies released similar products or services? What human rights risks and opportunities arose from the use of the product, and how were they mitigated and managed?
- How might risks, opportunities, and mitigation measures vary between different markets and countries?

The notion of identifying technical solutions to risks to human rights is not new. For example, it is not uncommon that when oil and gas facilities are constructed, their original plans can involve pipelines, roads, or installations that disrupt local communities, such as by cutting off routes to school for children or blocking access to sacred sites. In these situations, companies used a human rights due diligence approach to re-engineer alternatives. We are suggesting a similar approach to AI, and we are already seeing efforts by engineers to do this:

» **A research team (including from the Fairness, Accountability, Transparency, and Ethics in AI, or FATE,[3] team at Microsoft Research)** have proposed and prototyped the idea of "datasheets for datasets," which would provide buyers and users of AI information about the data set and create transparency about the likelihood of in-built bias. They note that a data sheet would "focus on when, where, and how the training data was gathered, its recommended use cases, and, in the case of human-centric data sets, information regarding the subjects' demographics and consent as applicable."[4]

» **IBM Research and the MIT Media Lab** have developed an approach to enable algorithms to simultaneously align to user preferences and ethical guidelines set by the user. The researchers have expressed that the current tool is only in the initial phases and using the exact model in situations where the user is setting (or indeed overriding) her or his own ethical preferences will be more complicated. Nonetheless, they plan to explore if the approach can be used to address other issues, such as social media and screen addiction.

We also believe that the invention of new principles and practices for "human rights by design" should be informed by best practices in human rights due diligence in five areas:

» **Cross-functional collaboration:** It is increasingly common for companies to deploy cross-functional approaches to human rights oversight and due diligence activities. This will be necessary in any evolution of "human rights by design" so that legal, procurement, human resources, public affairs, engineering, research and development, and data science are collaborating to find solutions. This can avoid blind spots and increases the likelihood that important human rights risks and mitigation approaches are omitted. Furthermore, many of the adverse human rights impacts highlighted in this series of three papers take place during the product or service use phase, rather than during the manufacturing phase, and for this reason it is essential that legal, sales, and marketing functions are also involved in human rights due diligence. In addition, legal, sales, and marketing functions hold significant influence over factors such as product functionality, data use, product upgrades, warranties, and target customers, all of which shape the human rights risk profile of AI.

» **Integrating rights-holder perspectives:** One of the most profound and important elements of implementing human rights due diligence is the integration of rights-holder perspectives and experiences into the process. In the context of AI, conversations on the topic remain highly specialized and concentrated inside engineering research and development, and data science functions. If AI is to fulfill its potential while mitigating accompanying risks, it is essential that civil society, rights-holders, and vulnerable populations benefit from new channels to participate meaningfully.

[3] https://www.microsoft.com/en-us/research/group/fate/

[4] https://arxiv.org/abs/1803.09010

» **Focus on vulnerable groups:** Professional communities engaged in the development and deployment of AI would benefit from a much deeper understanding of rights-holder perspectives from vulnerable groups who can provide insights into actual and potential adverse human rights impacts and secure access to remedy in practice. Examples include children (or those able to represent their interests) on issues of digital advertising; groups advocating for racial justice on issues of disparate impacts in the financial services industry; refugees, migrants, and trafficking victims on how facial recognition technologies can advance or impair their interests; or international development organizations on how the health-care benefits of AI can be spread more widely.

» **Informed consent:** AI and machine learning rely on data generated by multiple sources, such as personal devices, cameras, business records, and sensors. This omnipresence raises important questions about consent, which today can be granted without full understanding of the implications. Additionally, the changing uses of data, the variable vulnerability of rights-holders, and the differing degrees of personalization of data, all combine to create different levels of human rights risk. This raises the question of whether tiered levels of informed consent are needed for different scenarios, with a higher bar established where there is greater risk of harm, or where the use case involves especially vulnerable populations.

» **Collaboration across value chains:** As described above, a deeper understanding of the policies, processes, and practices that different actors across the value chain can use to mitigate adverse human rights impacts is essential given the impact of AI across the whole value chain and the challenges of understanding who the principle actor is in a violation. This suggests that collaboration in human rights due diligence across a value chain—suppliers, contractors, business partners, and customers—could be an important innovation in a human rights by design approach.

