Fundamentals of a Human Rights-Based Approach to Generative AI

Guide 1 of the Responsible AI Practitioner Guides for Taking a Human Rights-Based Approach to Generative AI

February 2025

BSR®

Summary

A human rights-based approach to generative AI (genAI) means embedding respect for human rights into the development and deployment of genAI. This paper provides some foundational background on human rights and describes what a human rights-based approach (HRBA) to genAI means in practice. The components of the HRBA are elaborated upon in greater detail in the rest of the practitioner guides. All readers should start with this paper, as it is important for understanding and utilizing the other guidance papers. Sections include:

1. **Introduction to Human Rights:** Describes what human rights are, how they are defined, and how they are enforced.
2. **Human Rights Responsibilities of Companies:** Describes the human rights expectations for all companies defined in the UN Guiding Principles on Business and Human Rights and the OECD Guidelines on Responsible Business Conduct.
3. **Understanding Ethics vs. Human Rights vs. Trust and Safety-Based Approaches:** Describes the differences and similarities between ethics, trust and safety, and human rights-based approaches to Responsible AI.
4. **Defining a Human Rights-Based Approach to Generative AI:** Describes what it means to take a human rights-based approach to genAI.
5. **Key Resources:** Provides resources for further exploration on human rights.

Key Points

- Human rights are fundamental rights inherent to all human beings. They are defined in international human rights law and overseen by the UN human rights system. They have also been incorporated into regional human rights instruments (such as the American Convention on Human Rights) and national constitutions and laws.
- Companies have a responsibility to respect human rights, which is outlined in the UN Guiding Principles on Business and Human Rights (UNGPs). Core responsibilities for companies include having a policy commit-

ACCOMPANYING RESOURCES

A HRA of the GenerativeAI Value Chain
Overview of the Practitioner Guide
Guide 1: Human Rights Fundamentals
Guide 2: Governance and Management
Guide 3: Impact Assessment
Guide 4: Risk Mitigation
Guide 5: Stakeholder Engagement
Guide 6: Policies and Enforcement
Guide 7: Transparency and Disclosures
Guide 8: Remedy for GenAI Related Harms

ment to respect human rights, conducting human rights due diligence, and providing for or cooperating in remedy when they have caused or contributed to adverse impacts.

- Principles and approaches from the UNGPs are increasingly being integrated into laws and regulations around the world requiring companies to carry out human rights due diligence.
- The Responsible AI field has been largely shaped by ethics framing; with the growth in genAI, trust and safety-based approaches are also becoming increasingly common. Each of these approaches have common ground with human rights.
- Because the international human rights framework provides the most universally accepted standard, a human rights-based approach to Responsible AI in line with the UNGPs can serve as the foundation for assessing and addressing the risks to people and society upon which other approaches can be integrated.
- A human rights-based approach to genAI means embedding respect for human rights into the development and deployment of genAI. This can be done by integrating human rights principles and frameworks into the following Responsible AI processes, each of which corresponds to its own practitioner guide:
  - Governance and management
  - Impact assessment
  - Risk mitigation
  - Stakeholder engagement
  - Policy development and enforcement
  - Transparency and disclosure practices
  - Remedy for harm

1. Introduction To Human Rights

First, it's important to understand what human rights are, how they are defined, and how they are enforced.

- **Human rights are basic rights inherent to all human beings**, regardless of nationality, place of residence, sex, national or ethnic origin, color, religion, language, or any other status.
- Human rights are **universal, inalienable, indivisible, interdependent, and interrelated.**
  - Universal means that all humans are equally entitled to their rights.
  - Inalienable means that rights should not be restricted or taken away except in specific situations and according to due process.
  - Indivisible, interdependent, and interrelated means that all rights cannot be fully enjoyed without the others, and the violation of one right can lead to adverse impacts on other rights.
- The full list of rights is delineated in the following United Nations **international human rights "instruments"** (a collective term that refers to the international declarations, treaties, and other documents that form the basis for human rights), which are focused on state obligations:
  - **The Universal Declaration of Human Rights (UDHR)**, which is the foundational document that lists human rights
  - **The International Covenant on Civil & Political Rights (ICCPR)**, which codifies civil and political rights
  - **The International Covenant on Economic, Social & Cultural Rights (ICESCR)**, which codifies economic, social, and cultural rights
  - **Thematic conventions** related to specific rights areas (e.g., labor rights) or groups of people (e.g., women, children, people with disabilities). For example, the ILO Declaration on Fundamental Principles and Rights at Work, Convention on the Rights of the Child (CRC), Convention on the Elimination of All Forms of Racial Discrimination

(CERD), Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW), and Convention on the Rights of Persons with Disabilities (CRPD).

