EOS Digital Governance Principles

Navishka Pandit, Engager

Nick Pelosi, Engager

Ross Teverson, Regional Team Lead, Asia and GEMs

EOS at Federated Hermes Limited

Federated Hermes Limited

www.hermes-investment.com
For professional investors only

Table of contents

Key takeaways

1. The financial materiality of capturing emerging digital opportunities, while also mitigating regulatory and reputational is already significant and rising. New technologies offer huge potential for value creation, but risks associated with overlooking or ignoring unintended harms must be addressed.

2. Regulation of digital services and artificial intelligence (AI) is fragmented and in many areas lags the pace of development of technology. Companies should be proactive in establishing guardrails and in making constructive contributions to public policy development.

3. Governance, policies and culture are foundational to generating long-term returns on investment as well as positive outcomes for wider stakeholders. Companies' governance structures and policies should support compliance with current and anticipated future regulations, management of enterprise risks, and promote a commitment to ethics and mitigating/doing no harm.

4. Appropriate management of the risks associated with digital services can help mitigate wide-ranging unintended negative impacts – on young people, the environment, cybersecurity, and the functioning of civil society.

5. Companies should deploy AI in human capital management responsibly and use the digital revolution as an opportunity to reinforce their commitments to their workforce. Companies that promote a "just AI transition" should be able to capture opportunities, while also minimizing workforce-related disruption.

6. Use cases and associated risks differ across industries, necessitating sector-specific lenses.

Introduction

This paper builds upon the principles and expectations outlined in our previous publications and considers the rapidly evolving landscape for digital governance, including considerations relating to digital rights and the development and deployment of AI. We recognise that risks and opportunities related to digital products and services are dynamic and we welcome feedback on these expectations.

EOS has been engaging on digital rights since 2012, and the business and wider societal impacts of digital governance since 2018. Our first paper on this topic - Investor Expectations on Responsible AI and Data Governance (2019), set out a full engagement framework based on the six principles of **trust**, **transparency**, **action** (to avoid unintended consequences), **integrity**, **accountability**, and **safety**. We also advocated a 'three lines of defence' model for trusted AI implementation, with ethics as the first line of defence. Risk, governance and audit should form the second line of defence, and having the responsible use of AI embedded in strategy and operations provides a third line of defence. This framework and its emphasis on a responsible approach to AI data governance permeating through a company's culture, processes and operations remains entirely relevant today.

In 2022, we consolidated our approach digital governance under the wider sub-theme of digital rights, which we define as "human rights specific to digital products and services". Our EOS's Digital Rights Principles (2022) set out our core expectations of companies regarding privacy rights, freedom of expression, mitigation of negative societal impacts (including prioritising children), and the need to have robust AI governance and policies. Our stated expectations on AI governance and policies include disclosure of the range of purposes for which companies use algorithmic systems, explanations as to how they work, the enabling of user choice, and the elimination of unintended bias.

Since the publishing of our EOS Digital Rights Principles in 2022, many new developments (some of which we list on the following page) have given us cause us to reflect on what might constitute best practice for digital governance.

Key developments since mid-2022:

- The **release of OpenAI's ChatGPT3** (November 2022) and a succession of other large language models (LLMs), such as Anthropic's Claude and Google's Gemini, which raised awareness of AI's potential to have transformative impacts on business and society;

- **A proliferation of use cases** for AI, which extend beyond early adopting industries, such as technology and finance, into all sectors;

- **Record levels of investment** in AI, with research from the International Data Corporation suggesting worldwide spending on AI-enabled applications, infrastructure, and related services, will more than double by 2028 to reach $632 billion[1];

- **Record fines** issued by the EU in 2023 for mishandling of personal data[2];

- **Heightened concerns regarding youth impacts**, documented in Jonathan Haidt's *The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness* (2024), and brought into the spotlight by legal action against large tech firms[3]

- **New legislation to protect children** online, including the UK's Online Safety Act 2023[4], Australia's recent amendment to its Online Safety Act restricting access to certain platforms for children under 16, and Florida's H.B.3. Bill (2024), which prohibits children under 14 from creating social media profiles[5];

- **The passing into law of the EU's Digital Services Act**[6] **(2022) and the EU AI Act (2024)**, which allow for fines of up to 6% or 7% respectively of a company's annual global revenue for infringement, and China's AI-Deep Synthesis Regulations (2023);

- **Fears about the environmental impact** of the widespread deployment of AI tools that are highly energy intensive, with Morgan Stanley suggesting that by 2027, generative AI could use as much energy as Spain consumed in 2022[7], and news of a high-profile recommissioning of the Three Mile Island plant in a deal with Microsoft[8];

- US President **Donald Trump's announcement of the Stargate Project**, which he claimed would invest up to US$500 billion in developing US AI infrastructure[9];

- **Growing concerns about how AI is used by firms for recruitment**[10] and the unintended consequences of broader deployment within firms[11]

[1] Worldwide Spending on Artificial Intelligence Forecast to Reach $632 Billion in 2028, According to a New IDC Spending Guide
[2] Chart: EU Data Protection Fines Hit Record High in 2023 | Statista
[3] New York City sues social media platforms over youth mental health crisis | CNN Business
[4] Online Safety Act: explainer - GOV.UK (www.gov.uk)
[5] Florida Gov. Ron DeSantis signs bill banning children on social media - The Washington Post
[6] The EU's Digital Services Act (europa.eu)
[7] Powering the AI Revolution: AI Energy Demand and Opportunity | Morgan Stanley
[8] Why Microsoft made a deal to help restart Three Mile Island | MIT Technology Review
[9] The Stargate Project: Trump Touts $500 Billion Bid For AI Dominance | Forbes
[10] AI hiring tools may be filtering out the best job applicants - BBC Worklife
[11] Using AI in the workplace: ethical risks and policy responses - IFOW

Summary of EOS Digital Governance Principles

Topic	Investor Expectations
Oversight, principles, and approach to regulation	• Establish robust and transparent AI and data governance structures, ensuring appropriate oversight, expertise, and clear lines of accountability. • Publish ethical AI and data governance principles addressing: transparency and accountability; fairness and bias; privacy; and other salient risks. • Integrate a culture of responsibility throughout the company, reinforced by training and regular solicitation of feedback from a range of stakeholders. • Establish no-go areas – technologies and use cases that the company considers unethical or presenting an unacceptable level of risk. • Disclose lobbying activities and advocate for regulatory consistency across markets where feasible.
Protect privacy and freedom of expression	• Obtain user consent for collection, storage, and utilisation of data, including targeted advertising, and ensure responsible use of facial recognition technology. • Maintain clear policies and processes for responding to data requests that can impact the privacy or freedom of expression of users. • Ensure robust governance of cybersecurity risks.
Action on negative societal impacts	• Prioritise children and young people and promote child-safe AI in the development and deployment of AI models. • Disclose how content is moderated and report actions taken.
Mitigate upstream environmental and social impacts	• Embed awareness and evaluation of energy intensity at a model level, hardware level and firm level. • Encourage minimization of data center water intensity in own operations and/or service providers. • Consider data centre design and location with a view to minimising negative environmental and social impacts.
Consider digital rights and the workforce	• Encourage employee engagement on AI use case development and deployment and use the digital revolution as an opportunity to reinforce commitments to employees. • Disclose anticipated impacts of AI deployment, as well as retraining/reskilling initiatives aimed at supporting a "just AI transition".

Defining Digital Rights and AI

Digital rights are human rights specific to digital products and services. These products and services can enhance human rights by increasing access to information and services or highlighting previously hidden issues. However, they can also cause harm, such as misuse of personal data, reinforcement of bias, cybersecurity breaches, and the rapid spread of false or extremist content.

Artificial Intelligence (AI), as defined by the International Organization for Standardization (ISO), is "an engineered system that generates outputs like content, forecasts, recommendations, or decisions" based on human-defined objectives. AI learns to analyse large amounts of data, recognise patterns, and make predictions or decisions. [12] Although AI is not new, its use has become more widespread, especially with the advent of generative AI (genAI), which can create high-quality text, images, and other content. GenAI can amplify both the positive impacts and risks of digital products and services. It raises ethical concerns about the use of personal data for training models, increases cybersecurity risks, accelerates the spread of disinformation, and can reproduce or amplify existing biases and inequities, a phenomenon described by Ruha Benjamin as "streamlining marginalization".[13]

This paper provides guidance and an engagement framework for companies materially exposed to digital rights, including those developing or deploying AI. The UN Guiding Principles on Business and Human Rights (UNGPs) outline the corporate responsibility to respect human rights, including a policy commitment, a due diligence process, and access to remedy. As digital rights were a nascent concept when the UNGPs were published in 2011, our *EOS Digital Rights Principles (2022) and these EOS Digital Governance Principles* include guidance for contemporary issues that require companies' attention when fulfilling their broader obligations to the UNGPs. Companies whose business models misalign with the UNGPs may have salient adverse impacts on peoples' lives and face material financial risks.