Business Leadership for Collective Action and Public Policy

When considering the challenge of ensuring responsible development and use of AI, governance solutions—not just technical problem-solving—should be embraced and informed by companies. This requires CEOs and other senior leaders to be proactively and thoughtfully involved in developing shared solutions.

The UNGPs describe how companies should use leverage to mitigate adverse impacts to the greatest extent possible and note that one way to increase this leverage is to collaborate with others, including civil society and States. Senior leaders of technology and non-technology companies will increasingly be called upon to establish and use their leverage in the context of collective action efforts and the development of State policy and action.

The systemwide characteristics of the human rights risks and opportunities associated with AI make collaboration with others—both companies and other actors, such as governments, civil society organizations, and professional associations—especially important. For example, ensuring the responsible development and use of new technologies such as autonomous vehicles, facial recognition, and AI-driven recruitment and customer profiling will require new standards and ways of working.

We believe that companies developing and using AI should be prepared to see collective action and engagement with public policy developments as part of their commitment and responsibility to society. This will manifest itself in multiple ways, including:

» **Engagement in multicompany and multistakeholder collaboration**: Companies in diverse industries realize that they cannot eradicate risks to people connected to their business activities alone. Holistic responses often require the development of new standards, learning about new business policies and practices, and new accountability mechanisms. As a starting point, professional bodies and existing industry associations can play an important role. In the wider context of business and human rights, the International Bar Association has made efforts to try and establish how the legal profession should and can support responsible business behavior. Early examples of this in the context of AI are the IEEE Global Initiative on the Ethics of Autonomous and Intelligent Systems[5] and the ITI AI Policy Principles.[6]

  As we have seen in countless other industry and operating contexts, multistakeholder efforts will also be critical. Mining companies have turned to collective action to try and ensure that public and private security services do not abuse local communities peacefully protesting; apparel companies

[5] https://ethicsinaction.ieee.org/

[6] www.itic.org/public-policy/ITIAIPolicyPrinciplesFINAL.pdf

are working together and alongside trade unions, NGOs, and universities to address labor conditions in supply chains; and companies from many diverse industries are involved with civil society groups to eradicate forced labor around the world. Again, we are already seeing efforts in the context of AI such as the Partnership for AI and AI Now. The participation of non-technology companies in existing AI collaborative initiatives, and consideration of human rights and AI in existing industry forums, will be important.

» **Good-faith and thoughtful support of public policy initiatives:** In recent years we have seen business, civil society, and States refocusing on the critical role that governments should play in shaping responsible business conduct. The UNGPs have played a key role in this shift as they focus on the various ways in which States—individually and in unison—can apply a smart mix of measures to address governance gaps. Such measures might involve the use of a myriad of tools available to governments, including regulation, public procurement, corporate reporting requirements, investment contracts, export credit, and development finance.

Business leaders—in particular the C-suite—often have a "seat at the table" when States are discussing policy proposals or the content of new regulations. This presents a clear opportunity to influence how the State realizes its role. Yet, not having a seat at the table does not preclude business leaders from educating policymakers about the responsible deployment of AI, calling for laws and regulations needed to protect rights and ensure responsible business conduct, or speaking out against government proposals that have the opposite effect.

The specific role that companies play in such collaborations and when engaging States can vary, ranging from founding new initiatives to participating in efforts launched by civil society. But when doing so, companies and senior leaders should:

» Make this a board-level and CEO priority so that investments by the company and other stakeholders can lead to meaningful outcomes and change.

» Use the UNGPs as a basis to advocate for clear and practicable standards of what responsible business conduct looks like.

» Ensure that the business is proactively and progressively "getting its house in order" by operating with respect for human rights. This includes ensuring that private corporate lobbying efforts do not contradict or undercut useful and progressive State action.

» Be prepared to share successes and shortfalls to build knowledge about what is effective in delivering better outcomes for people at risk.

» Be prepared to advocate for the same standards and rules in all operating contexts, to create a level global playing field for business and to recognize that all people, everywhere, should be afforded the same rights and protections.

