Augmenting Ethical AI:

2023 Progress Report on the Collective Impact Coalition for Digital Inclusion

September 2023

Table of contents

Executive summary

In September 2022, the World Benchmarking Alliance (WBA) rallied a group of 33 investors and 12 civil society groups to launch the Collective Impact Coalition for Digital Inclusion – a coordinated engagement campaign aiming to push technology companies to advance ethical AI policies and practices. The campaign, commonly referred to as the Digital CIC, builds on the findings of WBA's Digital Inclusion Benchmark (DIB), which has revealed large transparency gaps in companies' disclosures on ethical AI.

This report charts the progress of the campaign on its first anniversary, distills common trends in companies' responses to engagement by investors under the campaign, and provides new, preliminary data on the state of the tech industry's commitments to ethical AI.

The report finds that:

- **Most companies are receptive to investor outreach on ethical AI, but a smaller group remain unresponsive or push back.** Out of the 44 companies that investors have attempted to engage since the inauguration of the Digital CIC, 28 – about two thirds – have responded to their collective outreach. 11 others remained unresponsive or explicitly declined to join a dialogue, with several arguing that the risks of AI were either not material or not applicable to their business.

- **Governance and oversight mechanisms specific to AI remain poorly explained and understood.** While companies have activated a range of board-level committees and subcommittees with a full or partial focus on ethical AI, the role of teams below the executive tier in designing the policies and practices of ethical AI is usually much less clear, as is their ability to escalate concerns to the board. External advisory committees with broad representation are rare.

- **Companies are not clear about how they monitor the AI tools they buy and sell.** Many companies purchase or subscribe to third-party AI tools rather than developing them in-house. The spread of generative AI models and their rapid incorporation into diverse products has further fueled this trend. But the engagements under the Digital CIC suggest that neither buyers nor vendors clearly and consistently outline what due diligence and monitoring they apply when they make their models and tools available to third parties or when they are the ones procuring such tools.

- **Very few companies publish consolidated AI principles, but progress on this front is accelerating.** Since the launch of the Digital CIC, 19 of the 200 companies evaluated in the 2023 Digital Inclusion Benchmark have announced their inaugural AI principles. CIC members had led outreach to all 19 throughout the year. As of September 2023, 52 companies – more than a quarter of those we evaluate – have a publicly available set of principles, up from 33 in the 2021 DIB. Several of these companies have been evaluated across three editions on the benchmark since 2020, which underscores the value of collective engagement.

- **New data reveals additional blind spots in ethical AI disclosures, most notably implementation and impact assessment.** Anchored in DIB's preliminary vision for a standalone ethical AI indicator, these findings signal the challenges ahead and the key shortcomings companies need to address. Out of the 52 companies with ethical AI principles, only six show evidence of considering AI-related impacts as part of structured human rights impact assessments. Eight companies clearly demonstrate how they operationalise their principles. Only three fulfilled all of the expectations set out in our preliminary ethical AI indicator.

- **More than half of companies' AI principles reference international human rights standards.** In total, 31 out of the 52 companies explicitly cite the protection of human rights as a component of their ethical AI principles. These companies' choice to ground their principles in a universally recognised framework sets them apart from their peers, offering a pillar of stability in a fluid and highly disjointed ecosystem of standards.

These findings strongly suggest that collective engagement on ethical AI needs to expand and evolve, applying additional layers of scrutiny to high-risk use cases, bringing in more underrepresented perspectives, and increasing coordination among initiatives with similar goals. The emergence of binding AI regulations in the European Union, voluntary commitments in the U.S., and numerous national-level regulatory efforts worldwide supports this need for greater alignment on what constitutes good corporate practice and how broadly it should be applied. These factors will chart the course for the next phase of the Digital CIC.

1. The Digital Inclusion Benchmark

The World Benchmarking Alliance (WBA) aims to build accountability for business performance on the United Nations' Sustainable Development Goals (SDGs). WBA publishes free and publicly available benchmarks across seven systems transformations and works with allies to drive systemic change to achieve the SDGs.

WBA's Digital Inclusion Benchmark (DIB) provides guidance to technology companies on what is expected from them in the context of the UN SDGs and human rights. It ultimately equips investors, governments, civil society, and individuals to engage with these companies. We create a system that recognises leadership and creates accountability for companies that lag behind.

The DIB ranks and scores 200 of the world's most influential hardware, telecoms, software and IT service companies on their contribution to four measurement areas: enhancing universal **access** to digital technology, improving school connectivity and all levels of digital **skills**, fostering safe **use** and respect for digital rights, and practicing open, ethical and inclusive **innovation**. Since 2022, we have also assessed these companies' performance on a set of Core Social Indicators that apply to companies across the industry spectrum. These include companies' efforts to conduct human rights due diligence, advance gender equality, provide decent work and living wages.

2. The Collective Impact Coalition for Digital Inclusion

WBA's mission is to build a movement. Our strength and success are directly reflected in the collective impact of our global, multi-stakeholder Alliance. The Alliance comprises more than 380 organisations aligned in pushing for systemic transformations that will help achieve the SDGs. Our Alliance also forms the bedrock of our Collective Impact Coalitions (CICs).

The Alliance was created to strengthen partnerships and coordination across the broad community of actors seeking to ensure that corporations improve their practices and to hold them accountable when they fall short. WBA's Allies include research institutions, benchmarks and standard setters, industry platforms, civil society organisations, consultancies, financial institutions, and government or multinational institutions. Our benchmarks continuously provide Allies with insights on key sustainability challenges and opportunities, flagging companies that embody good practice, calling attention to those that fall behind their peers, and opening avenues for collective pressure. This structured fusion of data and action has proven to be a productive framework pushing companies to engage and improve.