Companies are encouraged to follow the standards of the OECD AI Principles[14] (updated in 2024) and consider the aims of the UN Global Digital Compact[15], which opened for endorsement in 2024. The Compact commits governments to uphold international law and human rights online and to ensure a safe digital space. However, it also recognises the private sector's critical role in achieving these goals.

Important topics related to digital rights but outside the scope of this paper include closing the digital divide (the gap between those with and without reliable, affordable access to digital products and services), avoiding internet shutdowns (intentional disruption of digital services to specific populations) and respecting human rights in the technology sector supply chain.

[12] ISO - What is artificial intelligence (AI)?

[13] Race After Technology — Ruha Benjamin

[14] AI principles | OECD

[15] Homepage | Global Digital Compact

The business case for a responsible approach to digital governance

Digital technologies, especially AI, are driving a fourth industrial revolution and offering new business opportunities. However, they also pose ethical, reputational, and legal risks, such as privacy breaches, cybersecurity threats, bias, lack of transparency, misinformation, increased energy demand, and workforce disruption. To address these issues, global but fragmented regulations are emerging, including the EU's AI Act, US state-level bills, and China's AI-Deep Synthesis Regulations. These regional regulatory differences and the rapid proliferation of new technologies increase risks for businesses.

McKinsey (2023) suggests that, while companies may be inclined to take a "wait and see" approach to how regulation is implemented, "failure to handle AI and gen AI prudently can lead to legal, reputational, organizational, and financial damages" and "if the right governance and organizational models for AI are not built early…fixing a system after the fact can be both expensive and difficult to implement consistently across the organization".[16] Furthermore, the urgency to establish strong digital governance principles is underscored by potential advancements in emerging technologies, such as artificial general intelligence (AGI) – a term which describes AI systems that surpasses human abilities across a wide range of cognitive tasks, and commercial quantum computing. A framework established today will not only address near-term regulatory and reputational concerns, but also prepare organisations to navigate the complexities of future technological breakthroughs.

The business case for a responsible approach to digital governance goes beyond the mitigation of downside risks, with studies highlighting the opportunity to enhance a company's return on investment (ROI)." The potential for value creation is explained in a framework devised by IBM and Notre Dame University framework to evaluate ROI in AI Ethics[17], which identifies "the direct economic returns of such investments, the indirect paths to return through intangibles associated with organizational reputation, and real options associated with capabilities." Research conducted by Bain & Co found that companies with a comprehensive, responsible approach to AI earn twice the level of return on their investment in AI.[18]

"The business case for a responsible approach to digital governance goes beyond the mitigation of downside risks, with studies highlighting the opportunity to enhance a company's return on investment"

Companies should consider more than just the risks and opportunities that relate directly to digital services and AI, with growing calls for companies to address the environmental impact of data centers upon which digital services, and genAI tools in particular, are heavily dependent[19]. Investors and regulators are increasingly concerned that companies which operate or rely upon these data centres will either fail to meet their stated climate commitments or divert electricity from renewable sources

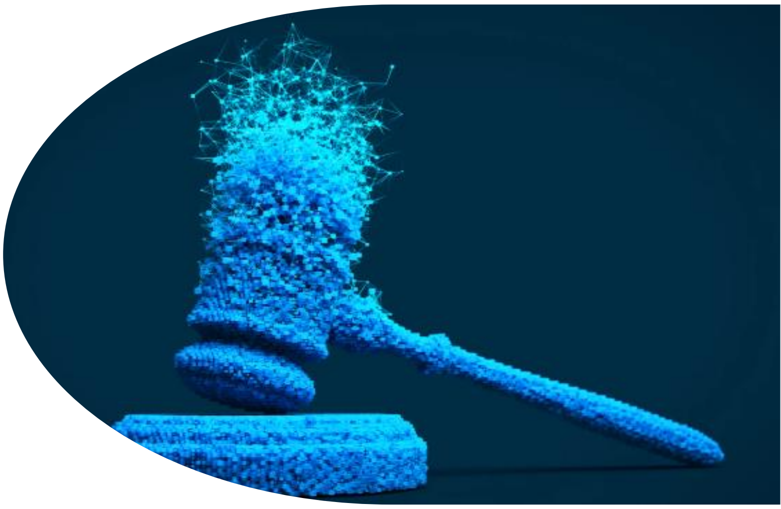
[16] Governance and regulation as generative AI advances | McKinsey

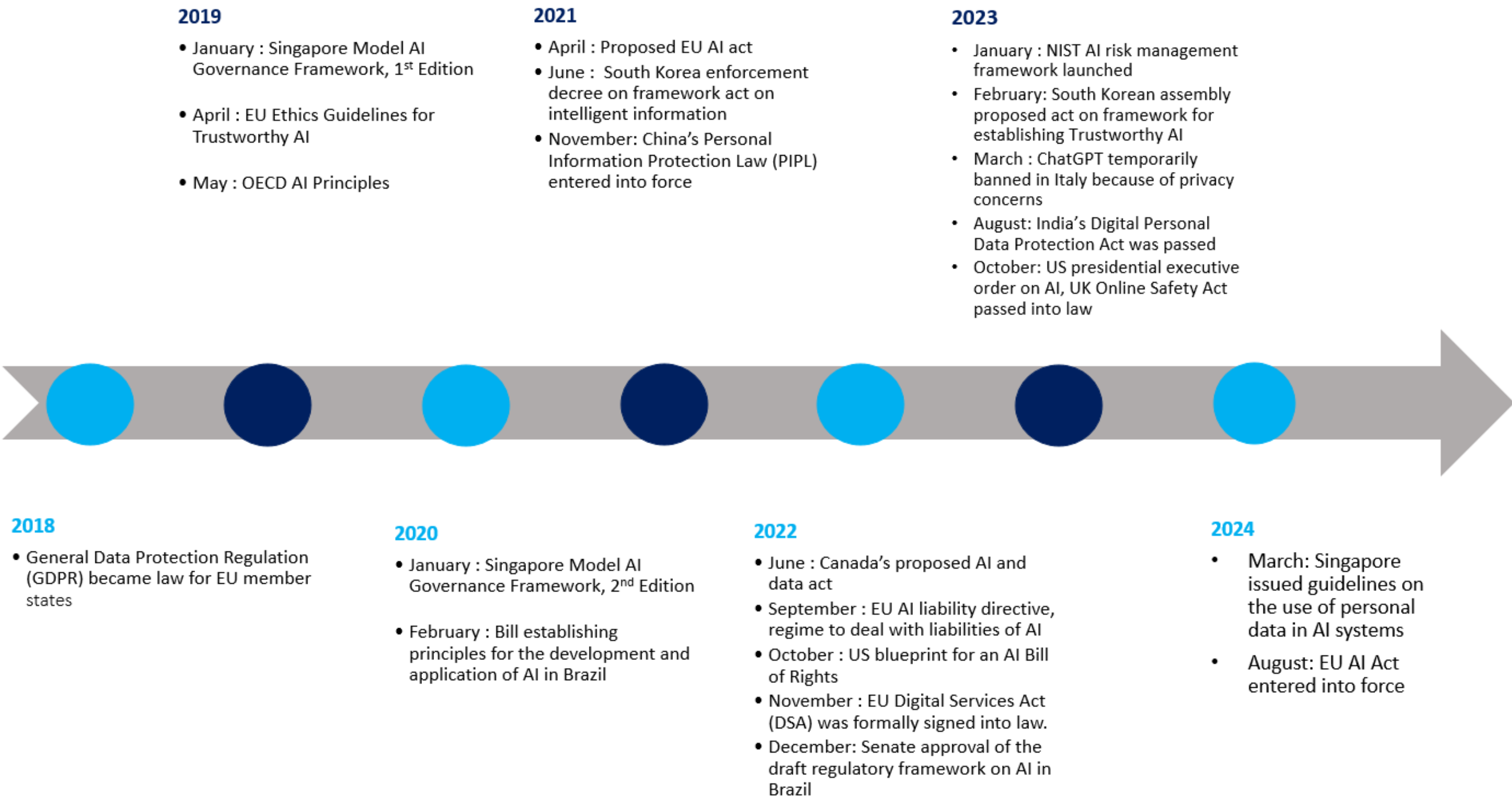
[17] The Return on Investment in AI Ethics

[18] Adapting Your Organization for Responsible AI | Bain & Company

[19] Growing data volumes drive need for ICT energy innovation | World Economic Forum

that would otherwise improve the sustainability of the energy mix in the broader economy. The high level of water consumption by data centres for cooling represents an additional risk, particularly in water-stressed areas[20]. Companies that are seen as part of the solution to these challenges, rather than part of the problem, should benefit from a stronger social license to operate and lower regulatory risk.





