redwheel

Greenwheel Insights

Responsible AI Framework

Jessica Wan
Social Research Lead, Greenwheel

Merle Jacobsen
Biodiversity and human rights fellow, Greenwheel

Executive Summary

There is a growing consensus on the definition of responsible AI based on international norms, regulations, and industry best practice. In a nutshell, responsible AI is balancing innovation with the necessary management of risks that underlie the development and deployment of AI products and services. Responsible AI requires a unique blend of human rights due diligence and product safety.

However, in practice, innovation is outpacing the development of safeguards and responsible AI practices. To date, there are more than 3000 recorded incidents of AI-related harms. Because regulations are lagging behind advances in AI systems as well as emerging risks and impacts due to unforeseen risks and misuse, businesses face the difficult task of applying international norms to new technologies.

The good news is that there are comprehensive guidelines and frameworks to help implement responsible AI principles. Though, for investors, existing frameworks are not tailored to their specific needs; they require "translation" into actions that investors can take in their pre- and post-investment due diligence processes.

To support investors in applying responsible AI principles, Greenwheel has developed an investor framework drawing on recommendations from international organisations, regulations, human rights experts, and company best practices. This tool translates responsible AI principles into the six due diligence steps that investee companies should take:

1. **Embed**: Uphold responsible AI policies and supply chain policies; define roles and responsibilities; and provide training on AI ethics, human rights, and safety;

2. **Identify**: Map regulatory changes and expectations; map human rights risks across the AI value chain; determine risk levels of AI systems;

3. **Address**: Implement safeguards to prevent and mitigate risks; conduct AI audits; engage with stakeholders (academics, civil society organisations, human rights experts, governments);

4. **Remediate**: Establish an effective operational grievance mechanisms; provide remedies to affected individuals and groups;

5. **Track:** Track and document AI system performance and impacts throughout the AI lifecycle; and,

6. **Report**: Openly communicate on company policies, and processes, including performance of AI systems and lessons learned.

CONTACT US

Please contact us if you would like to learn more about our strategies.

invest@redwheel.com | www.redwheel.com

redwheel

Preface: The Investor Need

A year ago, Greenwheel held an AI consultation lunch with representatives from all investment teams at Redwheel. A common concern raised by investment teams is how to balance the benefits we can maximise from AI while managing the foreseeable and unforeseeable risks and impacts.

We are only beginning to see the transformations from AI, especially for deployers. One of the key learnings we have from previous technological booms, including social media, is that innovation often outpaces regulations and best practices. Due diligence is often an afterthought, which can expose investors to human rights and environmental risks.

To ensure that our investment teams stay ahead of the regulatory requirements and normative guidance around AI, we have tasked Greenwheel with developing a Responsible AI Framework to help investors ensure that the relevant safeguards are put in place for both developers and deployers. Given the rapid changes we are witnessing, we anticipate for this Framework to continue evolving over time.

Arthur Grigoryants

Head of Investments

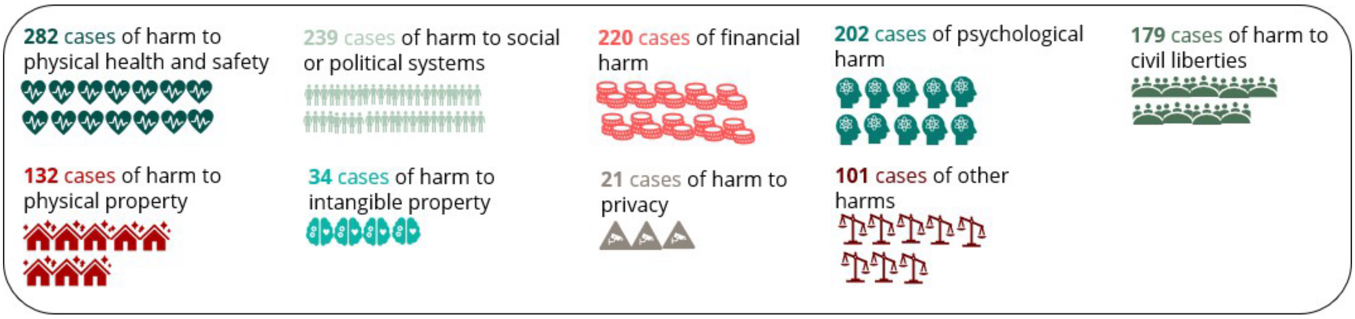
Defining responsible artificial intelligence

Artificial Intelligence (AI) systems are rapidly advancing. Already, **AI systems have facilitated the enjoyment of human rights**, leading to positive social outcomes across many sectors from health care to financial services in both developed and emerging and frontier markets.[1] For workers, the **AI value chain offers new job opportunities**; the deployment of AI is augmenting and transforming roles, freeing workers from repetitive tasks. For end-users, AI systems have spurred creativity and promoted access to information across the globe.[2]

However, **the adoption of AI tools is accompanied by human rights risks and ethical concerns**.[3] There are human rights risks and impacts identified across the AI value chain to workers, communities, and end-users. To date, the OECD Incidents Database shows that **there are over 3000 reports of AI-related harms** (Figure 1). In addition, the benefits are not distributed equitably as some communities remain underserved and excluded from the advancements of AI.

Businesses and investors are presented a unique **challenge in balancing innovation with the necessary management of risks that underlie the development and deployment of AI systems**. With the recent eruption of Generative AI (GenAI) and its unprecedented scale and speed in uptake, concerns are raised by academics, civil society, international organisations, journalists, policymakers, and trade unions. There is a growing demand for a more human-centred approach to AI development and deployment, including the positive role businesses and investors can and should play as part of their responsibility under international norms.[4]

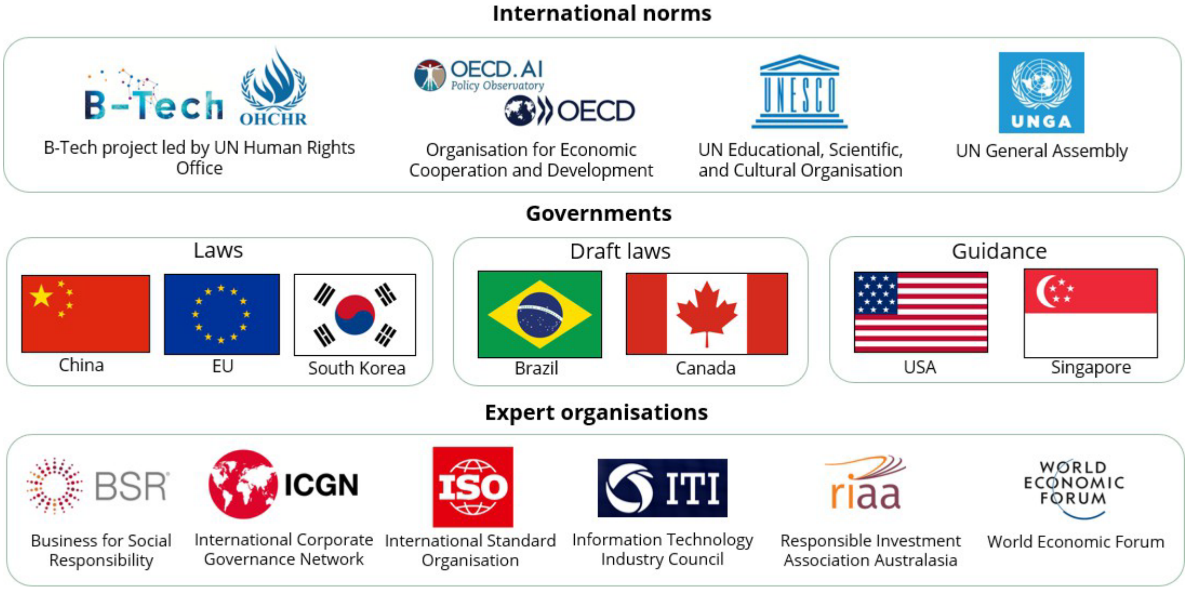
Figure 1: AI-induced harm



Source: AI Incident Database, 2025; created by Greenwheel.

To address the risks of AI systems and to maximise the potential benefits for society, **international organisations, governments, human rights experts, businesses, and investors have developed recommendations for the responsible development and deployment of AI** (Figure 2).

Figure 2: Norms and standards on responsible AI


International norms
B-Tech
OHCHR
B-Tech project led by UN Human Rights Office
OECD.AI Policy Observatory
OECD
Organisation for Economic Cooperation and Development
UNESCO
UN Educational, Scientific, and Cultural Organisation
UNGA
UN General Assembly
Governments
Laws
China
EU
South Korea
Draft laws
Brazil
Canada
Guidance
USA
Singapore
Expert organisations
BSR
Business for Social Responsibility
ICGN
International Corporate Governance Network
ISO
International Standard Organisation
ITI
Information Technology Industry Council
riaa
Responsible Investment Association Australasia
WORLD ECONOMIC FORUM
World Economic Forum

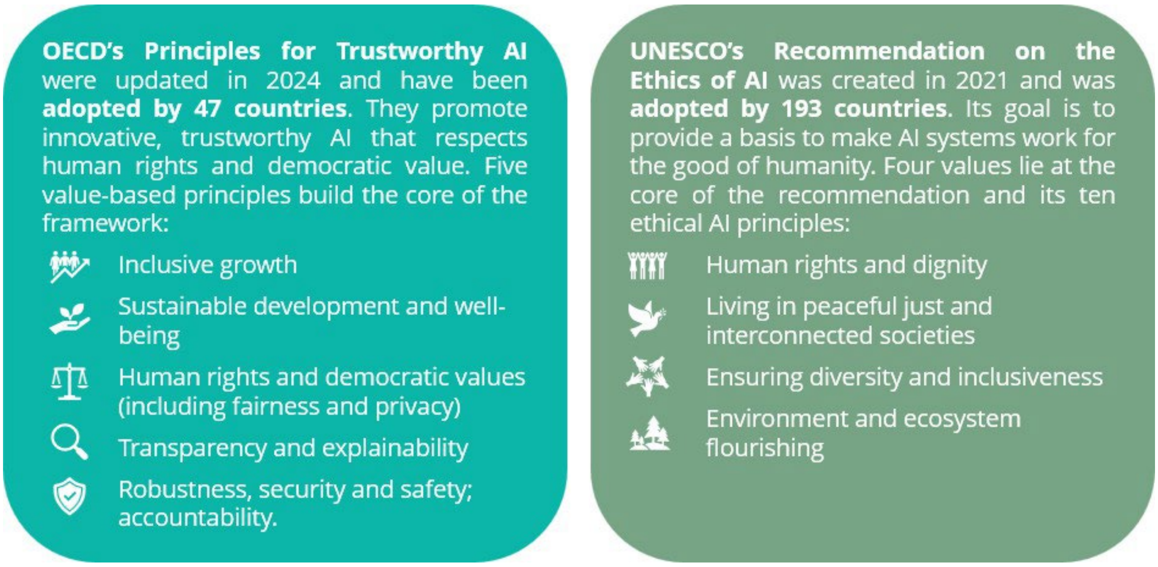
Source: UNESCO, 2021, B-Tech, 2023, NIST, 2023, ICGN, 2024, ITI, 2024, RIAA, 2024, South Korean AI Basic Law, 2024, AI Verify Foundation, 2024, UNGA, 2024, WEF, 2024, Yang, 2024, EU AI Act, 2025, Government of Canada, 2025, BSR, 2025, ISO, 2025, and, OECD, 2025; created by Greenwheel.