- The UDHR, ICCPR, and ICESCR are collectively known as the "**International Bill of Human Rights**." The UDHR is a nonbinding declaration and is the authoritative basis for the ICCPR and ICESCR, which are treaties and are therefore legally binding on the countries that have ratified them.

WHAT ABOUT "FUNDAMENTAL RIGHTS"?

The term "fundamental rights" and "human rights" are often used interchangeably. Fundamental rights is the term used for human rights in the European Union Charter of Fundamental Rights, which enshrines human rights for EU citizens into EU law. It applies to EU institutions and the member states when they act within the scope of EU law. It is grounded in the European Convention on Human Rights, which predates the establishment of the EU and applies to all states that are part of the Council of Europe (an international organization set up in 1949 to protect democracy, human rights, and the rule of law in Europe, not to be confused with the European Union).

The EU Digital Services Act and the EU Artificial Intelligence Act both contain references to and requirements related to the protection of fundamental rights as defined in the EU Charter of Fundamental Rights; thus the term "fundamental rights" is becoming more widely known in the Responsible AI field. For the purposes of taking a human rights-based approach to genAI, human rights and fundamental rights are the same thing.

WHAT ABOUT "CIVIL RIGHTS"?

The term "civil rights" or "civil liberties" are also often used interchangeably with human rights, but they have formally different definitions. Civil rights are rights granted by states to their citizens, whereas human rights are universal and apply to all people, regardless of their nationality. No nation may rightfully deprive a person of a human right, but different nations can grant or deny different civil rights. Human rights become civil rights when governments enshrine them in domestic law. For example, in the United States, the human right to freedom of expression as defined by the UDHR and ICCPR is defined and protected by the First Amendment of the Constitution.

- International human rights instruments are overseen by the UN Human Rights System,[1] which consists of a variety of bodies with two main tasks:

  1. **Further developing guidance on international human rights:** Human rights can and should evolve over time as society evolves. For example, the right to internet access and the right to a healthy environment have been declared as new human rights in recent years. Different UN human rights bodies also release updated guidance about what human rights mean in practice. For example, in the past few years there has been a greater focus on defining human rights in the digital age and the responsibility of technology companies to respect human rights.[2]
  2. **Monitoring and protecting human rights:** This includes monitoring whether states are respecting, protecting, and fulfilling the human rights they have committed to in the international instruments they have signed.

- International human rights instruments have also been incorporated into regional human rights instruments that are overseen by their own human rights systems, Including the Council of Europe's European Convention on Human Rights, the Organization of American States' American Convention on Human Rights, and the African Union's African Charter on Human and People's Rights.
- International human rights standards and principles have also been incorporated into national constitutions and laws. For example, Canadian case law requires judges to treat international human rights law as minimum standards for the interpretation of Canadian constitutional rights. Argentina's Constitution incorporates 22 international human rights treaties and conventions, and stipulates that they take precedence over domestic laws and other international treaties.
- Information about what each human right means in practice can be found here.

1 A detailed description of the UN Human Rights System and how it functions can be found here.

2 See, for example, the UN Human Rights Council report on the practical application of the UNGPs to the activities of technology companies and its addendum.

Summary List of Internationally Recognized Human Rights

The following list of human rights is a summarized version of the rights listed in all the international human rights instruments listed above. It includes those most relevant for AI/technology more broadly. Some rights have been combined for simplicity.