2018

- General Data Protection Regulation (GDPR) became law for EU member states

2019

- January : Singapore Model AI Governance Framework, 1st Edition
- April : EU Ethics Guidelines for Trustworthy AI
- May : OECD AI Principles

2020

- January : Singapore Model AI Governance Framework, 2nd Edition
- February : Bill establishing principles for the development and application of AI in Brazil

2021

- April : Proposed EU AI act
- June : South Korea enforcement decree on framework act on intelligent information
- November: China's Personal Information Protection Law (PIPL) entered into force

2022

- June : Canada's proposed AI and data act
- September : EU AI liability directive, regime to deal with liabilities of AI
- October : US blueprint for an AI Bill of Rights
- November : EU Digital Services Act (DSA) was formally signed into law.
- December: Senate approval of the draft regulatory framework on AI in Brazil

2023

- January : NIST AI risk management framework launched
- February: South Korean assembly proposed act on framework for establishing Trustworthy AI
- March : ChatGPT temporarily banned in Italy because of privacy concerns
- August: India's Digital Personal Data Protection Act was passed
- October: US presidential executive order on AI, UK Online Safety Act passed into law

2024

- March: Singapore issued guidelines on the use of personal data in AI systems
- August: EU AI Act entered into force

[20] Data centers draining resources in water-stressed communities - The University of Tulsa

Oversight, principles, and approach to regulation

The right oversight, policies and culture are foundational to generating positive outcomes for all stakeholders. Companies' governance structures and policies should support compliance with relevant regulations, management of enterprise risks, and promote a meaningful commitment to ethics and mitigating/doing no harm. As with other areas of material risk and opportunity for companies, appropriate board oversight and expertise is essential.

As digital technologies become increasingly important to a range of industries, many companies will find themselves materially exposed to risks and opportunities relating to digital services and/or AI. Given the complexity of associated risks and opportunities, only by having comprehensive and clearly articulated ethical AI and data governance principles can companies effectively communicate their approach to all stakeholders and ensure that their purpose, values and risk management aims are fully reflected in the products and services that they offer.

Our EOS Investor Expectations on Responsible AI and Data Governance state that companies should disclose the range of purposes for which they use algorithmic systems; explain how they work, including what they optimise for and what variables they take into account; and enable users to decide whether to allow them to shape their experiences.[21] Unintended racial, gender, and other biases have been identified within algorithms and can lead to inequitable outcomes[22] and companies should also take actions to eliminate these biases, including those recommended by the Equal AI Checklist to Identify Bias in AI.

Since the publishing of EOS's Digital Rights Principles in 2022, which reiterated the need for robust AI governance and policies, standard-setting organisations and AI safety institutes have produced useful guidance, which we encourage companies to consult. These include the introduction of ISO 42001[23] in December 2023, which supports the development of trustworthy AI management systems and the Artificial Intelligence Risk Management Framework[24] from National Institute of Standards and Technology (NIST) in January 2023.

AI training for employees is necessary for a company's approach and principles to be fully embedded within its culture, but there is currently a gap between the number of companies deploying AI and those providing employee training. A recent study by HR consultant, Randstad, showed that while 75% of companies assessed were adopting AI, only 35% of employees have received AI training over the course of the last year.[25] The Security and Exchange Commission's former Director of Enforcement, Gurbir Grewal, in a 2024 speech where addressed the risk of "AI washing" (making exaggerated or unsubstantiated claims about a company's use of AI) highlighted that importance of "proactive compliance", which "requires three things: education, engagement, and execution."[26]

[21] 2020 RDR Corporate Accountability Index | Ranking Digital Rights

[22] Artificial Intelligence Has a Racial and Gender Bias Problem | Time Magazine

[23] ISO/IEC 42001:2023 - AI management systems

[24] Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile

[25] AI Skills Gaps Threaten To Exacerbate Labor Shortages, Study Shows

[26] SEC.gov | Remarks at Program on Corporate Compliance and Enforcement Spring Conference 2024

Our expectations on oversight, principles, and approach to regulation:

Establish robust and transparent AI and data governance structures, ensuring appropriate expertise and clear lines of accountability.

There should be regular interaction between the board and those within the business with day-to-day responsibility for digital services and AI. The establishment of a technology ethics committee, chaired by a member of the board can help to ensure risks are appropriately mitigated. Boards should include directors with a strong understanding of evolving technologies and associated regulation in order to provide appropriate oversight and challenge.

Practical measures to enhance oversight include the carrying out of digital impact assessments, identification of clear metrics and establishment of reporting frameworks on digital governance that the board reviews, along with independent third-party audits of AI systems and digital infrastructure. Integration of digital governance targets into executive compensation can reinforce appropriate management alignment and accountability. In the case of AI specifically, which can introduce novel and complex risks, the need for appropriate expertise extends to model governance teams, which have direct oversight of AI model development and/or deployment. Model governance teams should include individuals that are able to provide not just technical but also legal/compliance and ethical perspectives. This is because what is technically possible may not always be complaint with regulations, and what is complaint with regulations may not always be considered ethical and could therefore expose companies to future regulatory or reputational risks.

Publish ethical AI and data governance principles addressing transparency and accountability; fairness and bias; privacy; and other salient risks.

The process of developing and agreeing on ethical AI and data governance principles helps to build a company's own internal understanding of how best to manage the associated risks. These principles should explain the structures for digital rights and AI governance, the ethical use principles to which a company adheres, examples of use cases, and explanations of how risks are identified and mitigated. For example, companies might consider the use of synthetic data to address bias resulting from unrepresentative datasets.

Integrate a culture of responsibility throughout the company, reinforced by training and regular solicitation of feedback from a range of stakeholders.

While ultimate accountability for the management of risks associated with digital services and AI should sit with a company's board and executives, an understanding of these risks and the company's approach to mitigating them should extend throughout the company. This can be achieved by having clear and accessible policies and through ongoing employee training and engagement (see "Digital rights and the workforce" section below).

Establish no-go areas – technologies and use cases that the company considers unethical or presenting an unacceptable level of risk.

Given the nascent and fragmented nature of regulation relating to digital services and AI, companies can reduce the risk of future regulatory infraction and reputational damage by proactively establishing 'red lines' for the development and deployment of technologies.

Companies should reflect on the potential pitfalls of new technologies and identify areas of unacceptable risk. For example, companies should commit to ensuring that users interacting with an AI tool will never be misled into believing that they are interacting with a human, and they should require the labelling of deepfakes to minimise disinformation risks. Companies should also consider whether to prohibit the use of facial recognition technology for real-time tracking of individuals in public spaces (a use that is already considered to be of unacceptable risk under the EU AI Act in most scenarios). A proactive approach to establishing no-go areas should support the building of a trusted brand and reduce the risk that companies fall foul of evolving regulation.

> **"A proactive approach to establishing no-go areas should support the building of a trusted brand and reduce the risk that companies fall foul of evolving regulation."**

Advocate for regulatory consistency across markets and disclose AI lobbying activities. Regulatory consistency will not always be achievable, due to divergent government priorities and agendas in different markets. However, advocating to reduce inconsistencies, along with adhering to a robust global approach should help minimise the risk that companies breach standards in any one country or region while conducting business internationally. This approach is also conducive to creating a transparent and level competitive playing field. There are important parallels with regulation of the financial industry, where for any given risk or issue companies may choose to adhere globally to the strictest rules of the multiple jurisdictions across which they operate. We encourage companies to report on their AI and data governance regulation lobbying activity, with the expectation that it should provide reassurance to investors that it aligns with a company's stated approach to responsible AI and data governance.

AI and data governance: examples of good practice

IBM's AI ethics board provides robust oversight to ensure applications maintain consistency with IBMs AI principles. The company is also active in the public policy space and has led the creation of an AI alliance board and contributed to the U.S. Department of Commerce's National Institute of Standards and Technology (NIST) AI Risk Management Framework[27].

DBS in Singapore has worked closely with the Monetary Authority of Singapore (MAS) to build an AI and responsible data use framework that includes governance across board, management and business unit levels, ensures regulatory consistency. The framework also evaluates the unintended bias against key stakeholders and ensures models and datasets undergo a compliance and audit assessment[28].

[27] 2023 ESG REPORT | IBM
[28] DBS' AI-Powered Digital Transformation | DBS Bank

Privacy and freedom of expression

Privacy and freedom of expression are human rights that may be put at greater risk by the emergence of new technologies. Companies collect, store, and use large quantities of data on their users.

Every day, companies receive requests from governments or other third parties that can impact the privacy or freedom of expression of their users. Companies face growing volumes of such requests, often from multiple countries and jurisdictions. Requests may be justified as tackling misinformation or seeking evidence against persons accused of crimes, but there is potential for misuse at the expense of the privacy and freedom of expression of individuals. In some jurisdictions, companies might also face direct access agreements, which are legal or technical agreements that enable governments to access data in bulk, without having to submit targeted requests.