International norms

The **OECD Principles for Trustworthy AI and the UNESCO Recommendation on the Ethics of AI are the two key international standards on responsible AI for policymakers and businesses** (Figure 3).[5] Both sets of principles provide a definition of responsible AI through the incorporation of a human rights approach for AI actors across the AI lifecycle as a guidance to both policymaker and businesses.

In 2024, the United Nations General Assembly adopted a resolution on the promotion and protection of human rights in the design, development, deployment and the use of AI. The resolution reaffirms international norms around human rights and sustainable development and calls upon all stakeholders, both policymakers and private sector actors (businesses and investors), to work together in ensuring that AI systems are safe, secure, and trustworthy.[6]

Figure 3: Two key international principles on responsible AI



OECD's Principles for Trustworthy AI were updated in 2024 and have been **adopted by 47 countries**. They promote innovative, trustworthy AI that respects human rights and democratic value. Five value-based principles build the core of the framework:

- Inclusive growth
- Sustainable development and well-being
- Human rights and democratic values (including fairness and privacy)
- Transparency and explainability
- Robustness, security and safety; accountability.

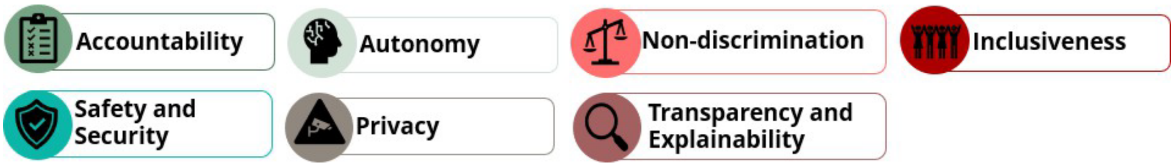
UNESCO's Recommendation on the Ethics of AI was created in 2021 and was **adopted by 193 countries.** Its goal is to provide a basis to make AI systems work for the good of humanity. Four values lie at the core of the recommendation and its ten ethical AI principles:

- Human rights and dignity
- Living in peaceful just and interconnected societies
- Ensuring diversity and inclusiveness
- Environment and ecosystem flourishing

Source: OECD, 2024 and UNESCO, 2021; created by Greenwheel.

Although there are some differences in how principles are presented across different global standards and guidances, **Greenwheel identified seven responsible AI principles that are commonly found** (Figure 4).

Figure 4: A global definition of responsible AI



Source: UNESCO, 2021, UN B-Tech and UNOHCHR, 2023, Council of Europe, 2024, OECD, 2024, and ISO, 2025; created by Greenwheel.

1. **Accountability: Effective governance structures and mechanisms should be in place to hold individuals, organisations, or entities accountable for ensuring the proper functioning of AI systems** and the responsible development and deployment of AI. Good AI governance ensures that an organisation's deployment of AI aligns with its strategies, objectives, and values.[7] This requires a clear set of rules, practices, processes, as well as roles and responsibilities.

2. **Autonomy**: Activities throughout the **AI system lifecycle should respect human autonomy** where humans are not treated as a "means-to-an-end".[8] This respect is upheld in three ways. Firstly, AI systems should be designed in a way that individuals can make choices and decisions free from manipulation, misinformation, and recommendation systems. In short, human decision-making is protected and respected.[9] Secondly, AI systems should avoid the *dehumanisation* of individuals or groups by reducing them to mere data points. Finally, the anthropomorphism of AI (e.g., projecting human qualities onto AI) should be avoided.[10]

3. **Non-discrimination**: AI actors should ensure that **AI systems and data do not reinforce existing biases or produce discriminatory outcomes** that disproportionately affect marginalised, vulnerable, or underrepresented individuals and groups. If such issues arise, they should be disclosed and further prevented by implementing appropriate safeguards.

4. **Inclusiveness**: The benefits of **AI systems should be accessible and inclusive**. Both the design and deployment of AI systems should consider the diverse needs of different groups such as ability, age, culture, gender, and language, with particular attention paid to marginalised and vulnerable communities. AI benefits should extend to users in both developed and emerging markets.

5. **Safety and security**: AI systems should be **robust, secure and safe throughout their entire lifecycle**. Developers and deployers are responsible for ensuring the appropriate functioning of AI systems under all use cases including normal, foreseeable or unforeseeable use or misuse. AI systems should avoid posing unreasonable safety or security risks for users and other impacted groups. Mechanisms should be in place to address harm caused or undesired behaviour, for instance, tackling disinformation and misinformation.

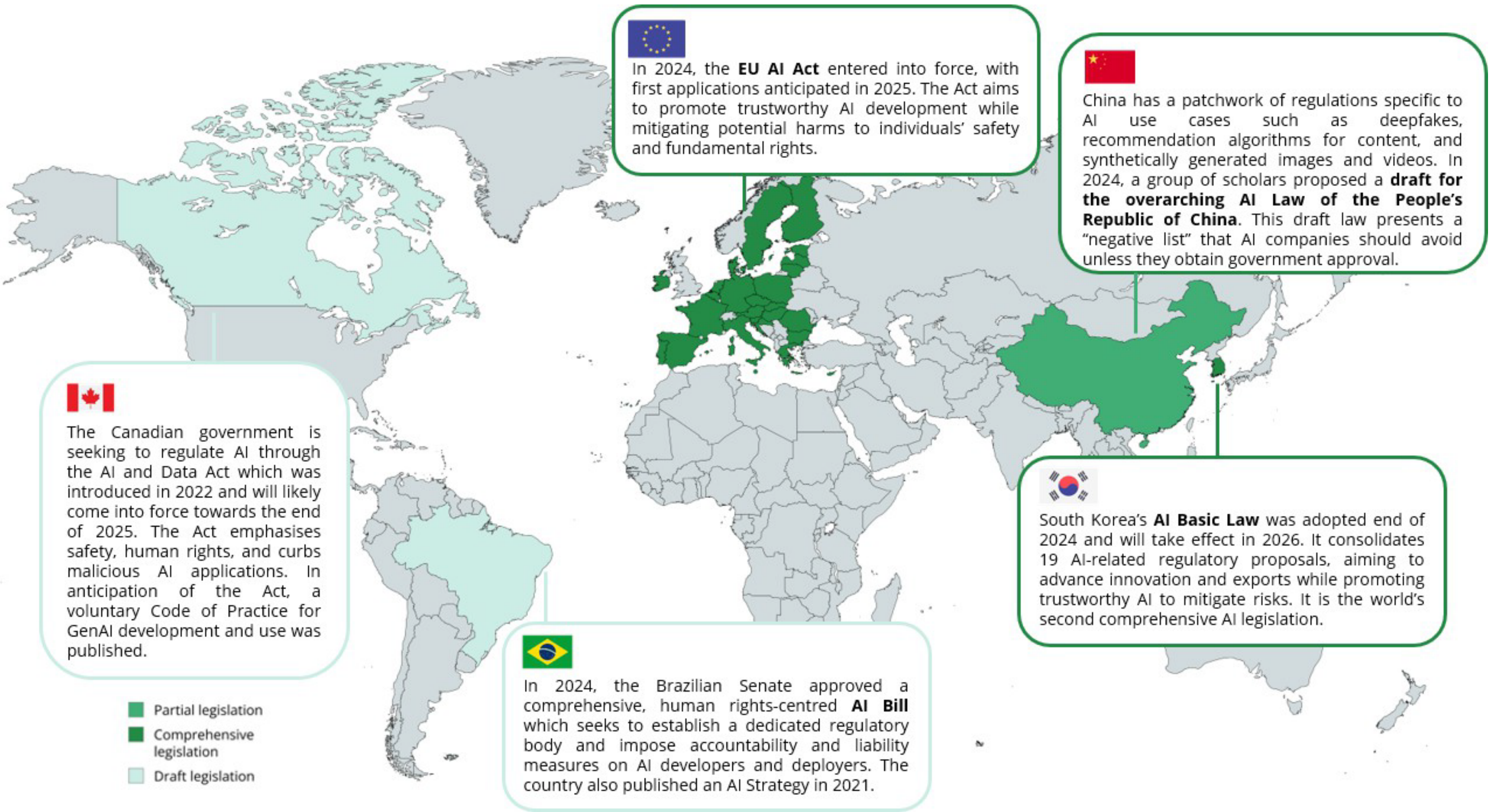
6. **Privacy**: AI systems should **safeguard personal and sensitive data**. Individuals should have control over how their information is collected, used, and disposed. Consent should be obtained by individuals *prior to the use* of their data in the training of AI models. In the case of GenAI, safeguards must be in place to prevent the misuse of generated content that may violate intellectual property rights.

7. **Transparency and explainability**: Finally, **governance structures, information on AI capabilities, datasets, models, limitations, and other factors influencing AI decision-making processes, should be well-documented, understandable and accessible** to relevant AI actors and stakeholders. At all times, users should be informed that they are interacting with AI. Those affected by AI outcomes should have access to simple and understandable explanations on how decisions were made and where they are able to challenge the results. However, transparency should be balanced with other competing demands such as protecting intellectual property (e.g., proprietary codes and datasets).

Legislation and guidance

Globally, **governments are developing regulatory frameworks to balance AI innovations with safeguards to protect human rights** (Figure 5).

Figure 5: National legislation and draft legislation on AI



In 2024, the **EU AI Act** entered into force, with first applications anticipated in 2025. The Act aims to promote trustworthy AI development while mitigating potential harms to individuals' safety and fundamental rights.

China has a patchwork of regulations specific to AI use cases such as deepfakes, recommendation algorithms for content, and synthetically generated images and videos. In 2024, a group of scholars proposed a **draft for the overarching AI Law of the People's Republic of China**. This draft law presents a "negative list" that AI companies should avoid unless they obtain government approval.

The Canadian government is seeking to regulate AI through the AI and Data Act which was introduced in 2022 and will likely come into force towards the end of 2025. The Act emphasises safety, human rights, and curbs malicious AI applications. In anticipation of the Act, a voluntary Code of Practice for GenAI development and use was published.

South Korea's **AI Basic Law** was adopted end of 2024 and will take effect in 2026. It consolidates 19 AI-related regulatory proposals, aiming to advance innovation and exports while promoting trustworthy AI to mitigate risks. It is the world's second comprehensive AI legislation.

In 2024, the Brazilian Senate approved a comprehensive, human rights-centred **AI Bill** which seeks to establish a dedicated regulatory body and impose accountability and liability measures on AI developers and deployers. The country also published an AI Strategy in 2021.

Source: EBIA, 2021, Brazil AI Act, 2024, CSET, 2024, IAPP, 2024, South Korean AI Basic Law, 2024, Yang, 2024, EU AI Act, 2025, Government of Canada, 2025a, and Government of Canada, 2025b; made with MapChart; created by Greenwheel.