Civil and Political Rights

- Right to life, liberty, and security (including both physical and psychological security)
- Freedom from torture, inhuman or degrading treatment
- Freedom from slavery
- Freedom from arbitrary arrest or detention
- Right to a fair trial and the presumption of innocence
- Right to equality and nondiscrimination
- Freedom of movement
- Right to seek asylum
- Right to privacy
- Freedom of expression and opinion
- Freedom of thought
- Freedom of religion
- Access to information
- Freedom of assembly and association
- Right to marry and have a family
- Right to personal property
- Right to political participation (e.g., to vote)

Economic, Social, and Cultural Rights

- Labor rights (e.g., just working conditions, fair wages, right to form unions)
- Right to social security
- Right to an adequate standard of living (e.g., adequate food, housing)
- Right to health
- Right to education
- Right to participate in cultural life
- Right to benefit from scientific advancement
- Right to a healthy environment
- Right to internet access

Thematic Rights

- Children's rights (e.g., to free expression, access to age-appropriate information, freedom from exploitation)
- Disability rights (e.g., accessibility)
- Indigenous people's rights

2. Human Rights Responsibilities of Companies

The international human rights instruments described above were written for governments, not companies.[3] However, over time it became clear that companies also significantly impact human rights, and so in 2011 the UN Guiding Principles on Business and Human Rights (UNGPs) were adopted to outline the human rights responsibilities of companies. These expectations were then also reflected in the OECD Guidelines for Multinational Enterprises on Responsible Business Conduct. The UNGPs and the OECD guidelines collectively form the baseline standards and guidance on business and human rights. Core elements of these standards are described below.

- **The corporate responsibility to respect human rights:** The UNGPs clarify that while states have a duty to protect human rights,[4] companies have a responsibility to respect all internationally recognized human rights. This means that companies should avoid infringing on human rights and should address adverse impacts that they are involved in.[5]
- This responsibility **applies to all companies** regardless of size, sector, or operational context. To meet their responsibility, companies should have policies and processes that are appropriate to their size and circumstances.[6]
- **Core responsibilities for companies include:** 1) having a **policy commitment** to respect human rights,[7] 2) conducting **human rights due diligence**,[8] 3) and providing for or cooperating in remedy for human rights harms when they have caused or contributed to adverse human rights impacts. (See Guide 4: A Human Rights-Based Approach to Risk Mitigation and Guide 8: Remedy for Generative AI Related Harms for more information.)[9]

3 States are the "duty bearers" in international human rights instruments, meaning states are responsible for protecting the rights therein.

4 The state duty to protect human rights is the first pillar of the UNGPs. See Principles 1-10 for more detail on state obligations and their role in the corporate responsibility to respect human rights.

5 The corporate responsibility to respect human rights is the second pillar of the UNGPs. See Principles 11-13 for more detail.

6 See UNGPs Principles 14 and 15 for more detail.

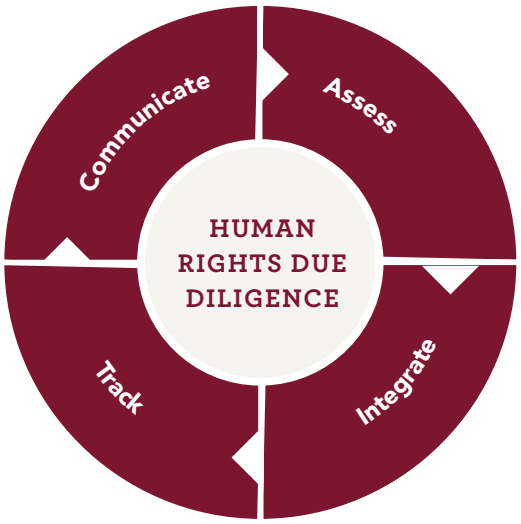
7 See UNGPs Principle 16 for more detail.

8 See UNGPs Principle 17 for more detail.

9 Remedy is the third pillar of the UNGPs. See Principles 22, 25-31 for more detail.

- **Human rights due diligence (HRDD) is an ongoing process** involving four steps:[10]

  1. **Identifying and assessing adverse impacts:** Identifying and assessing actual and potential adverse human rights impacts. In other words, carrying out some form of human rights impact assessment.
  2. **Integrating and taking action:** Integrating and acting upon the findings from impact assessments across relevant internal functions and processes to cease, prevent, or mitigate adverse impacts.
  3. **Tracking:** Tracking the effectiveness of measures taken to address identified impacts.
  4. **Communication:** Communicating publicly about how adverse human rights impacts are addressed.