In 2021, WBA launched the **Collective Impact Coalitions (CICs)**, aiming to provide a space for Allies to take forward cross-sector, collaborative action based on data and evidence provided by WBA benchmarks. Where our benchmarks provide the evidence, CICs offer a framework to bring together Allies around key issues that are catalytic to systemic change. Their key functions are to:

1. *Prioritise*: Identify key issues based on WBA research & engagement;

2. *Mobilise*: Convene Allies, set ambition, determine roles & responsibilities;

3. *Execute*: Take action, partner with leaders, hold laggards to account.

The **Collective Impact Coalition for Digital Inclusion (Digital CIC)** was launched in September 2022, bringing together a varied group of stakeholders to focus on driving measurable progress on corporate commitments to ethical Artificial Intelligence (AI) and human rights principles. At the outset, the coalition set it sights on technology companies that were found to be lacking any such commitment in the 2021 Digital Inclusion Benchmark (DIB).

Why ethical AI? Automated systems, ubiquitous but often unnoticed, are already playing an enormous role in billions of people's lives. This is particularly evident in domains such as finance, health, media and entertainment, advertising, law enforcement, and human capital management. For example, we have seen that AI can help to improve accurate medical diagnosis or broaden financial inclusion. At the same time, AI may increase the risks of potential harms such as bias and discrimination; invasions of privacy; denial of individual rights; and non-transparent, unexplainable, unsafe outcomes. If principles of ethical AI are not well-considered, individual companies face reputational as well as revenue losses, and society as a whole faces tremendous risk.

"The launch of the Digital CIC Progress Report presents an opportunity to reflect on the progress we have made in advancing commitment to ethical AI principles. We have been encouraged by the response to investor outreach on ethical AI. However, as technology becomes more sophisticated, we are committed to driving progress at both the laggards and those currently considered leaders. Close collaboration between investors, civil society and corporates on this issue will allow us to build on the success to date and move towards just and sustainable technology uses."

Emilie Goodall
Head of Stewardship – Europe
Fidelity International

Ethical AI is a critical area of digital inclusion that requires systemic change. The 2021 DIB found that very few companies had public commitments to responsible and ethical AI, thus failing to meet one of the most fundamental high-level expectations that can be applied to AI. A basic commitment to ethical AI principles can serve as a gateway to building trust with users and to reducing risks and harms to individuals, societies, and companies themselves.

With this in mind, the **objectives of the Digital CIC** have been threefold:

1. to **raise awareness** on the importance of responsible and ethical AI;

2. to **increase understanding** of the state of play and leading best practices;

3. to **Improve digital technology companies' commitments** to ethical AI.

The Digital CIC's investor members, representing over 6.9 trillion USD in assets, initiated a collaborative engagement to drive the adoption of ethical AI and human rights principles among the

companies in their portfolio. The investor component of the initiative is anchored in the Investor Statement on Ethical AI (April 2022), while the Joint Statement on the Responsible and Ethically Beneficial Design, Development and Use of AI (September 2022) serves as the bedrock of its civil society arm, which include non-governmental organizations, research institutions, think tanks, and business consulting groups.

The Digital CIC currently comprises **34 institutional investors** and **12 civil society groups**, including:

	Investors	Civil society groups
1	**Fidelity International (co-lead)**	**Women at the Table (lead)**
2	**Boston Common Asset Management (co-lead)**	Bluenumber
3	Aviva Investors	Centre for Artificial Intelligence Research (CAIR)
4	Macquarie Asset Management	E-Governance and Internet Governance Foundation for Africa (EGIGFA)
5	Ethos Foundation	EthicsGrade
6	Sycomore Asset Management	Global AI Ethics Institute (GAIEI)
7	Raiffeisen Schweiz	Initiate: Digital Rights in Society (Paris Peace Forum)
8	Ethos Engagement Pool International	Paradigm Initiative
9	Öhman Fonder	The Internet Commission
10	Sarasin & Partners	Transcendent
11	Mercy Investment Services	Thomson Reuters Foundation
12	SDG Invest	Data Economy Policy Hub (DepHUB)
13	Church Commissioners for England	
14	EdenTree Investment Management	
15	Christian Brothers Investment Services	
16	NEI Investments	
17	Cardano	
18	HSBC Asset Management	
19	Ausbil Investment Management	
20	Ethical Partners Funds Management	
21	GAM Investments	
22	Robeco	
23	DNB Asset Management	
24	Amundi	
25	Sustainability Group of Loring, Wolcott & Coolidge	
26	Schroders	
27	EQ Investors	
28	Glasswing Ventures LLC	
29	Thematics Asset Management	
30	Acadian Asset Management	
31	Candriam	
32	Pictet Group	

3. Engaging companies on ethical AI

Investor engagement and impact

WBA's 2021 Digital Inclusion Benchmark found that, while many technology companies spent pages citing the benefits and potentials of AI, few expressed concerns about the risks. Just 20 out of 150 (13%) disclosed a commitment to abide by a set of ethical AI principles.