China is the first country in the world to develop a set of regulations around AI. Supplementing the sets of regulations, the Chinese government has issued guidelines regarding the ethical use of AI and in addressing security risks. There are explicit references to anticipating the abuse and misuse of AI; prohibiting the infringement of basic human rights (personal, privacy, and property rights); addressing discrimination and biases; and, ensuring roles and responsibilities are appropriately assigned.[11] Additionally, there are supplementary guidance for the application of AI technology at the sectoral level (e.g., automobile, financial services, health care).[12]

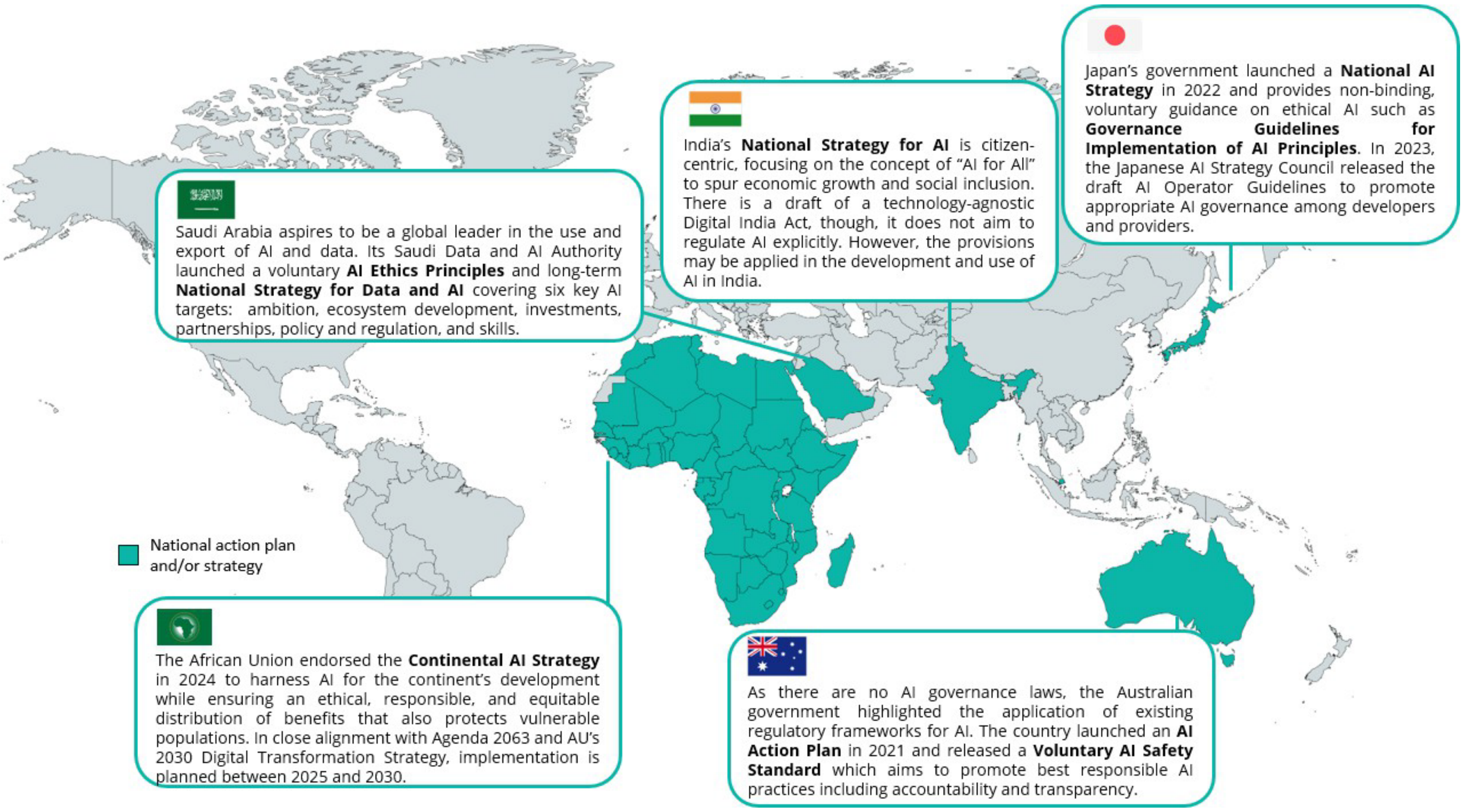
The European Union is the first to launch a comprehensive AI regulation. Its scope extends beyond the geographical boundaries of the EU to include any AI system used within the EU as well as systems that affect individuals in the EU regardless of where the system is developed or deployed.[13]

The EU AI Act takes a product safety approach where products must meet a minimum set of safety requirements prior to rollout in European markets. This Act balances the level of compliance burden for companies and protection of rights, particularly for small and medium enterprises.[14] The EU AI Act identifies "high-risk" AI systems: biometrics, critical infrastructure, education, employment, access to essential services (public and private), law enforcement, immigration, and administration of justice and democratic processes.[15] Depending on the risk-level of a given AI system, there are different mandatory requirements in the adoption and maintenance of a quality and risk management system.[16]

Brazil's draft AI Bill mirrors the EU AI Act in establishing a risk-based approach to AI systems. AI systems deemed to pose excessive risks are prohibited. High-risk systems are heavily regulated. All other systems have to meet a basic set of requirements. However, in contrast, to the EU AI Act, the Brazilian draft Act provides a clearly defined set of values for companies to adopt in their AI policies.[17] Brazilian civil society argues that their draft legislation is more explicit and aligned with human rights due diligence compared to the EU regulation.[18]

The regulatory landscape is expected to continue evolving swiftly in the next years. **Across all regions of the world, there are national and regional level strategies in addition to guidance on the ethical development and deployment of AI systems** (Figure 6).

Figure 6: National strategy and voluntary guidance on AI



Saudi Arabia aspires to be a global leader in the use and export of AI and data. Its Saudi Data and AI Authority launched a voluntary **AI Ethics Principles** and long-term **National Strategy for Data and AI** covering six key AI targets: ambition, ecosystem development, investments, partnerships, policy and regulation, and skills.

India's **National Strategy for AI** is citizen-centric, focusing on the concept of "AI for All" to spur economic growth and social inclusion. There is a draft of a technology-agnostic Digital India Act, though, it does not aim to regulate AI explicitly. However, the provisions may be applied in the development and use of AI in India.

Japan's government launched a **National AI Strategy** in 2022 and provides non-binding, voluntary guidance on ethical AI such as **Governance Guidelines for Implementation of AI Principles**. In 2023, the Japanese AI Strategy Council released the draft AI Operator Guidelines to promote appropriate AI governance among developers and providers.

The African Union endorsed the **Continental AI Strategy** in 2024 to harness AI for the continent's development while ensuring an ethical, responsible, and equitable distribution of benefits that also protects vulnerable populations. In close alignment with Agenda 2063 and AU's 2030 Digital Transformation Strategy, implementation is planned between 2025 and 2030.

As there are no AI governance laws, the Australian government highlighted the application of existing regulatory frameworks for AI. The country launched an **AI Action Plan** in 2021 and released a **Voluntary AI Safety Standard** which aims to promote best responsible AI practices including accountability and transparency.

Source: NITI, 2018; Saudipedia, 2020, Australian Government, 2021, Expert Group on AI Principle Implementation, 2022, NICT, 2022, SDAIA, 2023, AI Verify Foundation, 2024, AU, 2024, Australian Government, 2024, CIPIT, 2024, Clifford Chance, 2024, IAPP, 2024, and White Case, 2024; made with MapChart; created by Greenwheel.

Although there are differences in the various legislation, strategies, and guidance, there are some common themes:

Safety versus innovation: Governments recognise the potential trade-off between safety and innovation. There is a concern that stringent requirements on transparency and explainability may adversely impact and burden smaller AI actors such as small and medium enterprises.[19]

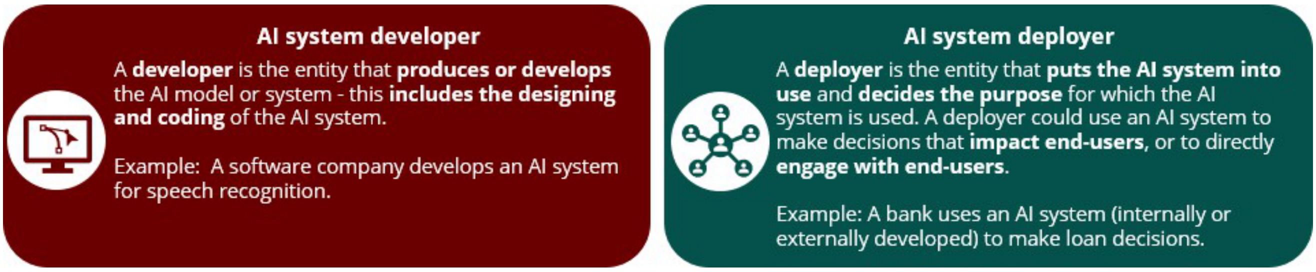
A value chain approach: Similar to the recommendations in international norms, the regulations, strategies, and voluntary guidance are relevant to all actors across the AI value chain. Though, the actors identified and classified across regulations may differ. For instance, the EU AI Act refers to providers, deployers, importers, distributors, and product manufacturers of AI systems while Canada's draft AI and Data Act references AI designers, developers, deployers and operators. Meanwhile, the OECD refers to developers and vendors.

A tiered approach to risks: Not all AI systems are equal. Some systems are higher risks due to the nature of the solution (e.g., biometrics) or where solutions are deployed (e.g., access to essential services). Consequently, the onerousness of risk management systems should be proportional to the risks posed to individuals and society.

Human rights impact recognised: Governments clearly recognise the risks of AI systems to human rights and the rule of law, which necessitates a risk management system.[20] While the prominence of human rights varies across legislation and strategies, there is universal recognition of the impact of discriminatory outcomes and biases.

The case for responsible AI for investors

Figure 7: AI Developers and deployers



AI system developer

A **developer** is the entity that **produces or develops** the AI model or system - this **includes the designing and coding** of the AI system.

Example: A software company develops an AI system for speech recognition.