- **Stakeholder engagement:** The UNGPs direct companies to carry out stakeholder engagement with potentially affected stakeholders (i.e., rightsholders) to inform all four elements of HRDD.[11]
- **Emphasis on vulnerable groups:** The UNGPs also direct companies to pay particular attention to the rights, needs, and challenges of those from groups or populations that are vulnerable or marginalized or at risk of becoming so.[12]
- **Issues of context:** The UNGPs recognize that context may affect a company's human rights risks and ability to address them. For example, in some cases domestic context, such as restrictive local laws, may make it impossible for companies to fully meet their responsibility to respect human rights. In these cases, companies are expected to respect human rights to the greatest extent possible under the circumstances. The UNGPs also recognize that certain operating contexts, such as conflict-affected areas, may increase the likelihood a company is associated with certain adverse human rights impacts, and should therefore involve enhanced HRDD.[13]

The UNGPs are now considered a global standard for companies, expected and inquired about by investors, policymakers, and other external stakeholders. Although the UNGPs themselves are voluntary, the principles and approaches described in the UNGPs are increasingly being integrated into laws and regulations around the world, such as the EU Corporate Sustainability Due Diligence Directive, the EU Corporate Sustainability Reporting Directive, and technology-focused regulation such as the EU Digital Services Act and the

10 More information about the HRDD Steps can be found in Principles 17-21 of the UNGPs and in the OECD Due Diligence Guidance for Responsible Business Conduct.

11 See Principle 18 of the UNGPs for more information.

12 See the "general principles" section of the UNGPs for more information.

13 See Principle 23 of the UNGPs for more information.

EU Artificial Intelligence Act. They are also part of multi-company and multi-stakeholder organization codes or principles, such as the Global Network Initiative (GNI) and the Global Internet Forum to Counter Terrorism (GIFCT).

Because the UNGPs apply to companies in all industries, it can be challenging to figure out exactly what they mean for technology products and services, including those powered by AI. The UN therefore created the B-Tech Project in 2019 to provide authoritative guidance and resources for implementing the UNGPs in the technology space. More recently, B-Tech has been focused on developing guidance about how the UNGPs apply to genAI and has released the following papers, which BSR assisted with:

- Advancing Responsible Development and Deployment of Generative AI
- Taxonomy of Generative AI Human Rights Harms
- Overview of Human Rights and Responsible AI Company Practices

Additionally, the OECD is releasing Due Diligence Guidance for Responsible AI for companies involved in the development and use of AI. Although the guidance has not yet been published as of the time of this writing, the guidance will build upon the OECD Due Diligence Guidance for Responsible Business Conduct and will highlight where similar guidance exists in other AI risk management frameworks to support companies navigating requirements across multiple jurisdictions.

3. Understanding Ethics Vs. Human Rights Vs. Trust and Safety-Based Approaches

The Responsible AI field has largely been shaped by ethics framing centered around a set of high-level principles—such as fairness, accountability, transparency, explainability, and human oversight. **These principles are generally compatible with human rights**. For example:

- Fairness maps closely with the right to equality and nondiscrimination
- Explainability, transparency, and accountability map closely with the right to access information, freedom of opinion, and the right to remedy.
- The principles of human autonomy and control over AI are not themselves human rights, but they are foundational concepts that underpin human rights more broadly.

Many ethical principles are also core to a human rights-based approach. A human rights-based approach (HRBA) is a conceptual framework that puts human rights at the center of decision-making, and focuses on strengthening protection and respect for human rights by "duty-bearers," such as states or companies, as well as empowering rightsholders to exercise their rights. A HRBA is underpinned by five key human rights principles:

- **Participation** —Everyone is entitled to active participation in decision-making processes that affect the enjoyment of their rights.
- **Accountability** —Duty-bearers are held accountable for failing to fulfil their obligations toward rightsholders. There should be effective remedies in place when human rights breaches occur.
- **Nondiscrimination and equality** —All individuals are entitled to their rights without discrimination of any kind. All types of discrimination should be prohibited, prevented and eliminated.
- **Empowerment** —Everyone is entitled to claim and exercise their rights. Individuals and communities need to understand their rights and participate in the development of policies that affect their lives.