Companies may also be faced with ethical and legal questions regarding their own approach to privacy and freedom of expression. Data collected from users it not always used simply for the fulfillment of service – it may be used to generate additional revenue through targeted advertising and other personalised offerings. Data can be further monetised if it is shared with third parties and some business models depend heavily on these functions, while others use data to generate revenue beyond their core purpose. Material differences in privacy laws across jurisdictions and the risk of incurring substantial penalties for regulatory breaches creates a need for robust policies and strong oversight of associated risks.

Obtain user consent for collection, storage, and utilisation of data, including targeted advertising, and ensure responsible use of facial recognition technology.

Companies should obtain user consent for their own collection, inference, sharing, and retention of data. The GDPR requires companies to do so and stipulates that consent must be "freely given, specific, informed and unambiguous." Many companies obtain user consent by having users check the terms and conditions. However, checking the terms and conditions box, especially when lengthy, may not constitute consent that is "freely given, specific, informed and unambiguous."

Companies should disclose the full range of purposes for which they collect, infer, share, and retain data, including core business purposes as well as additional commercialisation purposes. For consent to be "freely given, specific, informed and unambiguous", terms and conditions should be easy to find and understand for close to the entire user base, which may require supplementing written text with videos and images. Companies should take actions to grant users heightened access to and control over their data.

Where companies are deploying Facial Recognition Technology (FRT) responsibly, they should take action to address the related digital rights risks which include racial and gender biases observed within algorithmic systems; questionable accuracy and lack of public testing; possible privacy or legal violations in the sourcing of photos for databases; and misuse by some governments, law enforcement agencies or others. Companies should disclose the accuracy of their technology after measurement by a recognised and relevant scientific assessment institution; disclose the sources of their image databases and demonstrate that their technology is constantly monitored to detect algorithmic biases, particularly with respect to protected characteristics including race, gender, or age; demonstrate proper due diligence of clients before making the technology available to them; and demonstrate that effective grievance mechanisms are in place to enable victims to report consequences and to access remedies.

Maintain clear policies and processes for responding to data requests that can impact the privacy or freedom of expression of users.

Since 2020, EOS has endorsed the Global Network Initiative (GNI), the leading multistakeholder forum for accountability, shared learning, and collective advocacy on government and company policies and practices at the intersection of technology and human rights.

> **"We encourage companies to endorse the GNI or demonstrate policies and processes of equal strength and rigour to GNI guidance."**

Under GNI guidance, companies should:

- Follow established domestic legal processes by screening for requests that violate basic norms or unduly infringe upon privacy rights.
- Work with governments to promote adherence to the idea that restrictions on freedom of expression should not be imposed except in narrowly defined circumstances.
- Encourage governments to be specific, transparent, and consistent in their requests to restrict content or communication.

- Request clarification or modification, seek assistance from outside expertise, or challenge government requests in courts if they appear overbroad or unlawful.
- Maintain transparent policies and processes for responding to requests from governments or other third parties.
- Keep proper records and notify individuals impacted by requests, to the extent possible.

Ensure robust governance of cybersecurity risks.

Companies should ensure robust governance and policies over cybersecurity. The scale and frequency of breaches continues to rise, and the World Economic Forum consistently ranks cybersecurity as one of the top five risks to businesses.

> **"Breaches can cripple business operations, create legal and regulatory risks, and have adverse human rights impacts - particularly when sensitive personal data is compromised."**

Guidance published by the law firm Wachtell, Lipton, Rosen and Katz suggests that boards should not be involved in day-to-day risk management, but have oversight mechanisms informed by sufficient expertise, including the appointment of directors with experience in technology and ongoing director education on these matters. This continued training is necessary given the evolving nature of cybersecurity threats. Response strategies should cover all categories of likely scenarios, as well as unlikely but plausible scenarios with extreme consequences and appropriate and compliant disclosure should be made if systems are materially compromised.[29]

Companies should also consider how AI tools might be used to enhance their cybersecurity defenses, with recent studies demonstrating the effectiveness of general-purpose AI models at identifying cyber vulnerabilities.[30]

Privacy and freedom of expression - examples of good practice:

Alphabet discloses data on government requests to restrict content or communications, from which legal authorities, and for what purposes.

Microsoft states six principles for developing and deploying facial recognition technology and recently updated them to add a new Limited Access policy, remove AI classifiers of sensitive attributes, and bolster investments in fairness and transparency.

Apple publishes a transparency report twice a year about government requests for customer data in each country and how it responds.

[29] Cybersecurity Oversight and Defense | A Board and Management Imperative
[30] International AI Safety Report 2025 | UK AI Safety Institute

Negative societal impacts

As technology may be deployed faster than negative societal impacts are fully understood, risk management should be proactive rather than reactive. Companies must balance freedom of expression with obligations to remove problematic content as well as government demands, laws, and regulations imposing censorship. The commoditisation of data creates risks to privacy rights, which may be infringed upon by governments, malicious actors, or companies themselves.

> **"...companies that pursue a short-term revenue maximising strategy, at the expense of adequate moderation, may face a greater risk of incurring regulatory fines or impairment of their brand in the future."**

Companies should acknowledge and take steps to understand where their business models may generate or contribute to negative social impacts. They should also be transparent about findings, take actions to mitigate negative societal impacts, and cede the appropriate authority to regulators where appropriate.

The spread of hate speech, disinformation, and violent, racist, or extremist content online has necessitated content moderation, along with responsibility on companies to define these terms. The spread of problematic content on social media may be exacerbated by businesses for which short-term revenue maximization is driven by higher quantities of clicks, likes, posts, and shares.

Children & young people: Companies should prioritise children and young people, as well as other vulnerable populations, when addressing negative societal impacts. Doing so is likely to produce better outcomes for all, as one in three internet users is underage. Children and young people face heightened vulnerability to exploitation, cyberbullying, and other risks online. The long-term physical and mental effects of technology on children and young people are rarely studied or explored according to UNICEF.[31] However, public awareness of these risks is growing and the regulatory risk to companies is increasing, as policy makers respond with new legislation. The widespread adoption of generative AI, including the LLMs which typically drive chatbots or can be used to create deep fakes, can amplify the risks that children and young people face. Dr. Nomisha Kurian highlights that the 'empathy gap', or the inability of LLMs to have real-world context, can lead to dangerous outcomes for children, and that companies should consider how content is moderated and monitored, along with monitoring mechanisms and reporting systems.[32]

Content moderation and tackling disinformation: Companies have an essential role in tackling the societal risks associated with "fake news" and other harmful content. Neuroscientist Professor Gina Rippon has defined fake news as "pernicious information that's causing social mal-effect"[33], with the potential for disinformation to cause widespread harm apparent in the violent riots seen in parts of the UK during August 2024[34].

[31] Investigating Risks and Opportunities for Children in a Digital World | UNICEF

[32] Full article: 'No, Alexa, no!': designing child-safe AI and protecting children from the risks of the 'empathy gap' in large language models | Nomisha Kurian

In many countries, companies are granted broad powers and legal responsibilities for removing hate speech, false or misleading information, and violent, racist, or extremist content online. Companies should explain how they fulfill this role and allocate sufficient resources to personnel, including proper training and clear guiding principles.

Computer scientists, Mazurczyk, Lee and Vlachos, state that disinformation has become one of the most potent cybersecurity threats for businesses and society, adding that recent breakthroughs in AI have made it easier to make and disseminate fake content at scale.[35] They suggest that, in addition to companies investing in their own moderation and detection capabilities, they should educate users about disinformation and empower them to detect and report incidents.

Our expectations on mitigation of negative societal impacts:

Prioritise children and young people and promote child-safe AI in the development and deployment of AI models.

We expect companies to comply with the “safety-by-design” recommendations within the OECD Council on Children in the Digital Environment’s Guidelines for Service Providers:[36] These include enhanced privacy measures such as ensuring terms and conditions are accessible to children and young people; limiting data collection to the fulfilment of service; and refraining from profiling underage users without compelling reasons and appropriate safeguards in place. Companies should also consult the UN Convention on the Rights of the Child[37] in considering how their products and services may impact on children and young people. While companies should establish minimum age requirements, they should also acknowledge that younger users frequently interact

[35] Disinformation 2.0 in the Age of AI: A Cybersecurity Perspective | Mazurczyk, Lee and Vlachos (2023)

[36] OECD GUIDELINES FOR DIGITAL SERVICE PROVIDERS

[37] UN Convention on Rights of a Child (UNCRC) - UNICEF UK

with AI in the absence of adult supervision, and AI algorithms that might be used by children should be child-safe by design. Critically, risks to child users should be tested and considered on an ongoing basis after deployment. Companies should report on enforcement of protections and percentage of revenue derived from underage users.