It is widely accepted that artificial intelligence poses numerous risks to people and societies, from the casual infringement of privacy rights to labor displacement and the pollution of information ecosystems with mass-produced falsehoods. Some of those risks have already manifested themselves. OpenAI, the developer of popular generative AI tool ChatGPT, has repeatedly come under fire from regulators and civil society groups for allegedly feeding its models with personal data, ignoring copyright law, generating misinformation, and facilitating the spread of inauthentic activity online. Online platforms are rapidly introducing AI-based virtual assistants, reportedly replacing human support staff en masse before fully testing their new tools. Rapid movement to regulate AI, particularly in the European Union and China, signals that policymakers are seeing the technology as a potential disruptor of the current order whose influence will be difficult to harness.

Stories such as these capture the imbalances that advancements in AI can generate in the digital space. But they also spark questions from investors and civil society about the principles that guide the development and deployment of AI. As an articulation of a company's ethos, such principles should serve as a starting point for more specific disclosures on how they are put into practice. But even in isolation, they are vital to illustrating structured thinking and direction in a company's approach to AI beyond maximizing profit at all costs.

Although nearly all of the companies evaluated in the Digital Inclusion Benchmark develop, use, procure, or sell AI tools, very few publish a consolidated set of AI principles. The Digital CIC arose directly from this persistent gap in corporate transparency.

After the launch of the 2021 Digital Inclusion Benchmark, WBA sent letters to the 130 companies found to be lacking commitments to ethical AI, asking them to fill this gap in disclosure. Since 2022, investors under the banner of the CIC have followed up on this by leading outreach on ethical AI **to 44 of the 150 companies assessed in the 2021 DIB**, focusing on those that did not have a public set of principles to steer their development and use of AI.

These engagements have helped push 14 additional companies to announce their AI principles in the 2023 DIB, bringing the total of those with disclosed principles up to 47 out of 150 (31%). WBA's research found that five of the companies newly benchmarked in 2023 (**Capgemini**, **HPE**, **Juniper Networks**, **NXP**, and **Panasonic**) also fulfilled this expectation. Overall, by September 2023, just over a quarter (26%) of all companies assessed in the 2023 Digital Inclusion Benchmark were able to demonstrate a set of freely accessible AI ethics principles.

Four questions guided investors' dialogue with the benchmarked companies:

1. How does the company **consider and define** artificial intelligence in its business and business strategy?

2. How are **ethical considerations** integrated in the development and deployment of AI?

3. How are the ethical aspects of AI **governed** in the company and who is responsible for overseeing these processes?

4. Is the company considering a **public commitment** to ethical AI principles or is it already in the process of making such a commitment public?

These engagements have generated several overarching trends.

Most (but not all) companies are willing to come to the table

In total, 28 of the 41 companies that investors engaged with responded to their collective outreach, while 11 remained unresponsive. Those that engaged comprised a cluster of companies in which ethical AI considerations are at varying degrees of maturity, from powerful "everything app" platform companies with little visible work on AI ethics to popular services that regularly share updates on algorithmic responsibility.

For the most part, the responsive companies were highly receptive to feedback. Their unresponsive peers came from a mix of industries: e-commerce, hospitality, transportation, social media, hardware and semiconductors, networking, streaming, and telecommunications. They were also spread across three continents: North America, Asia, and Europe. All 11 are prominent in the tech landscape of their home jurisdictions.

Among companies that chose not to engage, a recurring argument was that the risks of AI were either not material to their business or not applicable to it at all. Several companies across the industry spectrum made this claim, but it was especially prevalent among media and communications firms that build hardware or infrastructure for internet and television access. This contradicts the many known uses of AI in this broad group of companies, some of which are prominently displayed on their websites: personalized content recommendations, computer vision, speech recognition and voice control, smart home systems, and chatbots, to name a few.

This points to a problem that is vigorously discussed in the responsible investing community: there is no universally accepted position on what constitutes a "material issue." The last few years have brought a positive shift toward a system anchored in *double materiality*, under which companies are expected to consider and report on both issues that affect their bottom line (*financial materiality*) and those that affect the economy, the environment, and people (*impact materiality*). It is also widely accepted that materiality is not static—a point driven home by catalytic events such as the COVID-19 pandemic, which realigned the priorities of entire industries. These changes help bring the world of shareholders closer to lived realities where the deployment of technology can lead to human rights harms.

But AI poses a challenge even to this new perspective on materiality. Automation is often part of a company's fabric and is used extensively for purposes that may seem benign in isolation but can become deeply problematic when linked with other features, such as poor privacy protections. During the CIC engagements, one large cloud computing company noted that it steers clear of "controversial" uses of AI such as biometrics, even as it pursued multiple acquisitions and new integrations of AI. In such cases, the malleable definition of what is controversial can create major blind spots, exacerbated by the lack of public algorithmic principles or an operational overview of how the company uses AI.

Policies and oversight structures are unclear or hidden from view

Technology companies often argue that they define the extent of their responsible AI obligations differently according to the products they sell. This can mean, for instance, applying more scrutiny for high-risk tools and uses and lower scrutiny for deployments that have to do with routine system maintenance.

One of the problems with this approach is that it blurs our understanding of the company's governance and oversight structures on AI-related issues. Although several companies in dialogue with investors outlined board-level committees and working groups broadly charged with overseeing AI issues, it was rarely clear which management-level teams (e.g., product, policy, trust and safety, and human rights teams) were involved in designing ethical AI policies or how their concerns can be escalated to executives and the board.

Similarly, in the few cases where companies have convoked independent external advisory councils on ethical and human rights topics, the remit of their work and the extent of their influence on the company's ethical AI policies and practices is not well explained. Overall, it remains exceedingly difficult to understand who signs off on major strategic decisions involving AI, how the board of directors is involved, and which teams at the management level participate in the conversation.