» Be prepared to demonstrate the positive business case for industry leaders and the economy of clear and enforced norms for responsible behavior.

Access to Remedy

Providing access remedy when human rights violations result from difficult-to-explain decisions made by machines and algorithms, rather than humans, requires new research and innovation.

The third pillar of the UNGPs establishes that access to remedy should be provided for victims of business-related abuses. Further, the UNGPs set out a list of effectiveness criteria to judge whether access to remedy is fit for purpose. Among these effectiveness criteria are concepts such as “accessible” (that access to remedy is known to all stakeholder groups for whose use they are intended), “equitable” (that aggrieved parties have reasonable access to information, advice, and expertise), and “transparent” (that all parties are kept informed of progress).

However, the rapid development of AI raises three new challenges for securing access to remedy in ways that meet the UNGPs effectiveness criteria:

» Providing effective access to remedy when violations result from decisions made by machines and algorithms, rather than humans, and as a result are challenging to explain or even beyond the cognitive ability of human beings to understand.

» Providing operational grievance mechanisms when there are hundreds of millions of rights-holders and billions of decisions, and as a result securing access to remedy presents challenges of scale never previously experienced.

» Safeguarding access to remedy when dozens of companies, rather than a single corporate actor, are linked to a human rights violation via the interaction of different AI-based products and services—for example, developers, suppliers, and operators of AI solutions. Understanding who the “principle actor” is in a violation, who is accountable for remedy in the event of a harm, will be challenging—for example, between the creator of the algorithm, the designer of the overall system, or the customer making use of it.

While there may be opportunities to address these issues at the contracting stage, the UNGPs emphasis on leverage and collaborative action suggests that shared responsibility models may emerge. Google’s AI principles helpfully set out several factors relating to Google’s role, such as how closely the solution is related to harmful use, whether the technology in question is unique or generally available, and whether Google is providing general purpose tools or developing custom solutions.

At the time of writing, BSR does not have solutions to these challenges. Instead, we believe there is an urgent need to identify a variety of AI use cases (such as access to credit or employment applications) and create case studies for how access to remedy can be obtained in each case. These case studies would provide an essential and accessible resource for the effective implementation of the UNGPs in the context of AI.

Looking Forward

By calling out these five aspects of human rights due diligence and raising questions that relate to access to remedy, our aim is to offer a launch point for companies and their stakeholders to develop methodologies that progress business practice. We believe this can happen through individual companies stepping up to innovate and communicate openly about their successes and shortfalls; in the context of peer learning exchanges; through industry collaboration; and in existing (or new) multistakeholder partnerships.

Some questions to consider might include:

» What opportunities exist to experiment with strategic foresight and futures methodologies in the context of human rights due diligence?

» Can we more deliberately engage with the whole value chain during human rights due diligence?

» What methods can we deploy to more systematically identify opportunities to promote the realization of human rights through the use of AI?

» Can we pilot “human rights by design” approaches by broadening the focus of existing privacy by design processes?