AI system deployer

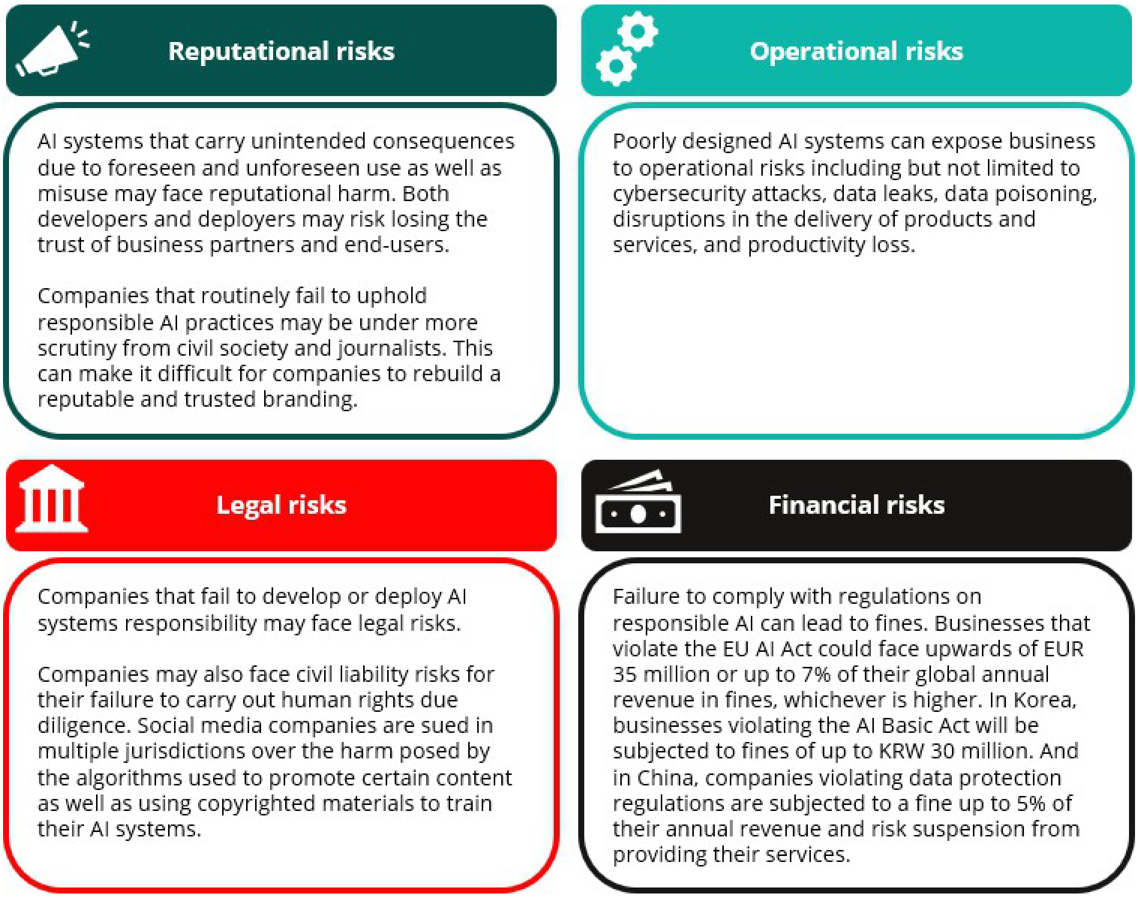
A **deployer** is the entity that **puts the AI system into use** and **decides the purpose** for which the AI system is used. A deployer could use an AI system to make decisions that **impact end-users**, or to directly **engage with end-users**.

Example: A bank uses an AI system (internally or externally developed) to make loan decisions.

Source: BSA, 2023 and ITI, 2024; created by Greenwheel.

Investors may be exposed to human rights risks across their portfolios if their holding companies fail to uphold the principles of responsible AI. Exposure can stem from companies developing and/or deploying AI systems (Figure 7). Failure to uphold responsible AI practices can carry significant human rights risks, which can also be financially material (Figure 8).[21]

Figure 8: The investor case for responsible AI



Reputational risks

AI systems that carry unintended consequences due to foreseen and unforeseen use as well as misuse may face reputational harm. Both developers and deployers may risk losing the trust of business partners and end-users.

Companies that routinely fail to uphold responsible AI practices may be under more scrutiny from civil society and journalists. This can make it difficult for companies to rebuild a reputable and trusted branding.

Operational risks

Poorly designed AI systems can expose business to operational risks including but not limited to cybersecurity attacks, data leaks, data poisoning, disruptions in the delivery of products and services, and productivity loss.

Legal risks

Companies that fail to develop or deploy AI systems responsibility may face legal risks.

Companies may also face civil liability risks for their failure to carry out human rights due diligence. Social media companies are sued in multiple jurisdictions over the harm posed by the algorithms used to promote certain content as well as using copyrighted materials to train their AI systems.

Financial risks

Failure to comply with regulations on responsible AI can lead to fines. Businesses that violate the EU AI Act could face upwards of EUR 35 million or up to 7% of their global annual revenue in fines, whichever is higher. In Korea, businesses violating the AI Basic Act will be subjected to fines of up to KRW 30 million. And in China, companies violating data protection regulations are subjected to a fine up to 5% of their annual revenue and risk suspension from providing their services.

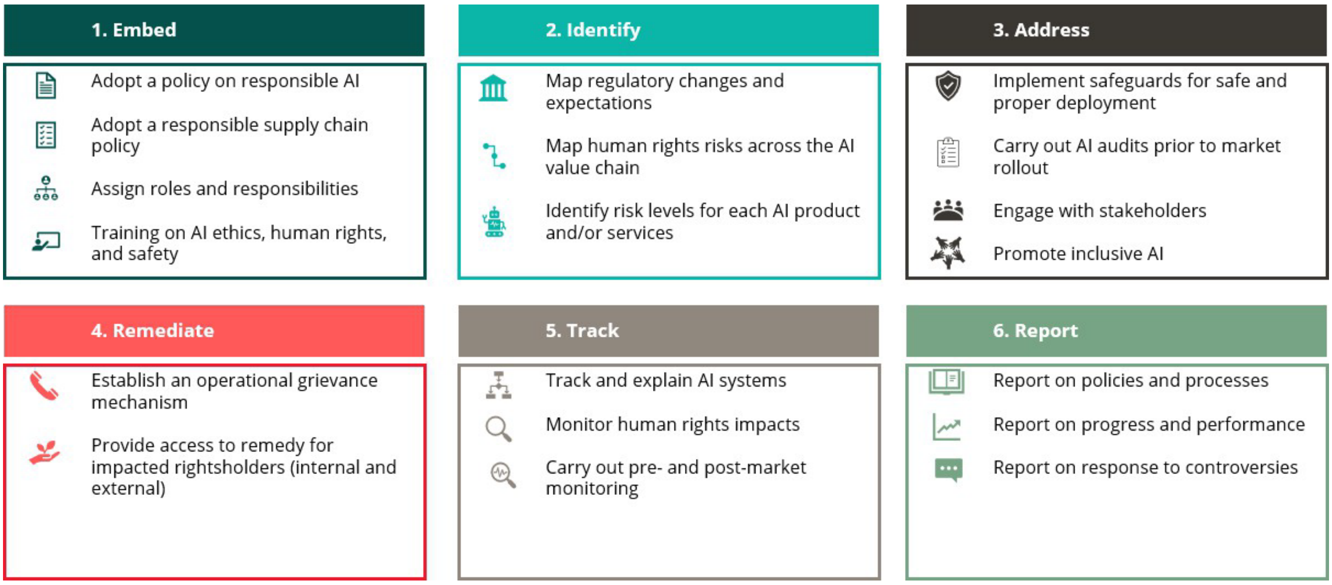
Source: EU AI Act, 2024, Naaia, 2024, RIAA, 2024, WEF, 2024, and Yap, 2025; created by Greenwheel.

The Greenwheel Responsible AI Framework for investors

Greenwheel developed a **Responsible AI Framework** to support investors in assessing the policies and processes of holding companies against norms and best practices on responsible AI (Figure 9). This Framework is grounded in the six steps of due diligence as part of responsible business conduct whilst drawing from a plethora of international norms, government regulations, and expert organisations. So far, the Framework targets two key AI actors: developers and deployers. Over time, Greenwheel seeks to add additional actors (e.g., data centres, data enrichment providers).

Although there are many emerging guidances and tools around responsible AI, many existing frameworks are not tailored to an investor audience; oftentimes, recommendations are provided to multiple stakeholder groups (e.g., government, businesses, and investors). Additionally, in our experience, investors find self-assessment tools useful if they can help identify red flags and gaps, practical company actions or key performance indicators, and engagement questions to companies. Finally, framing the tool against the six steps of standard human rights due diligence provides a consistent methodology for investors to apply international norms across an array of human rights issues.

Figure 9: Greenwheel's Responsible AI Framework



1. Embed
- Adopt a policy on responsible AI
- Adopt a responsible supply chain policy
- Assign roles and responsibilities
- Training on AI ethics, human rights, and safety

2. Identify
- Map regulatory changes and expectations
- Map human rights risks across the AI value chain
- Identify risk levels for each AI product and/or services

3. Address
- Implement safeguards for safe and proper deployment
- Carry out AI audits prior to market rollout
- Engage with stakeholders
- Promote inclusive AI

4. Remediate
- Establish an operational grievance mechanism
- Provide access to remedy for impacted rightsholders (internal and external)

5. Track
- Track and explain AI systems
- Monitor human rights impacts
- Carry out pre- and post-market monitoring

6. Report
- Report on policies and processes
- Report on progress and performance
- Report on response to controversies

Source: Greenwheel, 2025; created by Greenwheel.

1. Embed

Responsible AI Policy

A company can demonstrate alignment with internationally recognised human rights norms by **setting up a comprehensive responsible AI policy** (or equivalent, for instance, embedded into an existing human rights policy) that outlines a company's commitment to embed responsible AI practices.

- The policy should commit to upholding the defined responsible AI principles: accountability; autonomy; non-discrimination; inclusion; transparency; safety and security; and, privacy.

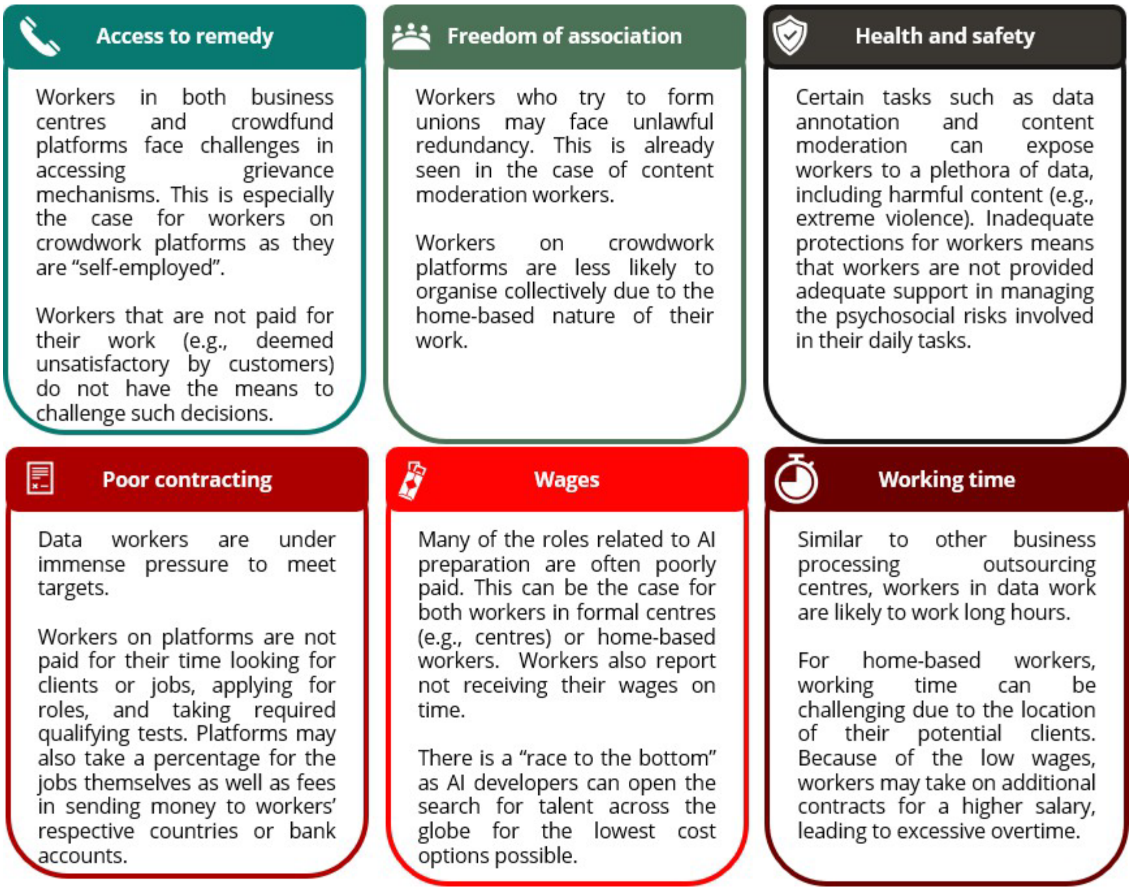
- The responsible AI policy should define clear roles and responsibilities as well as key actions taken to identify, mitigate, and remediate AI-related impacts.[22]