Disclose how content is moderated and report actions taken.

Social media companies and any businesses that rely in some way on user generated content should implement transparent content moderation rules and report on enforcement. They should disclose processes and technologies used to identify potentially harmful content; report volume and nature of actions taken; and offer users clear appeals mechanisms. Companies should apply more stringent standards to and require visible labelling of content or accounts produced, disseminated, or operated with the assistance of automated software agents ("bots").

Automation, outsourcing, and other cost-cutting measures may necessitate additional oversight. Companies should ensure fair pay and mental health support is offered to content moderation workers, many of whom are situated in lower-income countries. While AI tools are able to facilitate the process of content moderation, partly alleviating the mental health impact on workers, the ongoing need for people to input into the data labelling and model training process, means that humans will remain "in the loop" for the foreseeable future. Failure to adequately support the mental health of moderation workers not only leads to unintended societal harms but may also result in legal action against a company.[38]

> **"Failure to adequately support the mental health of moderation workers not only leads to unintended societal harms but may also result in legal action against a company."**

Mitigation of negative societal impacts - examples of good practice:

Apple publishes a transparency report for its App Store, detailing the number of apps reviewed, approved, or rejected in accordance with its policies seeking to keep the App Store a safe and trusted place for users[40].

Tencent's ESG report describes how the company implements its safeguards for children and teen users, including its Children's Privacy Protection Statement, and how it helped significantly reduce game time by underaged players in its domestic market. [39]

Alphabet's current disclosures on building generative AI responsibly discuss and outline the risks related to misuse. The company conducts an AI product risk assessment that considers the scale, severity and likelihood of risk [41] [42].

[38] 'You can't unsee it': the content moderators taking on Facebook | Financial Times
[39] Tencent ESG Report 2023
[40] 2023 App Store Transparency Report 5-16-24
[41] Google's Secure AI Framework - Google Safety Center
[42] Google introduces AI Red Team

Mitigation of upstream environmental and social impacts

The global race for AI supremacy has spurred a massive expansion in data centre infrastructure, driven by companies aiming to enhance efficiency and consumer experience. This growth presents environmental and social challenges, as training and using AI models is highly energy-intensive. For example, a ChatGPT query is reported to use ten times more energy than a standard Google search query, and AI-related computing, which currently accounts for 10-20% of data centre energy use, is expected to rise significantly[43].

Additionally, data centers impact water resources, public infrastructure, and commercial real estate footprints.

Water Use

A single large data centre can use millions of litres of water daily. Google's data centers in the U.S consumed around 12.7 billion litres of fresh water in 2021 to keep their servers cool.[44] In addition to this, data centers draw electricity from power plants that use large cooling towers that convert water into steam.

Energy demand

As AI adoption surges, energy demand forecasts are being revised. The Federal Energy Regulatory Commission reports that grid planners expect a 4.7% increase in nationwide electricity demand over the next five years, up from 2.6% in 2022 estimates[45].

Goldman Sachs predicts data centres will account for 8% of total US power demand by 2030, up from 3% today, and that to support this growth, 47 GW of power generation capacity will be needed, with 60% forecast to be from gas power and 40% from renewable sources.[46]

> **"Goldman Sachs predicts data centres will account for 8% of total US power demand by 2030, up from 3% today."**

Impacts of the increased power demand:

CO2 emissions: Data centres' energy usage often exceeds the capacity of clean energy sources[47], while grid and flexibility investments fail to keep pace[48] . As energy production ramps up to meet demand, CO2 emissions are expected to rise.

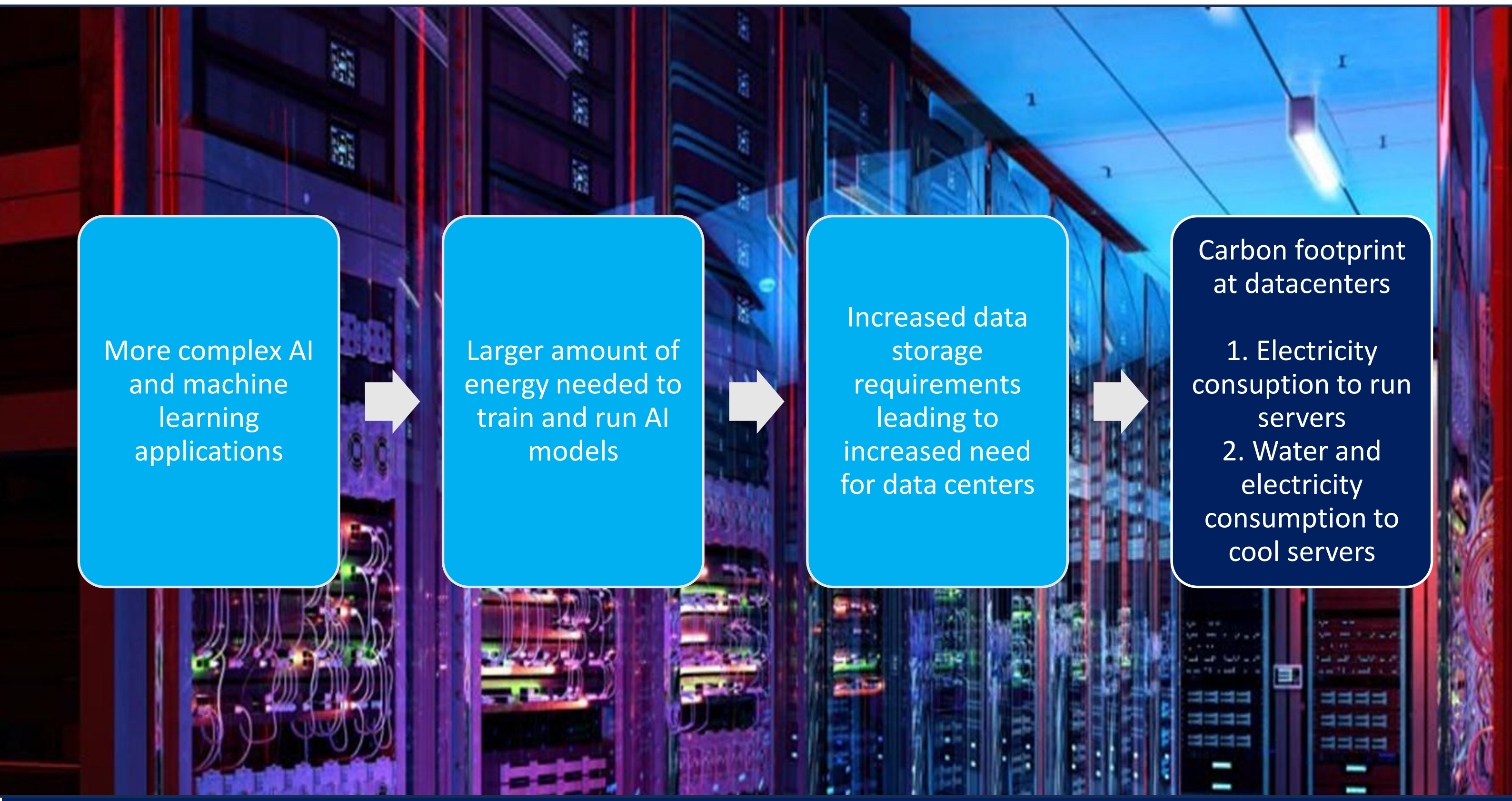
Coal retirement delays: Utilities are delaying coal plant retirements to meet data centre power needs. For example, North Omaha postponed a 2023 coal plant retirement to 2026 due to increased electricity demand from tech giants[49].

Local community impacts: Data centres often use diesel generators for backup power, causing localised pollution. They are frequently built in lower-income areas, leading to concerns about air and noise pollution.[50] Additionally, data centres have a large water footprint, using millions of litres daily, which can strain local water resources. Resources Center at Texas Tech University explains that the typical data center

[43] AI is poised to drive 160% increase in data center power demand | Goldman Sachs
[44] AI programs consume large volumes of scarce water | UCR News | UC Riverside
[45] US electricity load growth forecast jumps 81% led by data centers, industry: Grid Strategies | Utility Dive
[46] Generational Growth AI, Data Centers and the Coming US Power Demand Surge - Edward Conard
[47] DCC-PwC+Impact+Study.pdf
[48] AI execs who urgently need more energy to power their tech revolution are turning to fossil fuels | Business Insider India
[49] How Google and Meta data centers are keeping coal alive in Omaha - The Washington Post
[50] Environmental and Community Impacts of Large Data Centers - Gradient

consumes “the same amount of water as a city of 30,000-50,000 people.”[51]

Electricity rate increases: Rising data center power consumption may require over $2 trillion in energy generation resources, potentially increasing customer bills by 1% annually until 2032. [52]



More complex AI and machine learning applications

Larger amount of energy needed to train and run AI models

Increased data storage requirements leading to increased need for data centers

Carbon footprint at datacenters

1. Electricity consuption to run servers
2. Water and electricity consumption to cool servers

Our expectations on mitigation of upstream environmental and social impacts:

Embed awareness and evaluation of energy intensity and emissions at a model level, hardware level and firm level.