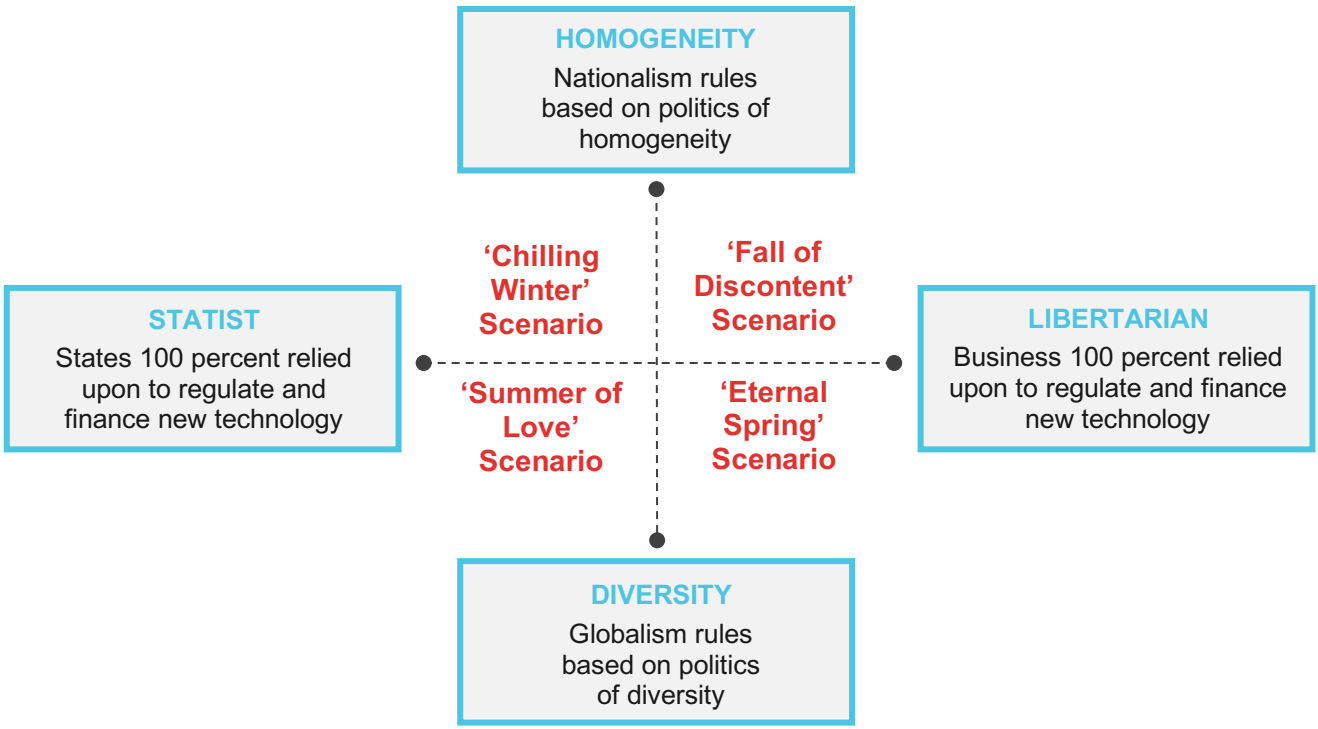
» What opportunities exist for responsible intervention in public policy development, or to increase the awareness of policymakers about the human rights impacts, risks, and opportunities associated with AI?

This series of papers has been developed as “working papers” to stimulate discussion and influence the ongoing debate about the responsible use of AI by business. The authors welcome feedback, comment, and dialogue on the papers, and we look forward to working with others to shape the next iteration of these ideas.

Please direct comments or questions to web@bsr.org.

Annex: High-Level Illustrations of Scenarios

This is an illustration of what a "two-by-two" scenario approach could look like. It should be noted that in real life these scenarios are the subject of extensive development, with each scenario described in greater detail than possible here.



'ETERNAL SPRING' SCENARIO

AI is having positive impact on job creation for everyone of working age by slowly rooting out structured bias that has been blocking work opportunities for minorities for centuries. Based on using AI to conduct highly individualized assessments of citizen health habits, medical histories, purchasing practices, financial status, social context, and personality scores, governments are able to direct tax dollars and social security to those that need it most. AI has also been used to automate 80 percent of all education and health provision in developing countries, allowing for public service provision to a young population that has doubled in the past decade and lives longer. All industries have become more efficient and we are all required to work less, with these efficiencies being channeled into universal basic income schemes that are allocated free from state interference.

This is all happening after decades of self-regulation by business leaders in all sectors voluntarily deciding to embed responsible practices in the design and use of new technology. Robust end-user agreements, when broken, automatically (using smart contract tools inspired by blockchain) stop access

to half the world's data insights and tools, so the costs of abusing data and privacy is too great for any company to bear.

At the same time, the generation of people over 70 (the so-called millennials) are left behind and struggling with endemic mental health problems due to lack of meaning and sense of self-worth. Some regret opting out of participating in how things are done due to a no- forgotten version of offline privacy. Age—not race, sexual preference, disability, or nationality—is a big dividing line in all societies. In addition, there is concern that low-income groups are systematically excluded from the benefits of AI.