- For deployers, it is particularly important to consider how the deployment of AI may impact its employees in its direct operations.

Responsible Supply Chain Policy

As AI actors have a responsibility to ensure that internationally recognised human rights norms are upheld throughout the entire AI value chain, it is **crucial for both developers and deployers to have a responsible AI supply chain policy** (or equivalent embedded within an existing supply chain policy) that considers any human rights risks their suppliers may face. For instance, developers may consider risks related to data generation, data labelling and annotation, content moderation, and data verification (Figure 10). Additional measures may be necessary to address the risks facing workers on crowdwork platforms, particularly around wages (Box 1).

Figure 10: A summary of commonly found human rights risks in the AI supply chain



Access to remedy

Workers in both business centres and crowdfund platforms face challenges in accessing grievance mechanisms. This is especially the case for workers on crowdwork platforms as they are "self-employed".

Workers that are not paid for their work (e.g., deemed unsatisfactory by customers) do not have the means to challenge such decisions.

Freedom of association

Workers who try to form unions may face unlawful redundancy. This is already seen in the case of content moderation workers.

Workers on crowdwork platforms are less likely to organise collectively due to the home-based nature of their work.

Health and safety

Certain tasks such as data annotation and content moderation can expose workers to a plethora of data, including harmful content (e.g., extreme violence). Inadequate protections for workers means that workers are not provided adequate support in managing the psychosocial risks involved in their daily tasks.

Poor contracting

Data workers are under immense pressure to meet targets.

Workers on platforms are not paid for their time looking for clients or jobs, applying for roles, and taking required qualifying tests. Platforms may also take a percentage for the jobs themselves as well as fees in sending money to workers' respective countries or bank accounts.

Wages

Many of the roles related to AI preparation are often poorly paid. This can be the case for both workers in formal centres (e.g., centres) or home-based workers. Workers also report not receiving their wages on time.

There is a "race to the bottom" as AI developers can open the search for talent across the globe for the lowest cost options possible.

Working time

Similar to other business processing outsourcing centres, workers in data work are likely to work long hours.

For home-based workers, working time can be challenging due to the location of their potential clients. Because of the low wages, workers may take on additional contracts for a higher salary, leading to excessive overtime.

Source: Tubaro et al., 2020, Fairwork, 2023, and ILO, 2024; created by Greenwheel.

Box 1: Living tariffs and the question of fair wages for gig workers in the AI value chain

Living tariff is a concept that refers to a more accurate and fairer approach to wages for gig workers. This approach accounts for the out-of-pocket costs of associated with gig work. For example, in the context of delivery drivers, their wages would factor costs including equipment, fuel, insurance, and unbilled hours waiting for tasks. Currently piloted in India, Indonesia, Kenya, and Pakistan, the Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH (GIZ) and Wage Indicator Foundation have developed a Living Tariff Calculator to help gig workers calculate the true cost of a gig.

Translating this to the context of AI value chain workers, calculating the living tariff for a worker would include costs such as accessing the internet, electronics (to access the platform), social security, time spent bidding for tasks on platforms, and trainings for upskilling (especially important in responding to the rapidly evolving demands in the AI landscape). Altogether, the living wage should be able to afford essentials such as education, food, healthcare, housing, and upskilling. This tool can the potential in helping businesses calculate fairer wages to promote decent work for gig workers.

Source: BMZ, 2024 and Gigpedia, 2025.

We believe that the responsible AI policy, responsible AI supply chain policies, or their equivalent should be **developed with inputs from internal and external stakeholders**. Where possible, actual or potentially affected rightsholders and/or their representatives should provide feedback and insights on the appropriate scoping of these policies. Finally, the **policies should be signed off by senior leadership and reviewed on a regular basis**.[23]

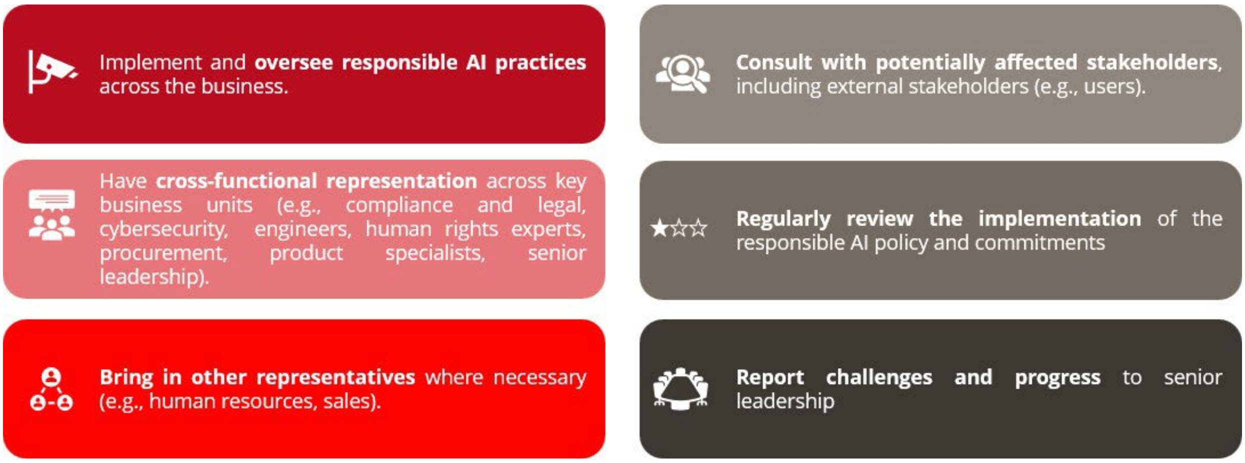
Assigning roles and establishing incentives

To oversee the implementation of the responsible AI policies and commitments, businesses may seek to **establish a responsible AI (or ethics) committee** (Figure 11). This committee should report challenges and progress to senior leadership on a regular basis.

For developers, this committee can bring together expertise from across the business to inform the responsible development of AI products and services. This allows engineering and product specialist teams to engage in dialogue with cybersecurity and human rights experts to understand potential and adverse impacts of AI systems; similarly, human rights experts can benefit from understanding the basics regarding any new or changes to products and services to better anticipate potential risks.

For deployers, the committee may draw in additional business units where relevant. For instance, the committee may bring in human resources if they are planning on deploying a new AI recruitment or performance review product.

Figure 11: Responsibilities of a responsible AI committee



Implement and **oversee responsible AI practices** across the business.

Consult with potentially affected stakeholders, including external stakeholders (e.g., users).

Have **cross-functional representation** across key business units (e.g., compliance and legal, cybersecurity, engineers, human rights experts, procurement, product specialists, senior leadership).

Regularly review the implementation of the responsible AI policy and commitments

Bring in other representatives where necessary (e.g., human resources, sales).

Report challenges and progress to senior leadership

Source: BSR, 2025b and ISO, 2025; created by Greenwheel.

Ultimately, **incentives play an important role in ensuring that responsible, safety, and security standards are implemented**.[24] Perverse incentives can encourage key actors within an organisation to forgo responsible AI practices. For example, tight deadlines and intense pressure on engineering and product teams would discourage the adoption of responsible AI practices, as this may cause delays in meeting their deliverables. Similarly, where the consideration of cybersecurity or human rights risks are not formalised in roles and responsibilities with adequate reward for compliance, employees may see responsible AI as largely voluntary.

Trainings

To support implementation of responsible AI policies and strong risk management systems, we believe, internal trainings should be provided to employees.[25] While it is important that all employees are adequately trained, **engineers and product specialists of developers, and procurement teams and affected business units of deployers in particular should understand the potential and adverse impacts of AI** in order to embed responsible practices into their day-to-day work.[26] Similarly, human rights experts and senior leadership are encouraged to have minimally a basic understanding of the products and services the company is developing or deploying to help them better assess and address potential risks and impacts.

2. Identify

Map regulatory changes

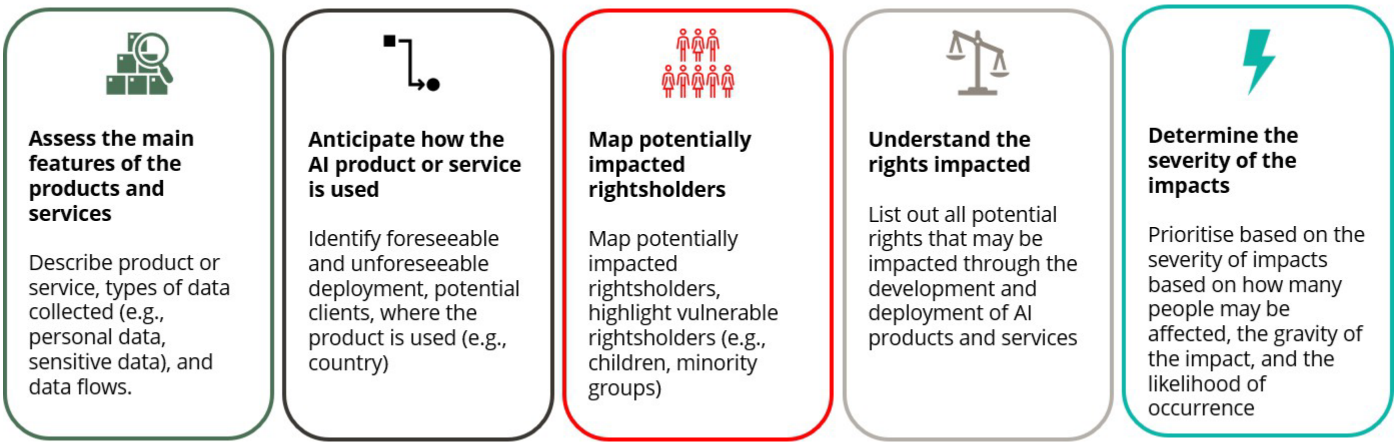
Emerging regulations on AI draw from international norms such as the UN Guiding Principles (UNGPs).[27] However, each **regulatory framework may have additional requirements for businesses**. The EU AI Act requires companies to define high-risk AI categories and carry out fundamental human rights impact assessments before deployment.[28] In China, there are requirements to clearly label content that is generated by AI; any sensitive research for AI development that involves human and animals or have significant societal impact require the establishment of an internal review committee and external experts to meet national ethical standards.[29]

Given the variances, following human rights due diligence may not be sufficient in meeting the diverse legal requirements. Businesses should conduct a mapping of the regulatory changes around the globe.