Addressing the emissions intensity of digital services, especially AI, requires a holistic approach involving the firms developing AI models, energy providers, utilities, data center infrastructure providers, real estate companies, and hardware developers. Awareness of energy use should be embedded from AI model inception through deployment to encourage energy-efficient solutions. For example, tools such as CarbonTracker[53], which predicts the carbon footprint of deep-learning models, are available to help model developers consider energy intensity, while for hardware, power capping (which involves limiting the percentage of maximum potential output at which any individual AI chip operates) can reduce server energy consumption by up to 15%.[54] All AI value chain players should engage suppliers on emissions reduction and, where relevant, consider Power Usage Effectiveness (PUE) metrics and emissions targets in due diligence for data centers and software as a service (SaaS) providers. All players of the AI value chain

[51] Drought-stricken communities push back against data centers

[52] Utilities Must Reinvent Themselves to Harness the AI-Driven Data Center Boom | Bain & Company

[53] GitHub - lfwa/carbontracker: Track and predict the energy consumption and carbon footprint of training deep learning models.

[54] The future of AI and energy efficiency | IBM

should be engaging with their suppliers on what it is doing to reduce emissions and whether targets have been set.

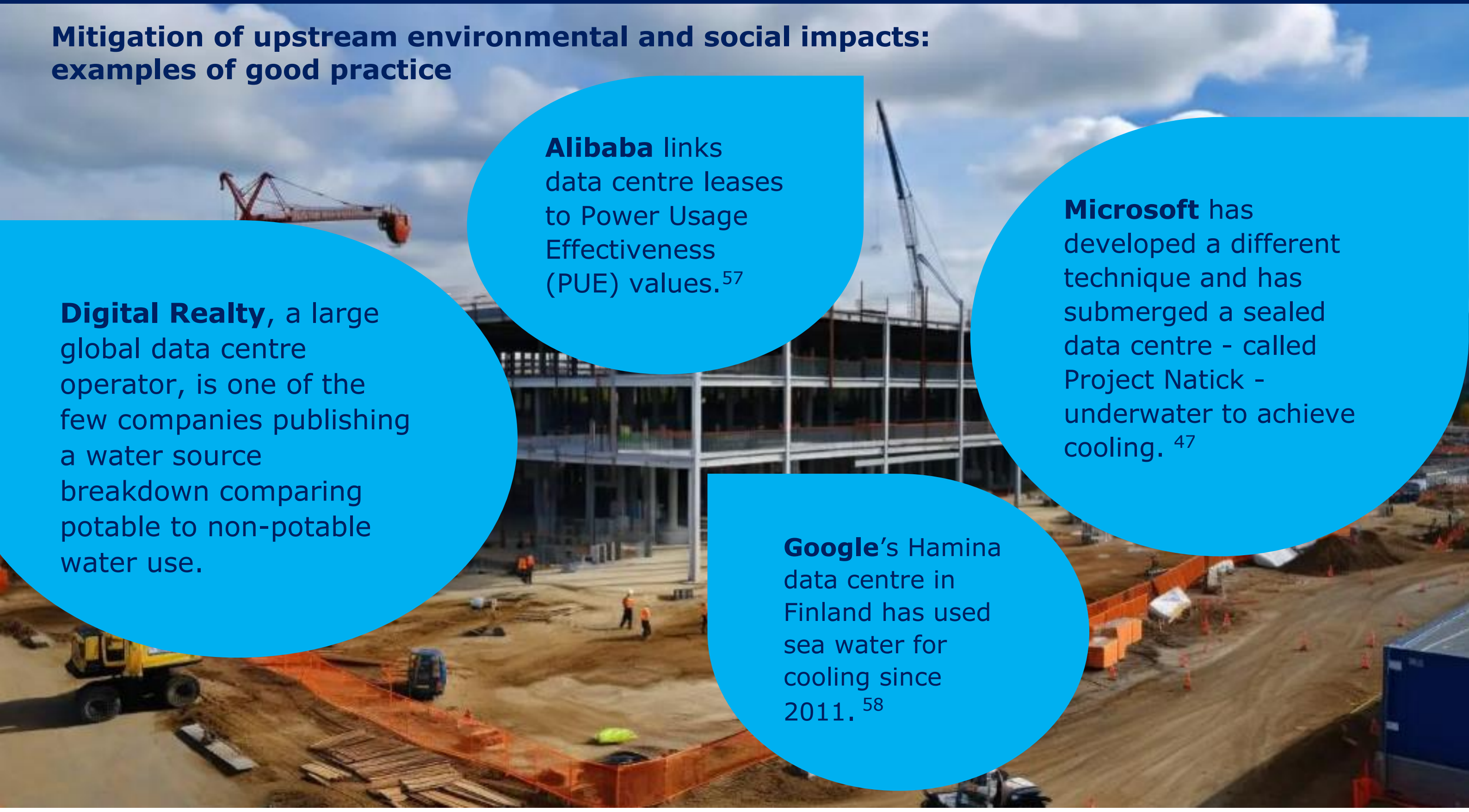
Encourage minimization of data center water intensity in own operations and/or service providers.

In addition to asking for PUE metrics, firms leasing data center space should evaluate water consumption measurements and incorporate these into vendor selection criteria.[55] For those owning or operating data centers, reducing water and power use can be achieved by optimising cooling operations through better ventilation, strategic equipment positioning, and proper insulation.[56] While the tech industry has made strides in publishing emissions data, it needs to do the same for water footprints. Operators and cloud vendors should calculate and publish detailed water usage, including potable water, to ensure transparency and community impact awareness.

Consider data centre design and location with a view to minimising negative environmental and social impacts.

Data centre developers should think strategically about data centre location if they are to achieve their emissions reduction targets. Locating data centres in carbon-intensive, under-invested grids is likely to have a greater adverse emissions impact than adding capacity to well-invested, flexible, and renewables-led grids.

Mitigation of upstream environmental and social impacts: examples of good practice



Digital Realty, a large global data centre operator, is one of the few companies publishing a water source breakdown comparing potable to non-potable water use.

Alibaba links data centre leases to Power Usage Effectiveness (PUE) values.[57]

Microsoft has developed a different technique and has submerged a sealed data centre - called Project Natick - underwater to achieve cooling. [47]

Google's Hamina data centre in Finland has used sea water for cooling since 2011. [58]

[55] Data centre water consumption | npj Clean Water

[56] Understanding Data Center Energy Consumption - C&C Technology Group

[57] 2024 Alibaba Group Environmental, Social and Governance Report-0809.pdf ((PUE values are a standardised measure of efficiency for power consumption in data centers and can be understood as the ratio of total energy used by a data centre to the energy used directly

[58] Data centre water consumption | npj Clean Water

Digital governance and the workforce

Digital technologies impact the workforce in many ways. AI and genAI in particular, is displacing certain jobs and changing many others. A recent IMF study revealed that AI will affect almost 40 percent of jobs around the world, replacing some and complementing others. In advanced economies, that number rises to 60 percent.[59] Also, digital technologies have increased the percentage of workers participating in the gig economy, defined as a labour market characterised by the prevalence of short-term contracts or freelance work as opposed to permanent jobs. AI is being deployed for basic human capital functions, and a survey by the Pew Research Center[60] found that most employees expect hiring, firing, and workplace assessment to be transformed by algorithms. Labour unions are becoming more vocal, and companies that don't respond appropriately to these challenges may be faced with reputational or legal impacts relating to their use of technology.

The impact of technology on the workforce is not a new debate. During the Industrial Revolution, similar concerns arose, but were in many instances assuaged, as a net positive impact on jobs and the economy proved to be sustainable. New technologies impact employment, wages, and working conditions through the displacement effect, in which they replace workers or suppress wages, and the productivity effect, in which they enhance workers' efficiency or create new jobs. Optimists assert that the productivity effect will offset the displacement effect as was the case during the industrial revolution and other technological advancements. We expect companies to "show" (demonstrate) rather than "tell" (claim) that this is the case.

[59] AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity. (imf.org)

[60] AI in Hiring and Evaluating Workers: What Americans Think | Pew Research Center

Our expectations on mitigation of upstream environmental and social impacts:

Encourage employee engagement on AI use case development and deployment and use the digital revolution as an opportunity to reinforce commitments to employees.