Another recurring blind spot investors reported was the blurry relationship between companies' high-level principles and operational policies. In some cases, this amounts to an open question: What should or should not be disclosed, and at what level – the company or the product? The answer to this question remains elusive in the absence of formalized and broadly adopted standards.

In one illustration of this problem, several of global tech giants have publicly revealed the existence of internal responsible AI guidelines but do not make them public, even if they openly discuss other components of their work on ethical AI. One such case is **Spotify**. The Swedish audio streaming pioneer stood out for its large body of published research on responsible AI and sharing lessons learned from a structured algorithmic impact assessment. But the company does not have a consolidated set of public AI principles, and its 2022 Equity & Impact Report discloses an "Algorithmic Policy, Guidelines, and Best Practices" that are not openly available. Contrary to the practices of many of its peers, it also does not publish any enforcement data to highlight the effectiveness and challenges of the algorithmic platform safety tools it acquired in 2022.

This "transparency imbalance" is far from an isolated case and speaks to the low degree of standardization in the area of AI ethics, which regulatory instruments such as the EU's Digital Services Act and emerging AI Act only partially mitigate.

Companies are not clear about how they monitor AI tools they buy and sell

AI tools and models are not always developed in-house. According to one estimate by Deloitte, 65% of companies either purchase packaged solutions or subscribe to "AI as a service" offers, with the remaining 35% building AI tools internally.

Companies also inherit these tools through mergers and acquisitions. Between 2020 and 2023, U.S. cloud computing company **ServiceNow** acquired at least three AI-focused startups or firms, including G2K, which is promoted as improving "on-site customer experiences" and increasing "employee productivity." **Microsoft**'s integration of automated moderation company Two Hat in October 2021

prompted a tenfold spike in policy enforcement actions such as user bans on the Xbox gaming platform in the quarter that followed the acquisition. **Airbnb**, **Spotify**, and other companies with large volumes of user-generated content (UGC) have all acquired similar startups offering AI-powered content moderation services.

Investors allied under the CIC asked multiple companies whether they require a commitment from commercial clients to follow their ethical AI guidelines and how they evaluated the ethical aspects of their AI-focused acquisitions. But the evidence for consistent processes of this kind remains scant.

Although companies routinely deploy AI tools developed by third parties in their operations, it is rarely clear how they are vetted and what principles they are expected to follow. This applies in the opposite direction as well: few companies deemed it their responsibility to monitor for compliance with their terms of use or adherence to ethical considerations when making a proprietary AI model or tool available to others. This is a major risk in the age of freemium generative AI models, which have already been extensively integrated into commercial products developed by others.

> **"Over the last twelve months, we have seen a lot of progress in active dialogue with companies, both in scope and in shared learning among investors, on how individual companies are advancing AI governance practices across the value chain of technology development and use. This initiative has motivated companies to advance disclosure of public ethical AI policies. But there is much more work to be done— some companies have declined to engage with this initiative, and some do not recognize the imperative for an ethical framework guiding the use of AI."**

> Lauren Compere
> Managing Director/Head of Stewardship & Engagement
> Boston Common Asset Management

Multi-stakeholder action

Other stakeholders in the Digital CIC, led by Women at the Table, have been instrumental in raising awareness on responsible and ethical AI. The working group has engaged in public events advocating for responsible and ethical AI, including publishing a joint statement calling on technology companies to commit to responsible development and application of AI, guided by ethics, respect for human rights and the principle of leaving no one behind.

The stakeholders also made a joint submission to the United Nations (UN) Global Digital Compact in early 2023 on digital inclusion topics, including AI regulation, following a series of consultations. The UN Secretary General's Common Agenda proposes a Global Digital Compact to be agreed upon at the Summit of the Future in September 2024 through a technology track involving all stakeholders: governments, the UN system, the private sector (including tech companies), civil society, grass-roots organisations, academia, and individuals, including youth. The Global Digital Compact is expected to

"outline shared principles for an open, free and secure digital future for all". Through this submission, the Digital CIC aims to influence policy making at the UN level on AI governance and hope our recommendations will be incorporated into the Secretary General's final report.

> **"The CIC on ethical AI is a unique coalition of civil society partners and private sector actors working together to help technology companies understand the imperative of taking a human rights-based approach to design and deployment of their innovations. Only if we work together across disciplines and sectors can we hope to achieve the critical ambitions of the Sustainable Development Goals – to leave no one behind as every corner of the globe is touched by some form of digital technology and AI."**

<div align="right">

Caitlin Kraft-Buchman
CEO and Founder
Women at the Table

</div>

4. New findings on ethical AI

In 2023, building on the findings of the 2023 Digital Inclusion Benchmark, we set out to develop a new, standalone indicator on ethical AI. Among other points for consideration is the focus on companies' commitment to those principles as well as how they are operationalised within an organisation. This section gives an overview of updated findings on ethical AI principles and the proposed AI indicator.