Map human rights risks across the AI value chain

Carrying out a human rights impact assessment or risk assessment prior to the market release of products and services can help **companies understand their most salient human rights risks**. This can help them prioritise key actions to take in response. This mapping should include risks in the company's direct operations, upstream activities (e.g., supply chain) as well as downstream risks (e.g., customers). As part of best practice, human rights impact assessment should involve the engagement of both internal and external stakeholders. Internal stakeholder engagement is recommended to get buy-in on the process and external engagement can gather insights from the perspective from potentially impacted rightsholders, their representatives, and experts.[30]

Figure 12: Assessing the human rights risks of AI products and services



Assess the main features of the products and services

Describe product or service, types of data collected (e.g., personal data, sensitive data), and data flows.

Anticipate how the AI product or service is used

Identify foreseeable and unforeseeable deployment, potential clients, where the product is used (e.g., country)

Map potentially impacted rightsholders

Map potentially impacted rightsholders, highlight vulnerable rightsholders (e.g., children, minority groups)

Understand the rights impacted

List out all potential rights that may be impacted through the development and deployment of AI products and services

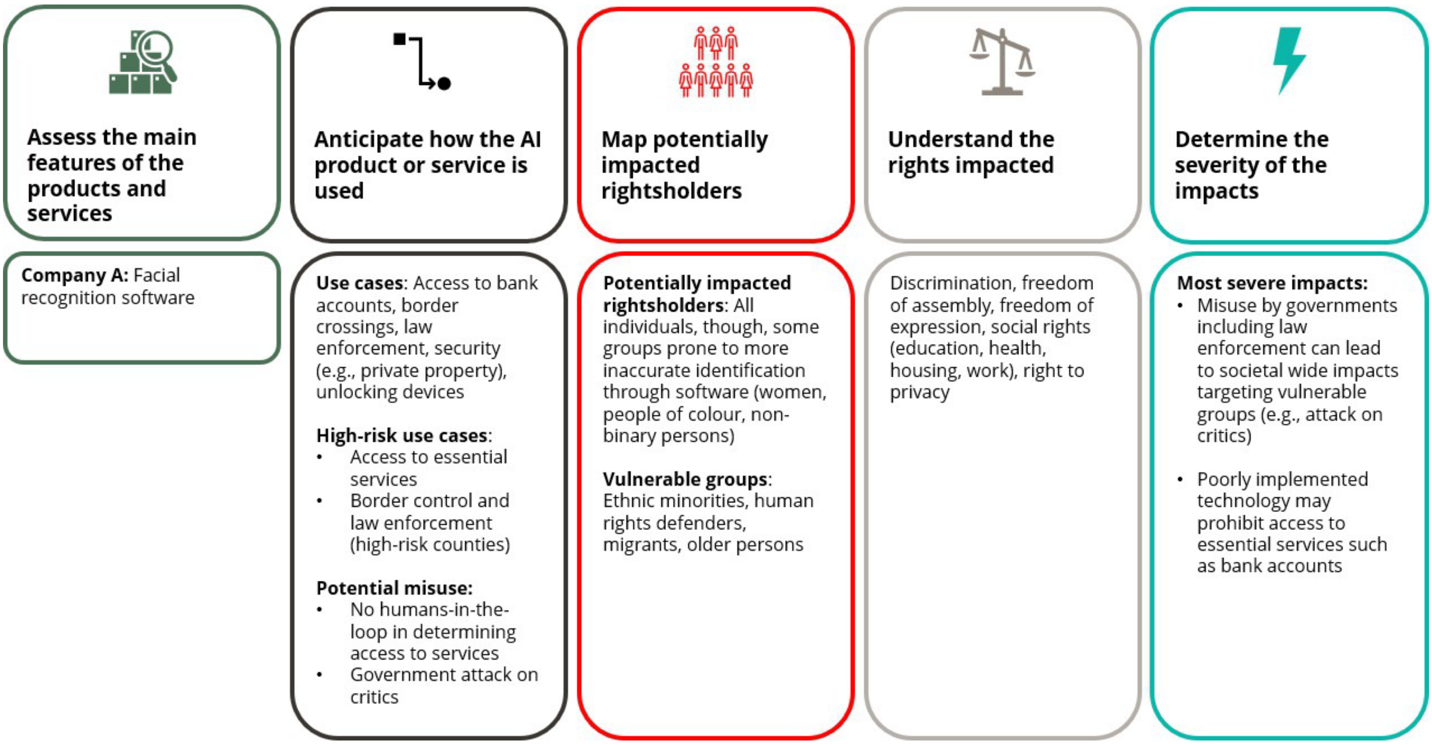
Determine the severity of the impacts

Prioritise based on the severity of impacts based on how many people may be affected, the gravity of the impact, and the likelihood of occurrence

Source: Mantelero and Esposito, 2021, United States Department of State, 2024, BSR, 2025d; created by Greenwheel.

Businesses should **identify the potential rights that may be impacted by AI products and services**, this includes civil and political rights and economic and social rights.[31] This assessment should consider the nature of the products and services, including the context for deployment (Figure 12). Businesses may consider assigning a risk level to their products and services based on this assessment to help prioritise actions and level of safeguarding required (Figure 13).

Figure 13: Assessing the risks of facial recognition software



Assess the main features of the products and services	Anticipate how the AI product or service is used	Map potentially impacted rightsholders	Understand the rights impacted	Determine the severity of the impacts
Company A: Facial recognition software	**Use cases**: Access to bank accounts, border crossings, law enforcement, security (e.g., private property), unlocking devices **High-risk use cases**: • Access to essential services • Border control and law enforcement (high-risk counties) **Potential misuse:** • No humans-in-the-loop in determining access to services • Government attack on critics	**Potentially impacted rightsholders**: All individuals, though, some groups prone to more inaccurate identification through software (women, people of colour, non-binary persons) **Vulnerable groups**: Ethnic minorities, human rights defenders, migrants, older persons	Discrimination, freedom of assembly, freedom of expression, social rights (education, health, housing, work), right to privacy	**Most severe impacts:** • Misuse by governments including law enforcement can lead to societal wide impacts targeting vulnerable groups (e.g., attack on critics) • Poorly implemented technology may prohibit access to essential services such as bank accounts

Source: Murray, 2024; created by Greenwheel.

3. Address

Implement safeguards

Safeguards should be in place across the AI lifecycle to address the plethora of adverse impacts including but not limited to biases, discriminatory results, privacy infringements, fraud, and the spread of disinformation and misinformation (Figure 14). Safeguards are especially important for generative AI due to the risks related to generating convincing yet false content (e.g., deepfakes) as well as adversely impacting an individual's intellectual property rights by copying existing work.

Conduct AI audits

AI audits enable businesses to assess their development and deployment of AI systems for legal compliance, technical safety, and ethics. An effective and trustworthy AI audit should provide comprehensive evidence about a system's responsible development and deployment.[32]

AI auditing can help businesses meet regulatory requirements.[33] For instance, under the EU AI Act, high-risk systems are required to undergo a conformity assessment prior to deployment carried out by an independent accredited third-party recognised by national authorities. Similarly, AI systems used to make employment-related decisions in New York City have to go through an independent audit.[34]

Figure 14: Safeguards for companies to adopt

	Developers	Deployers
Data sources	• **Carefully source data** and scrutinise data for quality (quality over quantity) • **Choose data based on representation and purpose** (e.g., language, representation, tasks, and topics) • **Make an informed choice** in selecting well-documented datasets • **Understand and document limitations**, especially when there are data gaps	• **Request developers for documentation** on datasets used to train models • **Understand the limitations** based on the data used to train the AI model, particularly gaps in representation that may lead to biases or copyright infringements
Data preparation	• **Analyse data for quality** (e.g., curated and edited data like Wikipedia and books versus social media) • **Analyse data for fairness** (e.g., biases, particularly against vulnerable groups) and assess potential human rights impacts • **Clean, filter, and review data** and check for accuracy and inaccurate labels • **Address concerns regarding copyright and licenses**	• **Assess the steps taken by developers** to address data quality and fairness
Model training	• **On-going assessment of downstream impacts** of training decisions • **Document training processes** for transparency • **Evaluate potential biases in model training** checking for algorithmic biases (e.g., optimisation techniques that favour majority group over minority group predictions)	• **On-going assessment of downstream impacts** of training decisions • **Document training processes**
Model evaluation	• **Evaluate model performance** against prescribed use cases • Anticipate and **expect unforeseen results and potential misuse** that can lead to adverse human rights impacts • **Red team to identify security risks and other vulnerabilities**, including potential risks to people	• **Request for data on model performance** from developers prior to purchase • **Evaluate the data and reporting** from developers through a human rights lens (e.g., potential impacts)
Model release and deployment	• Set up mechanisms to **address out-of-scope or misuse cases** • **Identify red lines and follow-up actions** in case of severe violations through misuse by deployers • **Carry out sales due diligence** in assessing the risk-levels of customers • **Provide guidance to deployers and users** on norms, limitations, error rates, and other relevant information to inform oversight • Establish an **adverse reporting mechanism** • **Create watermarking** where relevant (e.g., GenAI)	• **Adequately train deployers** in identifying harmful outputs and behaviours • **Provide human oversight in deployment** to intervene and address any biases, discriminatory outcomes, and errors • **Carry out on-going assessment** on adverse impacts as well as unintended consequences

Source: Longpre et al., 2025, BSR, 2025d, and BSR, 2025e, and Chapman University, 2025; created by Greenwheel.

Additionally, **AI audits can help assess the potential and actual environmental and social impact of AI systems as well as the mitigation measures adopted by a business**. On issues such as biases, AI audits can uncover flawed or non-representative datasets, test decisions made by AI processes, and engage with impacted rightsholders. Deployers seeking to adopt AI systems, from hiring processes to productivity tracking systems, can use AI audits to prevent unintended harm to workers. Particularly for high-risk AI systems used in sensitive personal data collection, facial recognition, and predictive policing, auditing can test systems for accuracy, fairness, and alignment with international norms.[35] AI audits can also evaluate a system's environmental impact on key metrics such as energy and water consumption, emissions, and e-waste.[36]

AI audits can be conducted internally or externally. **AI audits should be carried out before deployment and market rollout.** Auditing should occur iteratively and not as a one-time process.