Companies that encourage employee engagement on responsible AI, as well as AI use case development and deployment, are likely to benefit from improved risk mitigation and identification of opportunities.[61] As well as engaging employees on the use of AI, companies should take measures to eliminate bias and ensure accuracy of algorithms deployed for recruitment, performance assessment and other human resources use cases.

> **"Companies that encourage employee engagement on responsible AI, as well as AI use case development and deployment, are likely to benefit from improved risk mitigation and identification of opportunities."**

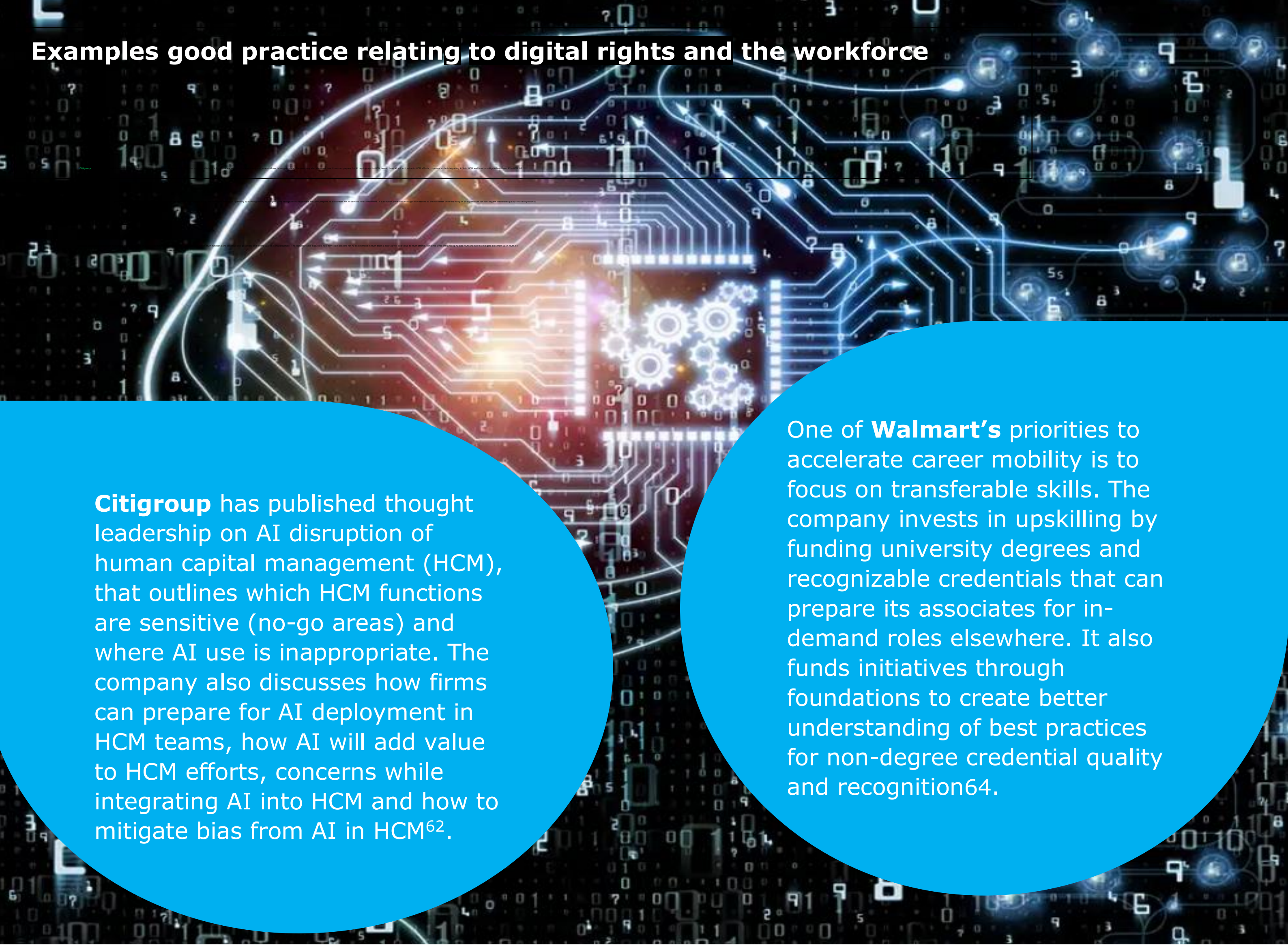
Our engagement plan states our expectation for companies to hire from the widest talent pool; and fair wages and benefits paid so all employees can afford a decent living standard, thereby promoting employee loyalty and productivity. Reinforcing or strengthening these commitments can help companies demonstrate the net positive benefits of AI on the workforce, and by extension, society. Furthermore, cognitive diversity is a prerequisite for identifying and eliminating unintended biases within algorithms.

Disclose anticipated impacts of AI deployment, retraining/reskilling initiatives aimed at supporting a "just AI transition".

To better inform investors, companies should disclose the types of skills its workforce will need to deliver the value of its investment in AI and automation/robotics tools to customers. This will require companies to evaluate whether the existing employee base has such skillsets, whether retraining is needed, or whether it anticipates cutting jobs, scaling back full-time opportunities, or prioritizing hiring for new skillsets. The company should be able to disclose its proportion of permanent and contracted workers and explain how it manages the benefits and risks of increasing reliance on contracted workers. Disclosure should provide quantitative and qualitative information about jobs displaced and other impacts to employment, wages, and working conditions; describe policies and practices for managing impacts such as ensuring workers are given sufficient notice and/or priority for other open positions; and demonstrate evidence of retraining, upskilling, and other forms of financial or technical support for workers impacted by the transition. Existing disclosure frameworks are beginning to consider these topics, for example, the Workforce Disclosure Initiative.

[61] Workers could be the ones to regulate AI (ft.com)

Examples good practice relating to digital rights and the workforce



Citigroup has published thought leadership on AI disruption of human capital management (HCM), that outlines which HCM functions are sensitive (no-go areas) and where AI use is inappropriate. The company also discusses how firms can prepare for AI deployment in HCM teams, how AI will add value to HCM efforts, concerns while integrating AI into HCM and how to mitigate bias from AI in HCM[62].

One of **Walmart's** priorities to accelerate career mobility is to focus on transferable skills. The company invests in upskilling by funding university degrees and recognizable credentials that can prepare its associates for in-demand roles elsewhere. It also funds initiatives through foundations to create better understanding of best practices for non-degree credential quality and recognition64.

Sector considerations

Sector use case: Technology

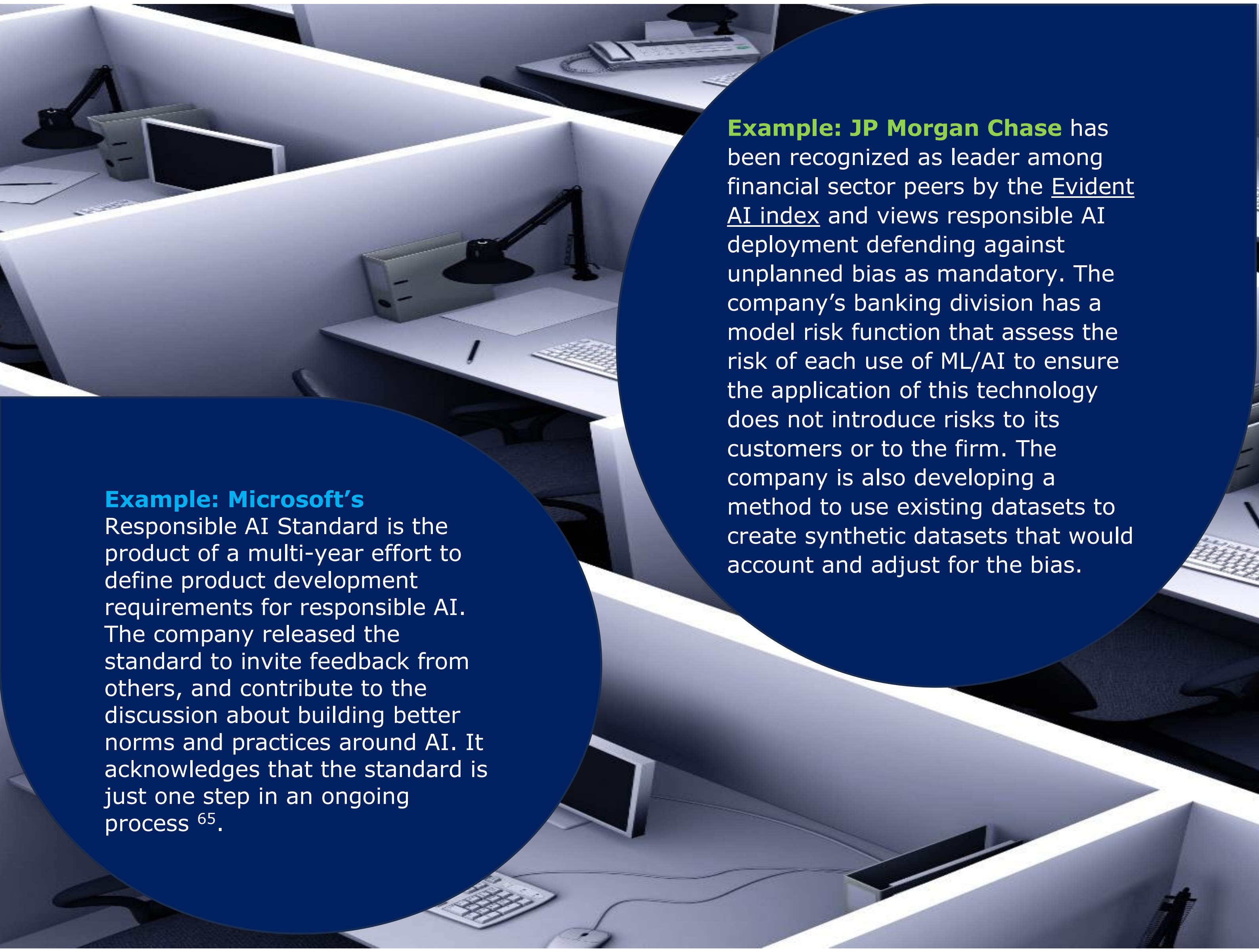
The technology sector is leading the development of AI itself while also using AI for numerous purposes. For example, social media companies use AI to curate, rank, and recommend online content, targeted advertising, search results, and political news. AI advances human development, but there is the potential for misuse. Tech companies are increasingly powerful in influencing users' behaviour or contributing to social segmentation, while exerting significant control over media consumed.[62] Recently, the US Surgeon General issued a new advisory on the effects of social media on youth mental health.[63] We expect companies to build trust in responsible AI through various actions. For example, we have asked companies to demonstrate that their business models do not incentivize problematic content and to ensure that human rights impact assessments cover all relevant digital products and services. We have also encouraged companies to enhance disclosure of the policies and processes they use to enforce child age restrictions and an assessment of their effectiveness.