The DIB methodology considers four measurement areas for digital inclusion: improving *access*, enhancing *skills*, building trust to foster beneficial *use*, and advancing openly, sustainable, and ethical *innovation*. Company disclosure related to ethical AI has been assessed under the *innovation* measurement area. In the 2023 benchmark, ethical AI was not a standalone indicator, but was part of four elements that fall under indicator **I4: The company practices inclusive and ethical research and development.** The elements in this indicator and their weights are as follows:

1. Does the company have a publicly available **ethical AI framework**? (0.25)

2. Does the ethical AI framework include **human rights considerations**? (0.25)

3. Does the company have a **committee dedicated to ethics**? (0.10)

4. Does the company have a committee that explicitly considers **ethics in R&D/AI**? (0.15)

In 2023, the DIB is conducting a public methodology review to ensure our work aligns with the most relevant topics and examples of best practice in a constantly changing sector. Given increasing focus on the risks posed by AI, we have proposed a standalone ethical AI indicator with the following elements:

1. Does the company have its own publicly available **ethical AI principles** that apply at the group level? (0.25)

     a. Does the company **clearly commit** to these principles? (0.25)

     b. Are the principles **the company's own**? (0.25)

     c. Are the principles disclosed in a **standalone document**? (0.25)

2, Does the ethical AI framework include **respect for human rights**? (0.25)

3. Does the company have a **committee** with oversight of ethical AI? (0.25)

4. Does the company describe how its ethical AI principles are **operationalised**? (0.25)

5. Does the company carry out **AI-focused human rights impact assessments**? (0.25)

A company can achieve a maximum score of 2. More details about the rationale for the elements can be found in WBA's 2023 *Digital Inclusion Benchmark Insights Report*.[1]

Preliminary results

Since the launch of the 2023 DIB in March, our ongoing research has yielded new data on the 200 digital technology companies in scope, using the proposed ethical AI indicator. Our continuous monitoring of company disclosures allowed us to track the pace of companies' progress in the course of the CIC's engagement efforts.

We found that, as of September 2023, **52 of the 200 of the companies in the DIB (26%) have adopted ethical AI principles**. This marks an increase from 33 companies when we launched the Digital CIC in 2022, with 19 additional companies joining the cohort that fulfils this basic expectation. Of the companies that have a publicly available ethical AI framework, only 24 have a clearly stated commitment to their principles.[2] Three companies – **Cisco**, **Softbank**, and **Spark** – participated in the launch of the Digital CIC and made their ethical AI principles publicly following the launch.

Most companies developed their principles in-house, but a small subset declared their adherence to an external framework. Similarly, 45 out of 52 companies that adopted ethical AI principles had them published in a standalone document. This is not trivial: companies often make key disclosures like this in voluminous ESG reports or in another part of a suite of annual reports whose audience consists primarily of investors. But the rules a company sets for itself in developing and deploying AI capabilities should be accessible to everyone, whether or not they are users of its products and technology.

Three companies – **Deutsche Telekom**, **Microsoft**, and **Telefónica** – achieved the highest possible score of 2, fulfilling all eight of our preliminary expectations. These companies exemplify leading practice by explaining how they operationalise their ethical AI principles while conducting human rights impact assessments (HRIAs) with an explicit focus on AI. But they are the exception to the rule: in total, only 6 of the 52 companies with AI principles disclosed that they conducted HRIAs on their development and use of AI tools, while only 8 detailed how they translated their principles into practice. While we are seeing an increase in companies developing principles and making them

publicly available, there is still room for improvement when it comes to transparency and protecting the human rights of users.

Figure 1: Preliminary ethical AI indicator scores among companies that publicly disclose their principles

Note: These preliminary findings were produced independently by the Digital Inclusion Benchmark's research team. They do not necessarily reflect the final form of the indicator.
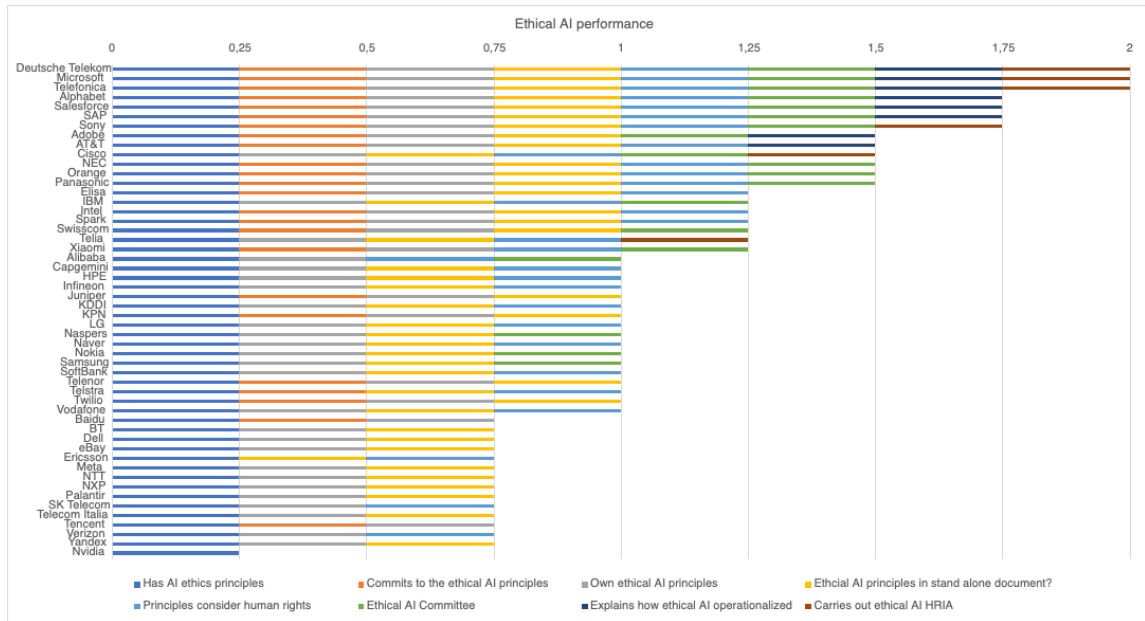
Note: AI HRIA = Artificial Intelligence Human Rights Impact Assessment. Data as of September 5, 2023.