- **Legality** —Approaches should be in line with the legal rights set out in domestic and international laws.

A rights-based approach is considered by some to be one approach to ethics, and is grounded in the idea that humans have the basic freedom to do what they choose with their lives and that freedom should be respected.

However, because there is no universal ethics framework or standards in the same way that there is a universal human rights framework and standards, **ethics-based approaches can vary significantly in both substance and outcome**. Further discussion of the interplay between ethics and human rights-based approaches to Responsible AI can be found in this paper.

Ethics-Based Approaches	Human Rights-Based Approaches
A framework for decision-making in situations where right and wrong, good and bad, are not clearly defined	Considered to be an ethics "school of thought" or "approach"
Often seek to address issues of fairness and social justice	Utilize an internationally defined framework for defining both government and company responsibilities
Have different "schools of thought" and standards that support different approaches and choices	Establish a list of rights that should always be protected / respected regardless of context or culture
Different traditions, cultures, countries, and religions may choose different outcomes	Focus on the experiences of the most vulnerable
	Establish a floor rather than a ceiling—e.g., respect for human rights is a minimum requirement

With the growth of genAI, trust and safety-based approaches to Responsible AI are becoming increasingly prevalent. Trust and safety professionals who previously worked at online platforms are increasingly being hired by genAI companies to help operationalize ethics and safety efforts.

- Trust and safety-based approaches **focus primarily on the practical application of ensuring genAI tools are "safe,"** and are bringing approaches and lessons learned from online platform content governance.

- Trust and safety-based approaches tend to be anchored in pre-established taxonomies of harm **based on how a tool may be misused or abused**.
- As with ethics, **many trust and safety goals are also closely aligned with human rights.** For example, efforts to prevent genAI tools from producing hate speech or instructions for fabricating weapons addresses risks to the right to nondiscrimination and the right to life, liberty, and security.

Ethics, trust and safety, and human rights-based approaches are not mutually exclusive and need not be in conflict. However, the responsibility to protect and respect human rights is a universal minimum standard for all companies and governments, and the international human rights framework provides the most universally accepted approach. This means **a human rights-based approach in line with the UNGPs can serve as the foundation for assessing and addressing the risks to people and society associated with AI upon which other approaches can be integrated**.

For example, international human rights provide a comprehensive foundation for identifying impacts to people in society. Using all human rights as a foundation for identifying how people and society might be impacted by an AI product often results in a more complete list of potential impacts, including many that may be missed by an ethics or trust and safety-based approach. However, ethics and trust and safety-based approaches can explore a wider variety of issues and decisions that may not be directly addressed by the human rights framework, such as the role of technology in society, the nature of relationships between humans and technology, or whether a certain use of technology is "ethical" or "desirable" outside of simply respecting human rights.

AI SAFETY AND HUMAN RIGHTS

The term "AI safety" is sometimes used interchangeably with Responsible AI—in fact some companies call their Responsible AI teams "AI Safety." However, AI safety also refers to a field of study focused on anticipating, preventing, or preparing for long-term, existential, and societal-level risks associated with AI, such as the risk of human extinction due to misalignment of superintelligent AI systems with human interests. Although human rights can also be utilized as a framework with which to analyze and consider trade-offs related to long-term and existential risks, these guides are focused on helping Responsible AI practitioners address the AI risks of today and the near future.

4. Defining a Human Rights-Based Approach To Generative AI

A HRBA to genAI means embedding respect for human rights into the development and deployment of genAI that is grounded in the five principles of a HRBA described above—participation, accountability, nondiscrimination and equality, empowerment, and legality. For Responsible AI practitioners, that means implementing the guidance laid out in the UNGPs to do the following, each of which is described in more depth in its own guide:

Integrate guidance for human rights governance into governance of genAI. Guidance on human rights governance from the UNGPs and OECD guidelines can be applied to the governance of genAI. This includes incorporating human rights into AI principles, enabling collaboration across Responsible AI and human rights teams, ensuring Responsible AI teams have human rights expertise, and embedding Responsible AI governance throughout the organization. See Guide 2: A Human Rights-Based Approach to Governance and Management.