Engage with stakeholders

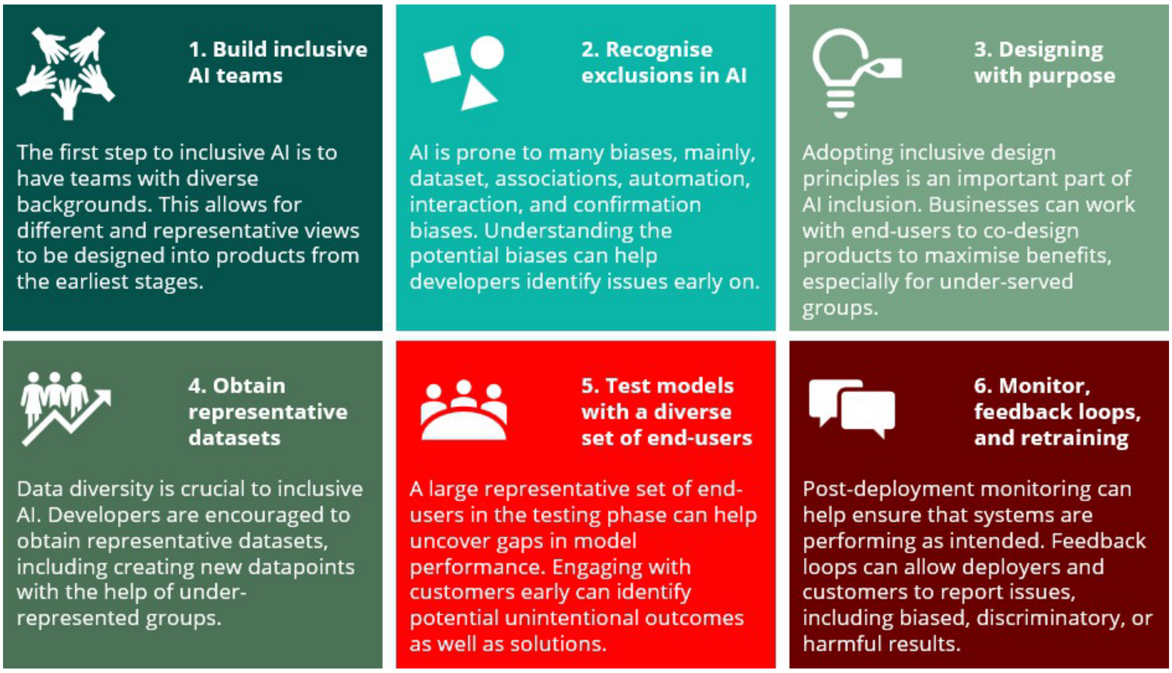
To inform the different stages of the development process from design to product rollout, external stakeholders such as academics, civil society organisations, human rights experts, governments should be consulted. Stakeholders can give inputs to companies on what works and potential challenges. Particularly where there are clear risks to specific groups of rightsholders (e.g., patients in a healthcare setting, own employees), engagement with the potentially impacted groups can help mitigate harm.[37]

To address any downstream adverse impacts related to deployment or use of the product or service, developers should communicate with deployers to understand how they plan to use the AI system. **Developers should inform deployers the potential risks and limitations to the given product or service**. Any unforeseen impacts after deployment should also be communicated to developers in an on-going basis to encourage continuous improvement.

Promote inclusive AI

To ensure that the benefits of AI can be reaped by everyone, **AI systems should be developed in a way that promotes equality, multilingualism, cultural diversity, and the inclusion of persons with disabilities**. To achieve this goal, businesses are recommended to think about inclusion as a lens across the whole AI lifecycle (Figure 15).

Figure 15: Promoting inclusive AI



1. Build inclusive AI teams

The first step to inclusive AI is to have teams with diverse backgrounds. This allows for different and representative views to be designed into products from the earliest stages.

2. Recognise exclusions in AI

AI is prone to many biases, mainly, dataset, associations, automation, interaction, and confirmation biases. Understanding the potential biases can help developers identify issues early on.

3. Designing with purpose

Adopting inclusive design principles is an important part of AI inclusion. Businesses can work with end-users to co-design products to maximise benefits, especially for under-served groups.

4. Obtain representative datasets

Data diversity is crucial to inclusive AI. Developers are encouraged to obtain representative datasets, including creating new datapoints with the help of under-represented groups.

5. Test models with a diverse set of end-users

A large representative set of end-users in the testing phase can help uncover gaps in model performance. Engaging with customers early can identify potential unintentional outcomes as well as solutions.

6. Monitor, feedback loops, and retraining

Post-deployment monitoring can help ensure that systems are performing as intended. Feedback loops can allow deployers and customers to report issues, including biased, discriminatory, or harmful results.

Source: Chou et al., 2018, Brayan, 2021, WEF, 2021, Microsoft, 2025, and WEF, 2025; created by Greenwheel.

To respect this inclusiveness throughout AI development, the active participation of individuals and groups regardless of gender, disability, ethnic origin, language, or age should be encouraged. For instance, one of the biggest challenges to inclusive AI is data inequity particularly from under-represented groups. **Companies can address this issue by working with governments, data suppliers, non-profit organisations, or academic institutions to build diverse training datasets**. An example emerging best practice is found in Japan, where a public-private partnership has been established to build training data originating in Japan to develop a LLM serving a Japanese audience.[38] These practices can help businesses tackle data gaps and overcome the challenges of developing AI systems that can capture cultural, linguistic, and local nuances.

4. Remediate

Operational Grievance Mechanisms

Operational grievance mechanisms should be in place so that any potentially or affected stakeholders (e.g., clients, suppliers, users) are able to report concerns and address complaints they have in the context of the development or deployment of an AI system. Grievance mechanisms should handle complaints from workers that are made redundant or adversely impacted by the adoption of AI systems internally. Grievance mechanisms (e.g., hotlines, online platforms) are the most efficient way to enable companies to directly respond to human rights-related concerns. They should therefore be effectively designed to encourage use by those stakeholder groups they were intended to be used by and guarantee non-retaliation.[39] Outcomes of grievance mechanisms should feedback into the development stages of AI.

As AI systems operate at a large scale and impact many users, it should be possible to raise grievances at an individual level (e.g., complaints regarding a specific AI-made decision), but also at a group-level (e.g., complaints regarding personal data collection or discriminatory outcomes).[40] To ensure that such grievance mechanisms are effective and remediation processes can be continuously improved, potentially and actually affected stakeholders should be consulted in the design process.

Access to remedy

Even with the implementation of safeguards, businesses may be unable to avoid all adverse impacts. As such, in line with international human rights norms and regulations, **AI developers and deployers should have appropriate remedy processes in place** (Figure 16).

A single point of contact should be responsible for the coordination of remedy processes across the entire value chain.[41] While this is usually the actor in the value chain with a direct connection to the impacted rightsholder, other actors in the value chain should still use their leverage to drive remediation of adverse human rights impacts.[42]

Figure 16: Steps for effective remedy

Remediation step	Example
After **identifying the source of the harm**, the associated system or specific functions may have to be **temporarily halted** to avoid additional adverse human rights impacts. A **commitment and plan to fix the mistake** should be developed and executed prior to making the system or functions available again.	Automated decision-making by an AI system in hiring processes found to exhibit biased results should be temporarily restricted. The company should document the issue identified and fix the issue prior to resuming use.
Apologising for and **acknowledging the harm** an AI system caused can provide victims with a form validation and emotional relief. It can also help rebuild the trust and reputation of a company.	AI deployers may send an apology in the form of an email to the individual or group of customers affected by a harmful deployment of a given AI service due to unforeseen issues.
Sanctions (contractual sanctions or penalties) may **be imposed on the AI actors responsible** for the human rights harm in line with laws and regulations.	Under the EU AI Act, non-compliance with various provisions relating to responsible AI could subject AI actors to fines up to 15 million EUR or 3% of a company's turnover.
To the furthest extent possible, **what has been lost** to a rightsholder due to a harm related to an AI product or service **should be restored** and the **state of the rightsholder** should be **returned to its condition prior to the occurred harm**.	An AI system misdiagnoses medical conditions in certain patient groups, so corrective services should be offered a reevaluation of affected patients and follow-up appointments with a medical professional.
To **support an affected rightsholder's recovery** from having incurred an AI-related harm, **rehabilitation** may include providing access to needed medical, psychological, legal and/or social services. If the AI actor is not able to provide this, affected rightsholders should minimally be **redirected to expert organisations**.	A social media company using a recommendation system that promotes extremely harmful content can provide access to counselling services or other psychological support to affected users.
In addition to restitution and rehabilitation, **adequate compensation to victims** of AI-related harms may be necessary. Compensation may be **financial or non-financial**, although it may sometimes be challenging to quantify damages.	When an AI system used in a multiplayer game falsely penalises and bans players based on flaws in the algorithm, players could be compensated with in-game currency.
Taking necessary steps, implementing appropriate measures and making policy changes **to ensure the same harms do not occur again** is crucial. As AI systems such as GenAI, are built on probabilistic algorithms, harmful outcomes may still occur despite safety controls and risk mitigants.	A GenAI developer may incorporate new safety fine-tuning methods to reduce hallucinations in future outputs and document changes made to explain to relevant stakeholders.