Companies with ethical AI principles (as of September 2023)

Note: These preliminary findings were produced independently by the Digital Inclusion Benchmark's research team. They do not necessarily reflect the final form of the indicator.

Company	Policy reference	Year adopted	Commits to its principles?	Developed its own principles?	Standalone document?	Principles have human rights considerations?	AI ethics committee?	Explains how principles are operationalised?	Conducts AI-focused HRIAs?
Adobe	Adobe's Commitment to AI Ethics	2021	Yes	Yes	Yes	Yes	No	Yes	Yes
Alibaba	Science and technology ethics (ESG Report)	2022	Yes	No	Yes	No	Yes	Yes	No
Alphabet	Artificial Intelligence at Google: Our Principles	2018	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AT&T	Artificial Intelligence at AT&T: Our Guiding Principles	2019	Yes	Yes	Yes	Yes	Yes	No	Yes
Baidu	AI for Social Good (ESG Report)	2023	Yes	Yes	Yes	No	No	No	No
BT	Our responsible tech principles	2020	Yes	No	Yes	Yes	No	No	No
Capgemini	Our Code of Ethics for AI	2021	Yes	No	Yes	Yes	Yes	No	No
Cisco	The Cisco Responsible AI Framework	2022	Yes	No	Yes	Yes	Yes	Yes	No
Dell	Dell Technologies Principles for Ethical Artificial Intelligence	2022	Yes	No	Yes	Yes	No	No	No

Company	Policy reference	Year adopted	Commits to its principles?	Developed its own principles?	Standalone document?	Principles have human rights considerations?	AI ethics committee?	Explains how principles are operationalised?	Conducts AI-focused HRIAs?
Deutsche Telekom	Guidelines for Artificial Intelligence	2018	Yes	Yes	Yes	Yes	Yes	Yes	Yes
eBay	Third Party Code of Business Conduct and Ethics	2022	Yes	No	Yes	Yes	No	No	No
Elisa	Ethical principles for Data and Artificial Intelligence	2019	Yes	Yes	Yes	Yes	Yes	No	No
Ericsson	EU Ethics guidelines for trustworthy AI	2019	Yes	No	No	Yes	Yes	No	No
HPE	AI Ethics and Principles	2021	Yes	No	Yes	Yes	Yes	No	No
IBM	AI Ethics	2018	Yes	No	Yes	Yes	Yes	Yes	No
Infineon	Responsible AI at Infineon	2023	Yes	No	Yes	Yes	Yes	No	No
Intel	Responsible AI	2021	Yes	Yes	Yes	Yes	Yes	No	No
Juniper	AI Innovation Principles	2021	Yes	Yes	Yes	Yes	No	No	No
KDDI	AI R&D and Utilization Principles for KDDI Group	2021	Yes	No	Yes	Yes	Yes	No	No
KPN	Responsible use of AI for KPN	2023	Yes	Yes	Yes	Yes	No	No	No
LG	LG AI Ethics Principles	2021	Yes	No	Yes	Yes	Yes	No	No

Company	Policy reference	Year adopted	Commits to its principles?	Developed its own principles?	Standalone document?	Principles have human rights considerations?	AI ethics committee?	Explains how principles are operationalised?	Conducts AI-focused HRIAs?
Meta	Facebook's five pillars of Responsible AI (Blog post)	2021	Yes	No	Yes	Yes	No	No	No
Microsoft	Microsoft responsible AI principles	2016	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Naspers	Approach to AI Ethics	2022	Yes	No	Yes	Yes	No	Yes	No
Naver	AI Ethics Principles	2021	Yes	No	Yes	Yes	Yes	No	No
NEC	NEC Group AI and Human Rights Principles	2019	Yes	Yes	Yes	Yes	Yes	Yes	No
Nokia	Responsible AI	2022	Yes	No	Yes	Yes	No	Yes	No
NTT	Our Approach to the Use and R&D of AI	2021	Yes	No	Yes	Yes	No	No	No
Nvidia	Trustworthy AI (CR Report)	2021	Yes	No	No	No	No	No	No
NXP	Artificial Intelligence (Code of Conduct)	2021	Yes	No	Yes	Yes	No	No	No
Orange	EU Ethics Guidelines for Trustworthy AI	2019	Yes	Yes	Yes	Yes	Yes	Yes	No
Palantir	Palantir Technologies' Approach to AI Ethics	2023	Yes	No	Yes	Yes	No	No	No
Panasonic	Panasonic Group's Efforts to Ensure	2022	Yes	Yes	Yes	Yes	Yes	Yes	No