Assess the human rights impacts of genAI. When carrying out impact assessments for genAI, practitioners can utilize human rights as their starting point—identifying and assessing risks to all internationally recognized human rights, not just a predetermined list of risks or harms. Human rights assessment methodology can also be used to help assess and prioritize risks. See Guide 3: A Human Rights-Based Approach to Impact Assessment.

Incorporate human rights principles into risk mitigation. When identifying risk mitigations, Responsible AI practitioners can utilize relevant human rights guidance. This includes not offsetting risks against benefits, addressing new risks to human rights that might arise from implementing a mitigation, and determining how to address situations where rights are in tension with each other. See Guide 4: A Human Rights-Based Approach to Risk Mitigation.

Carry out meaningful and effective stakeholder engagement. Entities across the genAI value chain should engage with stakeholders to inform their understanding of risks and how to address them. This engagement should be conducted in line with established best practices. See Guide 5: Conducting Stakeholder Engagement.

Incorporate human rights principles into product policy development and enforcement. This includes ensuring policies reference human rights when relevant and address risks to all human rights, ensuring policies and enforcement approaches adhere to the human rights principles, and engaging with stakeholders to inform policy development and enforcement, among others. See Guide 6: A Human Rights-Based Approach to Policies and Enforcement.

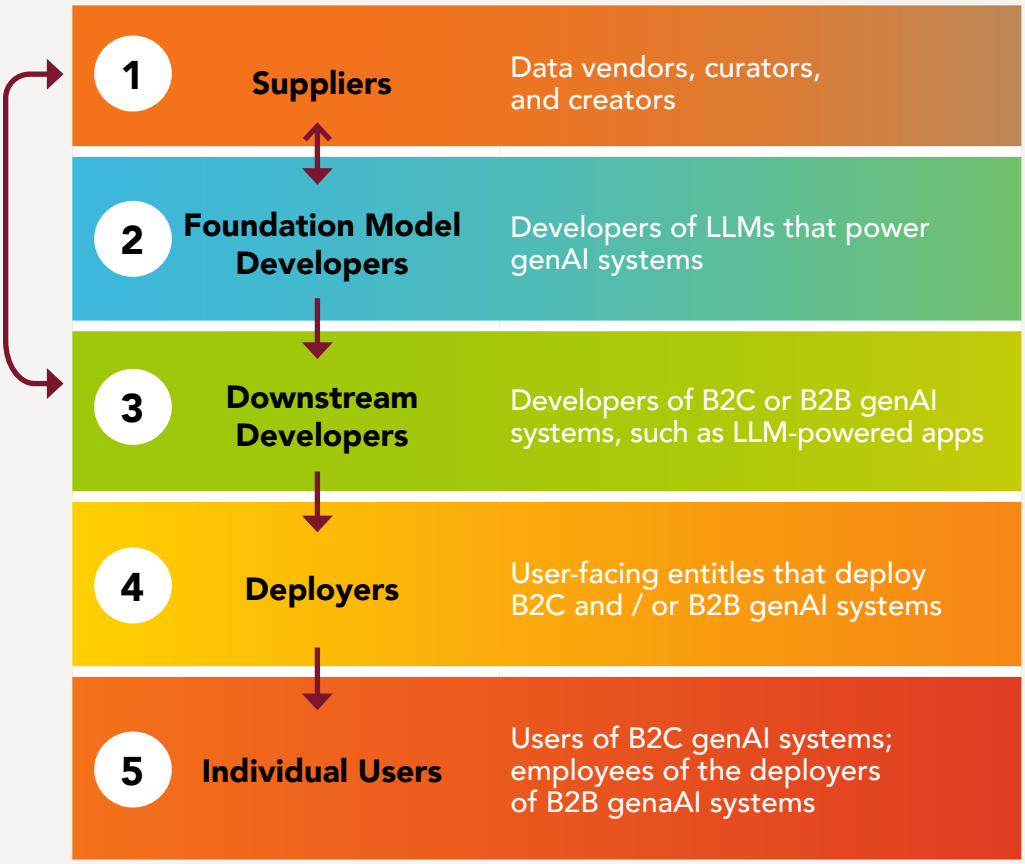
Align transparency and disclosure practices with human rights responsibilities. Entities across the genAI value chain can ground their disclosures and transparency efforts in best practices for human rights and sustainability disclosure that are already well established. Disclosures should focus on risks to people that are the most salient, in accordance with the UNGP criteria of severity and likelihood. See Guide 7: Aligning Transparency and Disclosure Practices with Human Rights Responsibilities.