[62] AI Disruption of human capital management
[63] Our People: Associate Experiences and Paths to Opportunity
[64] AI Disruption of human capital management

In the 2024 proxy season we saw a record number of AI related shareholder proposals at technology companies touching upon AI oversight in the audit committee charter, risks related to AI generated misinformation, human rights assessment of AI driven targeted as policies etc. These proposals received significant support, speaking to the continuously increasing focus on AI. A large number of technology companies have a growth and investment strategy that revolves around AI. As such, AI amplifies all other company opportunities and risks, which is why robust oversight is key.

Sector use case: Financial services

AI is being widely deployed all across the financial services market today. A number of specific applications include risk management, chatbots, virtual assistants, underwriting, fraud detection and algorithmic trading. One of the most concerning issues in AI deployment is the potential for gender and other forms of bias. We have engaged on this issue to gauge how companies are thinking about it. We have also asked companies to publish ethical AI principles that their AI models follow and to consider conducting a bias assessment.



Example: JP Morgan Chase has been recognized as leader among financial sector peers by the Evident AI index and views responsible AI deployment defending against unplanned bias as mandatory. The company's banking division has a model risk function that assess the risk of each use of ML/AI to ensure the application of this technology does not introduce risks to its customers or to the firm. The company is also developing a method to use existing datasets to create synthetic datasets that would account and adjust for the bias.

Example: Microsoft's Responsible AI Standard is the product of a multi-year effort to define product development requirements for responsible AI. The company released the standard to invite feedback from others, and contribute to the discussion about building better norms and practices around AI. It acknowledges that the standard is just one step in an ongoing process [65].

[65] Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf

Sector use case: Healthcare

Effective AI tools are heavily dependent upon good quality data, and this is especially true for AI tools and models in the healthcare sector. At the moment AI is being used to improve population health management, operational improvement and to strengthen innovation. Historically, the data available in the sector has not been representative, leading to limitations and a greater risk of gender and other bias. We are beginning to ask companies in the healthcare sector how they working to eliminate these risks. Our engagement has focused on:

1. AI governance
2. AI Bias assessments
3. AI use cases
4. Public policy alignment
5. Impact of the EU AI act on the company

We are seeing increasing litigation related to AI deployment. Most recently a lawsuit was filed against a healthcare company that alleged a faulty AI algorithm was used to deny elderly patients critical coverage and did not include a human in the decision making loop. This highlights the need for companies to do more to anticipate unintended negative impacts from the use of their products in an environment where regulation may fail to keep pace with technological developments and the range of use cases. A number of these use cases include high stakes domains, especially within healthcare, where companies need to be more prudent about ensuring adequate oversight.

Example: Gilead deploys two specific approaches to control the potential for biases in its use of AI: (i) to account fo the historic biases seen in many scientif and healthcare datasets, it in-licenses a uses data sets with a fair representatio of under-represented individuals when training AI algorithms; (ii) the clinical tr sites proposed by the AI algorithms are reviewed by the Clinical Trial Operation team to control for any potential biases that arise.

Example: GSK has developed a responsible AI use policy that includes explicit detail on which board members have oversight of AI usage and clarity on what reporting structures and procedures are in place for AI use. [67]

Conclusion

These EOS Digital Governance Principles are intended to provide a robust framework for companies to appropriately consider and manage the risks and opportunities associated with digital services and AI and, in doing so, protect long-term value. We welcome feedback from all stakeholders, recognising that this is a fast-changing and dynamic space and that understandings of best practice will evolve over time.

We believe that companies should recognise not only the potential for value creation through new technologies, but also the necessity of addressing unintended harms, and that this dual focus is essential for sustainable long-term growth. We view the establishment of robust governance structures, ethical principles, and clear lines of accountability as crucial. Companies should embed a culture of responsibility, ensuring that ethical considerations are integral to their operations. We also encourage companies to go beyond simple compliance with current regulations, and to anticipate future regulatory developments and contribute constructively to public policy.

Protecting privacy and freedom of expression is a fundamental aspect of digital governance. Companies must obtain user consent for data collection and use, maintain clear policies for responding to data requests, and ensure robust cybersecurity measures. Transparency in these areas builds trust and aligns with ethical governance principles.

The scale of the environmental and social impacts of digital services, particularly AI, are increasingly evident. Where relevant, companies should consider the energy and water consumption of data centres, the carbon footprint of AI models, and the broader societal impacts of their digital value chains. Mitigating these impacts is important to maintaining a social license to operate and reducing regulatory risks.

AI and digital technologies are transforming the workforce, creating both opportunities and challenges. We encourage companies to support a "just AI transition" by engaging employees in AI development, ensuring fair and unbiased use of AI in human capital management, and providing retraining and reskilling opportunities.

Finally, different industries face unique challenges and opportunities related to digital governance and this means that companies should adopt sector-specific lenses to effectively manage these issues.

For professional investors only. This is a marketing communication. Hermes Equity Ownership Services ("EOS") does not carry out any regulated activities. This document is for information purposes only. It pays no regard to any specific investment objectives, financial situation or particular needs of any specific recipient. EOS and Hermes Stewardship North America Inc. ("HSNA") do not provide investment advice and no action should be taken or omitted to be taken in reliance upon information in this document. Any opinions expressed may change. This document may include a list of clients. Please note that inclusion on this list should not be construed as an endorsement of EOS' or HSNA's services. EOS has its registered office at Sixth Floor, 150 Cheapside, London EC2V 6ET. HSNA's principal office is at 1001 Liberty Avenue, Pittsburgh, PA 15222-3779. Telephone calls may be recorded for training and monitoring purposes.

Federated Hermes

EOS

Federated Hermes

Federated Hermes is a global leader in active, responsible investing.

Guided by our conviction that responsible investing is the best way to create long-term wealth, we provide specialised capabilities across equity, fixed income and private markets, multi-asset and liquidity management strategies, and world-leading stewardship.

Our goals are to help people invest and retire better, to help clients achieve better risk-adjusted returns and, where possible, to contribute to positive outcomes that benefit the wider world.

Our investment and stewardship capabilities:

- **Active equities:** global and regional
- **Fixed income:** across regions, sectors and the yield curve
- **Liquidity:** solutions driven by five decades of experience
- **Private markets:** private equity, private credit, real estate, infrastructure and natural capital
- **Stewardship:** corporate engagement, proxy voting, policy advocacy

Why EOS?

EOS enables institutional shareholders around the world to meet their fiduciary responsibilities and become active owners of public companies. EOS is based on the premise that companies with informed and involved shareholders are more likely to achieve superior long-term performance than those without.

For more information, visit **www.hermes-investment.com** or connect with us on social media: in X