Company	Policy reference	Year adopted	Commits to its principles?	Developed its own principles?	Standalone document?	Principles have human rights considerations?	AI ethics committee?	Explains how principles are operationalised?	Conducts AI-focused HRIAs?
	Responsible AI-Utilizing Operations								
Salesforce	Trusted AI	2020	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Samsung	Samsung AI principles	2019	Yes	No	Yes	Yes	No	Yes	No
SAP	SAP's Guiding Principles for Artificial Intelligence	2018	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SK Telecom	Ethics of AI (Annual Report)	2021	Yes	No	Yes	No	Yes	No	No
SoftBank	SoftBank AI Ethics Policy	2022	Yes	No	Yes	Yes	Yes	No	No
Sony	AI Engagement within Sony Group	2018	Yes	Yes	Yes	Yes	Yes	Yes	No
Spark	Spark's Artificial Intelligence Principles	2022	Yes	Yes	Yes	Yes	Yes	No	No
Swisscom	AI Ethics Principles	2020	Yes	Yes	Yes	Yes	No	Yes	No
Telecom Italia	Ethical Principles for AI	2019	Yes	No	Yes	Yes	No	No	No
Telefonica	AI Principles of Telefonica	2018	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Telenor	EU Ethics Guidelines for Trustworthy AI	2019	Yes	Yes	Yes	No	No	No	No

Company	Policy reference	Year adopted	Commits to its principles?	Developed its own principles?	Standalone document?	Principles have human rights considerations?	AI ethics committee?	Explains how principles are operationalised?	Conducts AI-focused HRIAs?
Telia	Guiding Principles on Trusted AI Ethics	2019	Yes	No	Yes	Yes	Yes	No	No
Telstra	Australian Government AI Ethics Principles	2019	Yes	Yes	No	Yes	Yes	No	No
Tencent	AI for Good (ESG Report)	2021	Yes	Yes	Yes	No	No	No	No
Twilio	Twilio's approach to trusted Customer AI	2023	Yes	Yes	Yes	Yes	No	No	No
Verizon	Responsible AI Program (ESG Report)	2022	Yes	No	Yes	No	Yes	No	No
Vodafone	Vodafone Group's Artificial Intelligence (AI) Framework	2019	Yes	No	Yes	Yes	Yes	No	No
Xiaomi	AI Technology (Privacy White Paper)	2022	Yes	Yes	Yes	No	Yes	Yes	No
Yandex	Principles	2022	Yes	No	Yes	Yes	No	No	No

[1] World Benchmarking Alliance. 2023. "4 Assessing company performance on ethical artificial intelligence." In *Digital Inclusion Benchmark Insights Report*. https://www.worldbenchmarkingalliance.org/research/2023-digital-inclusion-benchmark-insights-report/

[2] More details on commitment language that fulfils our standards can be found in the 2021 DIB Scoring Guidelines.

5. Emerging policies on AI governance

AI has radically transformed digital companies, and its societal impacts at once highly tangible and poorly understood. While it has the potential to help solve global problems, it also poses real risk to people and societies. Without regulation, it can reproduce real-world discrimination and bias, fuel hate speech, and threaten fundamental freedoms and rights. Surging awareness of AI-related risks has prompted a wave of national policies as well as frameworks and initiatives by intergovernmental and non-governmental organizations.[1]

At the forefront of these efforts is the **European Union**, where a legal framework that regulates the development and use of Artificial Intelligence is currently taking shape through the **AI Act** (AIA). This regulation aims to ensure that the use of AI in the EU is safe, transparent, traceable, non-discriminatory, and environmentally friendly. On June 14, 2023, the European Parliament voted in favour of its negotiating position on the AI Act.

The AI Act proposes a risk-based approach that categorizes AI technologies by the levels of risks they pose to human rights. Depending on risk level, providers and users will be expected to comply with different obligations. Some AI systems like social scoring, real-time and remote biometric identification, as well as cognitive behavioural manipulation of people are banned under this approach due to the unacceptable risk they are deemed to represent.

In the **United States**, more than 500 AI companies were founded in 2022 alone, representing half of all new AI companies that year worldwide. In May 2023, the White House published a Blueprint for an AI Bill of Rights, which identifies five principles to guide the design, use, and deployment of automated systems. It applies to automated systems that have the potential to significantly affect rights, opportunities, or access to critical resources or services for the American public. The five principles include safe and effective systems; algorithmic discrimination protections; data privacy; notice and explanation; and human alternatives, consideration, and fallback. The AI Bill of Rights aligns with the EU AI Act in that it aims to mitigate discrimination and protect human rights; however, it does not carry the force of law.

Another important hub of AI innovation is **Asia**. In 2022, China reported more than 13 billion USD in private investment toward developing AI capacities. The following year, it published and put into effect a set of guidelines for generative AI tools that looks to regulate the booming sector. The new guidelines focus on promoting fairness, robustness, privacy, and transparency in the sector. However, the framework only applies to entities that offer AI services to the public inside the country and follows the strictures of China's regulatory regime. Japan's approach to AI regulation promotes voluntary initiatives by businesses. Within the ASEAN region, six member countries — Indonesia, Malaysia, the Philippines, Singapore, Thailand, and Vietnam — have AI guidelines in place.

While the remaining ASEAN countries lack specific policies that incorporate AI language, the collective membership of ten has reached a consensus to formulate an ASEAN Guide on AI Governance and Ethics (referred to as the ASEAN 'AI Guide'). This initiative aligns with global trends in AI regulation and aims to encompass three pivotal areas. These areas encompass the advancement of AI as a technology to stimulate economic growth and development, the augmentation of capacities—such as human capital capabilities—to harness the advantages presented by AI applications, and the establishment of ethical and governance frameworks for these applications of AI. If a formal region-

---

[1] The OECD maintains a live repository of national AI strategies and policies.

wide policy on AI is developed within ASEAN, it will likely be rooted in the concept of 'best practices by design' rather than relying on the legally binding mechanisms that characterise the EU's AI Act.