'SUMMER OF LOVE' SCENARIO

AI is being used in much the same way as in the Eternal Spring scenario. However, after two decades of failures by the private sector to build and use AI responsibly, governments now strongly oversee all technological development and control the market in AI. In several places around the world—including the United States and the European Union—the major corporations known previously as "tech giants" have been dismantled.

Many companies using data-driven technologies and AI—especially in health care, education, transport, and urban design—are being nationalized or restructured as public-private partnerships in the vein of public utilities. Governments are also experiencing a spike in trust, as they are seen to allocate resources where they are needed most, with a focus on serving poorer segments of society, while being fully transparent to opposition parties and citizens.

Private companies do have a major role to play. Many bid for contracts and several have become successful at doing so, becoming trusted household names and generating revenues at similar levels to those back in 2020. The use of AI for commercial purposes is still common, but privacy violations and failures to respect human rights result in suspensions of legal licenses and major fines. Certain commercial uses of AI come with a "data tax" linked to the volume of data collected and stored by a company, as well as a "computer carbon tax," which strongly disincentivizes the private sector from building and maintaining energy-consuming data centers.

Accumulated government funds are used to finance university research into R&D, basic income provision, technology training centers, and even international technology aid supporting poorer countries to invest in AI. However, corruption is becoming endemic and irresponsible companies are finding ways to avoid fines and tax payments.

'AUTUMN OF DISCONTENT' SCENARIO

AI is being used to deliver public good and support private innovation. However, the most talked-about use of AI is abuse by over 30 governments to serve their ideological goals and nationalist political agendas. Many of the abuses are like those in the Chilling Effects scenario (below), though in this scenario non-State actors from across ideological groups can access the same technologies and use these to wage wars of intimidation and incite physical violence, with actual incidents being reported in three major capital cities and many rural communities around the world.

While some segments of the population are pleased with a focus on their livelihood and well-being, in this scenario business leaders and technology companies are the focus of widespread public discontent. Many note the promises by big technology firms that they could self-regulate and stop their technologies from being uses for nefarious ends. Many in the international community are calling for sanctions to be

placed on any technology that can be abused by States, even where such technology has proven to aid with addressing poverty and enabling whole communities to prepare for climate-change-related natural disasters. Human rights organizations are calling for legal investigations into three technology companies, alleging complicity in a case of mass killings enabled by their AI and drone technologies.

Investment and company valuations are still high. The economics of AI are still working to support growth, in part due to the incredibly high margins secured doing business with government. Business leaders are being pressured to not do business with these same governments—however, responsible companies cancel lucrative contracts, only to find that competitors step in to replace them.