Provide, or cooperate in providing, remedy for harms. This includes cooperation in the broader "remedy ecosystem" of businesses in the genAI value chain, nongovernmental organizations, and public entities. Where direct remediation by businesses is required, and that remedy requires collaboration across the genAI value chain, identify that relevant point of contact for coordinating the remedy. See Guide 8: Remedy for Generative AI Related Harms.

ONGOING HUMAN RIGHTS DUE DILIGENCE

The various components of integrating human rights into Responsible AI should be carried out on an ongoing basis. This is part of the ongoing human rights due diligence required by the UNGPs and the OECD Due Diligence Guidelines, based on the notion that because technologies, use cases, and human rights context changes over time, HRDD cannot just be conducted at one moment in time, but rather should be considered an ongoing process. For example, while a standalone risk or impact assessment is a helpful foundation for identifying and understanding risks, there should be a process for identifying new or evolving risks. Similarly, risk mitigations need to be continually evaluated to ensure they are effective and to identify when new mitigations are needed to respond to emerging risks.

What a human rights-based approach to genAI looks like in practice will vary according to where a given entity falls within the value chain. Each of the practitioner guides linked above describe the specifics for each component of the genAI value chain where relevant, a summary of which can be found below. For further explanation of the genAI value chain, see Section 4 of the HRA.



Suppliers of genAI system components, such as hardware or infrastructure, are not in scope for this HRA.

Some entities straddle multiple categories. For instance, OpenAI is a supplier, a foundation model developer, and a deployer

The double arrows between foundation model / downstream developers and suppliers indicates that these actors work closely, particularly during the training and evaluation process. Both types of developers may go back to suppliers multiple times for additional data to satisfy data needs for model fine-tuning.

Other actors, such as regulators or investors, may be "value-chain adjacent." This means that their activities have impacts across the value chain. They are excluded from this image for simplicity.

5. Key Resources

Practitioners can consult the following key resources on human rights responsibilities of companies both generally and related to the development and deployment of technology for additional detail not provided in the guide.

- **UN Guiding Principles on Business and Human Rights** (UNGPs): A series of principles that set out the human rights responsibilities of companies.
- **UNGPs Interpretive Guide**: Provides guidance on how to interpret the UNGPs, alongside examples.
- **OECD Guidelines for Multinational Enterprises on Responsible Business Conduct**: OECD standards for all companies related to "responsible business conduct."
- **OECD Due Diligence Guidance for Responsible Business Conduct**: A subcomponent of the OECD guidelines that describes how companies should carry out due diligence. These guidelines are in line with the UNGPs.
- **OECD Due Diligence Guidance for Responsible AI** (forthcoming): Builds on OECD Due Diligence Guidance for Responsible Business Conduct to provide guidance for companies developing and using AI.
- **The UN B-Tech Project**: A UN initiative to help interpret the UNGPs for the technology sector, related to the development and deployment of technology.

BSR™ is an organization of sustainable business experts that works with its global network of the world's leading companies to build a just and sustainable world. With offices in Asia, Europe, and North America, BSR™ provides insight, advice, and collaborative initiatives to help you see a changing world more clearly, create long-term business value, and scale impact.

www.bsr.org

Copyright © 2025 by Business for Social Responsibility (BSR)
All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the publisher, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law.