In other regions of the world less commonly associated with frontier uses of AI, regulation is generally scant. For example, **Latin America and the Caribbean** was found to be the third lowest ranked region in the 2021 Government AI Readiness Index, with only 5 countries maintaining discrete national-level AI strategies. Despite a shortage of legislation in many Latin American countries, companies and society at large have enthusiastically embraced AI. In 2019, around 80% of large Latin American businesses were using AI, buttressed by strong public support. Existing regional AI strategies place a focus on the potential for positive development impact that can derive from the use of this technology. In the Caribbean, UNESCO has sponsored the Artificial Intelligence Initiative, which aims to promote and provide recommendations for a global normative instrument to ensure the application of human-centred AI. This initiative also contemplates public-facing education around AI topics.

As of 2022, no country in **Sub-Saharan Africa** had dedicated AI regulations, while only Mauritius had partial legislation on the subject. At the regional level, the African Union's Digital Transformation Strategy for Africa (2020-2030) addresses the need to keep pace with advances in technology. Under this document, different stakeholders made policy recommendations around how to legislate AI technology in African countries. It advocates for including AI as part of broader national strategies and maintaining a balance between an enabling environment and ethical, legal and governance considerations.

More mature legislation might hamper the development of new AI tech in regions that are at a different stage of development. Nevertheless, there is widespread need for a universal set of standards that provides a benchmark for companies to follow. In the absence of any legislation, the more vulnerable risk being the most affected by unethical AI practices. In this context, initiatives such as the Digital CIC can foster industry progress and providing a reliable assessment on companies' ethical AI development.

The Digital CIC's other members have played a vital role in contributing to policy discussions surrounding the development of trustworthy and responsible AI governance within multilateral forums. Collaboratively, we worked with stakeholders to create a policy brief titled "Why the G20 Should Lead Multilateral Reform for Inclusive Responsible AI Governance for the Global South." This document was then submitted to the 2023 G20 Presidency during India's leadership.

During the Indian G20 Presidency, there was a strong emphasis on advocating for AI regulation, focusing on the potential misuse of AI technologies by malicious entities and the necessity for transparent and accountable AI governance frameworks. The government recognized the importance of these frameworks in ensuring the responsible use of AI technologies.

Our collaboration with the CIC members on this policy briefing is significant because it stems from the realization that **the Global South confronts specific challenges with the emergence of artificial intelligence.** Addressing these challenges requires urgent multilateral reforms to establish a comprehensive and inclusive approach to AI governance. Such an approach should engage numerous stakeholders and ensure the legitimacy of international cooperation. It must also address concerns like the concentration of power and control, promote fair distribution of digital benefits, and mitigate the subset of AI risks with the potential to have sweeping effects on communities and societies. Every global initiative to steer AI toward accountability, whether current or prospective, should consider the socioeconomic panorama of both countries in the Global South and those in the Global North.

6. The road ahead

Soon after the Digital CIC launched in 2022, the technology industry was rocked by the debut of OpenAI's ChatGPT, an advanced chatbot trained on troves of data scraped from the Internet. Its unmatched speed, adaptability, and precision in answering complex prompts set off a surge of interest and investment in generative AI (GenAI) – the automated creation of media from user inputs.

Generative AI immediately shifted the priorities of the industry's most influential players, triggering an arms race that has seen dozens of companies integrating third-party GenAI tools or announcing their own. At the same time, mass layoffs swept through the U.S. tech industry, costing hundreds of thousands of jobs and generating knock-on effects around the world. Some companies implemented cost-cutting measures that explicitly focused on replacing human workers with AI tools, without showing how they plan to mitigate any resulting harms.

Such events have contributed to widespread concern that we are entering a new era of "moving fast and breaking things." In response to this challenge, both the Digital Inclusion Benchmark and the Digital CIC will expand and evolve in the months ahead. Our next steps in the campaign will be driven by three interlocking needs continuously flagged by both members of the CIC and our other Allies:

- **Richer data:** In early 2024, WBA will release an updated methodology for the Digital Inclusion Benchmark, with a new indicator on ethical AI whose preliminary version comprises the data on company disclosures in this report. In light of the fast changes in the landscape of AI, we plan to update this data on a regular basis. In addition to better reflecting the pace of progress, this will help prioritise companies for individual and collective engagement.

- **Higher expectations:** Although more than a quarter of the 200 companies ranked in the Digital Inclusion Benchmark now have ethical AI principles, our findings show that these principles do not always translate into transparency on how they are implemented. We plan to expand the formal scope of the CIC to seek not just the disclosure of high-level principles, but also evidence of how they are operationalised and how companies factor AI into their human rights due diligence. This will also open the path for investors and other stakeholders to initiate engagement with more companies.

- **More coordination and representation:** The breakneck pace of AI development is likely to spark or inform multiple collective campaigns. We seek to work with other engagement initiatives to strengthen the feedback loops between us. These initiatives often cover complementary aspects of ethical AI, from the digital rights-focused engagements overseen by the Investor Alliance for Human Rights to efforts focused on selected topics and high-risk deployments of AI technology. Digital CIC members are already leading some of these campaigns. At the same time, we will draw up a strategic plan to build out the civil society cohort of the CIC, with a focus on experts and organizations from underrepresented countries and communities. Our goal is to increase interaction and coordination between CIC investors and other stakeholders, continuously channeling expertise into action.

We anticipate that the need for ethical AI practices will only continue to grow. These steps will help us chart the course for the future of corporate accountability efforts aiming to improve the practices of companies that develop or deploy frontier technologies. Collaboration and inclusion will be our top principles as we advance toward this goal.