'CHILLING WINTER' SCENARIO

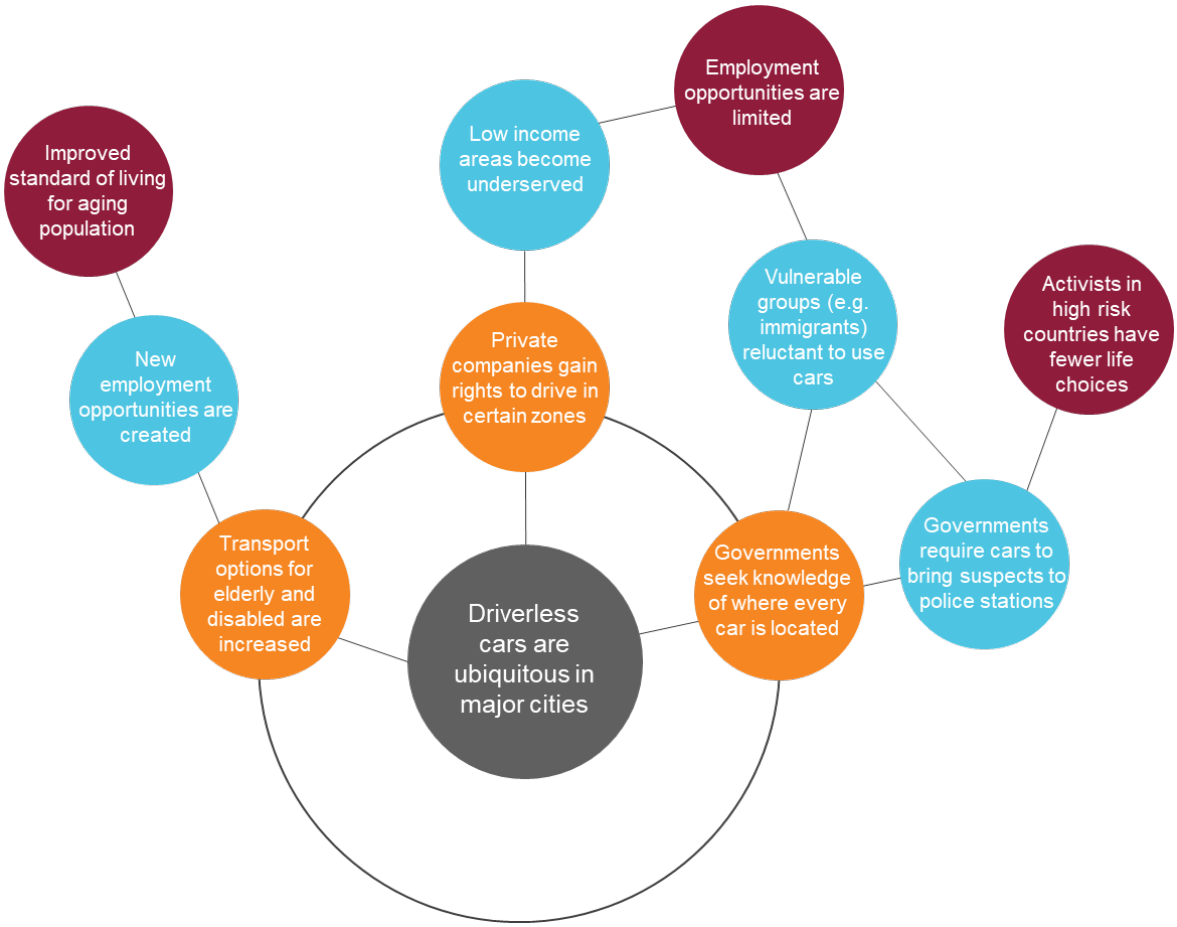
States discover that AI technologies can dramatically increase their ability to regulate who enters and lives in their country. Thanks to AI that can identify all and any illicit or anti-social behaviors of a given applicant, authorities can ensure that even legal immigration processes result in refusing anyone who is not like the indigenous population.

Meanwhile, racial, national, and sexual minorities already living in the country experience a reduced quality of life due to AI-driven biased distribution of social security benefits and access to education, housing, and health care. A new wave of "predictive policing" and constant State surveillance via cameras and sensors allow the State to ostracize community leaders and groups working to protect minorities. The intent is to subtly push these groups to leave and reduce immigration.

At the same time, the majority of citizens and taxpayers across the developed and developing world feel like technology is being used to protect their rights and way of life. It is, in their mind, turning back the ills of the late 20th century faith in economic and political globalization. Slowly but surely, economic well-being is coming back to some local areas left behind.

The State incentivizes technological innovation, via public contracting and tax breaks, that can be used to meet its political ends. Companies seeking to operate responsibly struggle and some are forced to compromise to maintain market share. Working in AI start-ups and for large technology companies becomes seen by many in society as a morally questionable profession, slowly leading to a generation of young people not wanting to join in. The best minds steer clear from AI and innovation is limited.

This is a high-level illustration of what a “futures wheel” scenario approach could look like, showing possible cascading impacts from an event or development—in this case, that driverless cars are ubiquitous in major cities. It should be noted that in real life these scenarios are the subject of extensive development, with each scenario described in greater detail than possible here.



Page left intentionally blank.

About BSR

BSR is a global nonprofit organization that works with its network of more than 250 member companies and other partners to build a just and sustainable world. From its offices in Asia, Europe, and North America, BSR develops sustainable business strategies and solutions through consulting, research, and cross-sector collaboration. Visit www.bsr.org for more information about BSR's 25 years of leadership in sustainability.

www.bsr.org

BSR®