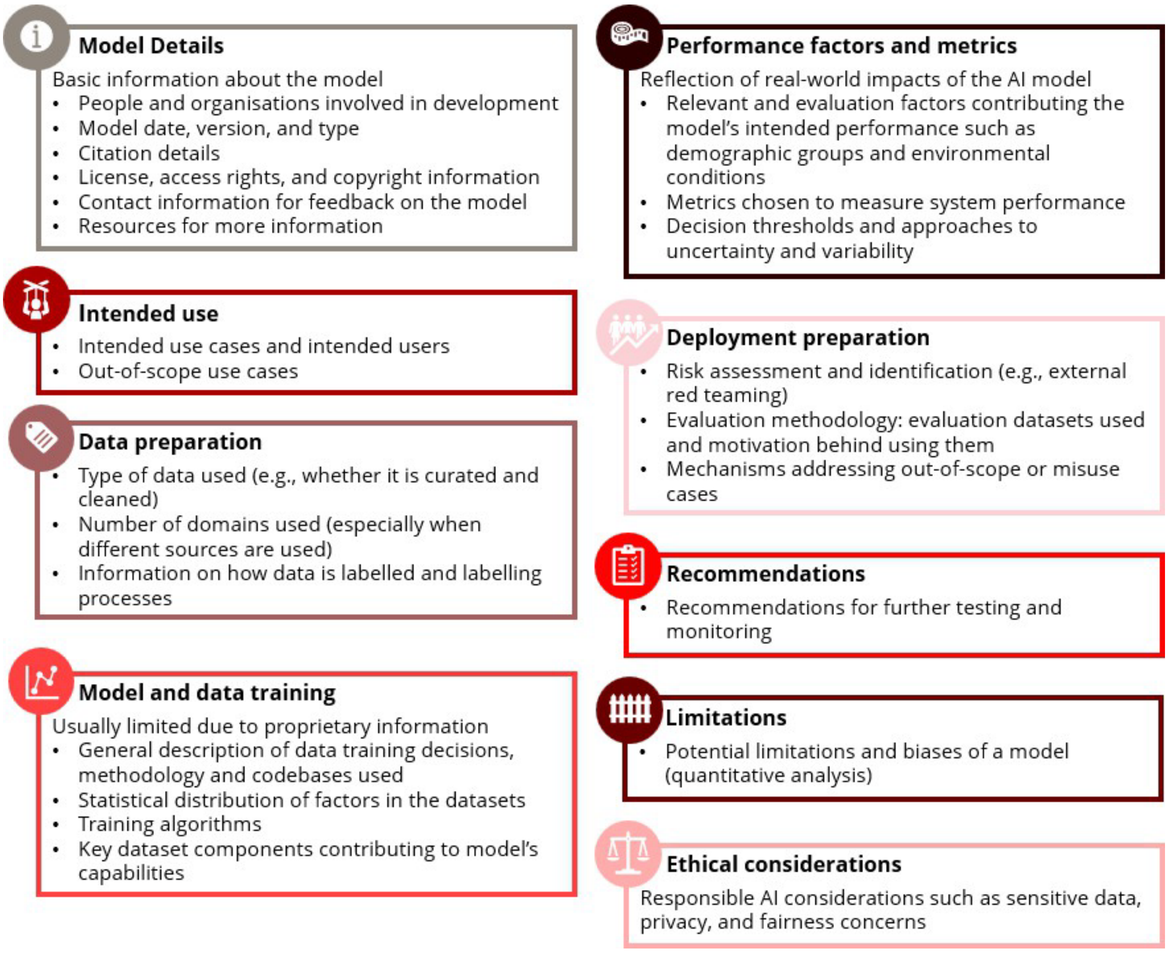
Source: B-Tech, 2021, EU AI Act, 2024, Hudson, 2024, BSR, 2025f, and UNEP, 2025; created by Greenwheel.

5. Track

Tracking from data to inference

AI developers should track and document how, why, and with what inputs an AI system is developed. **All AI actors in the value chain should be able to explain the various relevant elements and implications of AI systems** (Figure 17). To the furthest extent possible, easily understandable information about these AI system processes should be available to stakeholders affected by the outputs of the system so that they can understand and challenge outputs if necessary.[43] Documentation tools such as model, system or service cards are used in practice to help in the "data-to-inference" tracking process and explainability of an AI model.[44]

Figure 17: Recommended metrics to capture in a model card[45] based on existing best practice



Model Details

Basic information about the model

- People and organisations involved in development
- Model date, version, and type
- Citation details
- License, access rights, and copyright information
- Contact information for feedback on the model
- Resources for more information

Intended use

- Intended use cases and intended users
- Out-of-scope use cases

Data preparation

- Type of data used (e.g., whether it is curated and cleaned)
- Number of domains used (especially when different sources are used)
- Information on how data is labelled and labelling processes

Model and data training

Usually limited due to proprietary information

- General description of data training decisions, methodology and codebases used
- Statistical distribution of factors in the datasets
- Training algorithms
- Key dataset components contributing to model's capabilities

Performance factors and metrics

Reflection of real-world impacts of the AI model

- Relevant and evaluation factors contributing the model's intended performance such as demographic groups and environmental conditions
- Metrics chosen to measure system performance
- Decision thresholds and approaches to uncertainty and variability

Deployment preparation

- Risk assessment and identification (e.g., external red teaming)
- Evaluation methodology: evaluation datasets used and motivation behind using them
- Mechanisms addressing out-of-scope or misuse cases

Recommendations

- Recommendations for further testing and monitoring

Limitations

- Potential limitations and biases of a model (quantitative analysis)

Ethical considerations

Responsible AI considerations such as sensitive data, privacy, and fairness concerns

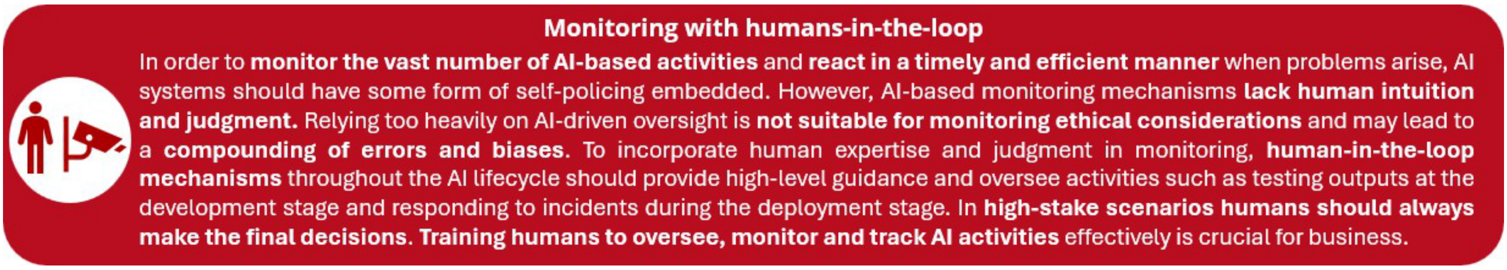
Source: IAPP, 2023, Hewson, 2024, OpenAI, 2024, Longpre et al., 2025, BSR, 2025d, BSR, 2025e, Chapman University, 2025, and Google, 2025; created by Greenwheel.

Monitoring human rights impacts

Human rights violations occurring throughout the AI value chain should be monitored and documented. This includes tracking any complaints raised, including the number of complaints and type of complaint (e.g., right violated or harm perpetrated), and closely following remediation processes, including the time taken to investigate and remedies offered. Tracking this information closely can be advantageous for companies to prepare for external evaluations and regulatory mandates.[46]

Companies should also continuously monitor the effectiveness of their safeguards implemented and mitigation measures taken to ensure that AI systems are not having unintended negative impacts. Lastly, as AI systems also have many positive impacts on individuals and society, AI actors are encouraged to track the beneficial outcomes of the developed and deployed AI systems.

Figure 18: Keeping humans-in-the-loop in monitoring AI systems



Monitoring with humans-in-the-loop

In order to **monitor the vast number of AI-based activities** and **react in a timely and efficient manner** when problems arise, AI systems should have some form of self-policing embedded. However, AI-based monitoring mechanisms **lack human intuition and judgment.** Relying too heavily on AI-driven oversight is **not suitable for monitoring ethical considerations** and may lead to a **compounding of errors and biases.** To incorporate human expertise and judgment in monitoring, **human-in-the-loop mechanisms** throughout the AI lifecycle should provide high-level guidance and oversee activities such as testing outputs at the development stage and responding to incidents during the deployment stage. In **high-stake scenarios humans should always make the final decisions. Training humans to oversee, monitor and track AI activities** effectively is crucial for business.

Source: OECD, 2023 and Security, 2024; created by Greenwheel.

Post-market monitoring

Once the AI product or service reaches the market, AI actors should keep track of the performance of the AI system through post-market monitoring.[47]

For developers this includes tracking what and how many use cases there were for the system and the percentage of incorrect or misleading results the system produces. There should be continuous engagement with the deploying entity and direct feedback should be acquired to include into further system development and improvement.

For deployers post-market monitoring includes tracking what use cases there were for the system; the number of users of the system; and the number of queries that were resolved. To understand the impact AI deployment has on employees, customers, society and the business as a whole, companies should compare the performance to the situation prior to the introduction of AI.

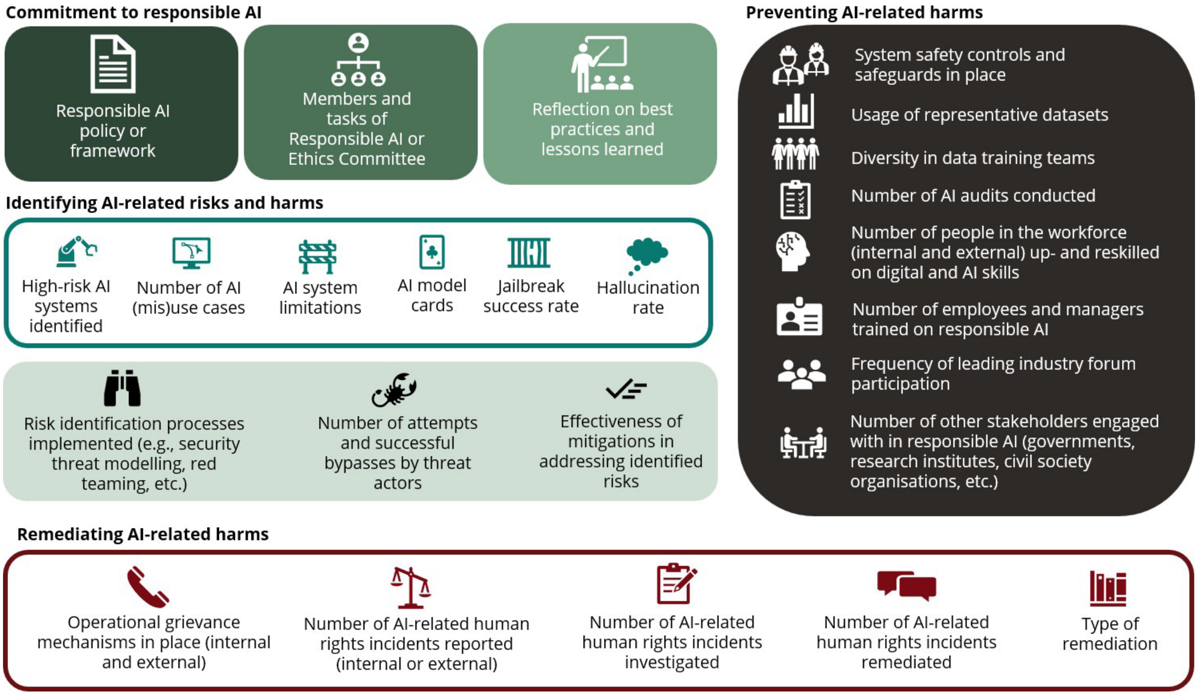
6. Report

Public reporting

To openly demonstrate their commitment to responsible AI practices, minimally, **companies should make their responsible AI policies publicly available**. In line with the transparency principle and the UN Guiding Principles on Business and Human Rights, companies should disclose the findings from their impacts or risks assessments, mitigation efforts in addressing salient human rights issues, and an evaluation of their effectiveness (Figure 19).[48]

Findings from AI audits and post-market monitoring should be reported. For developers, these may include the number of users and number of queries resolved, whereas developers may report against indicators such as number of intended use cases, number of misuses, and rates of hallucination.[49] As a best practice, companies should explain how their performance stands against the targets set and document successes, challenges, and lessons encountered when implementing a responsible AI strategy. This will showcase that implementing such a strategy is an iterative process that can continuously be improved.

Figure 19: Key Performance Indicators to capture in public reports



Commitment to responsible AI

- Responsible AI policy or framework
- Members and tasks of Responsible AI or Ethics Committee
- Reflection on best practices and lessons learned

Identifying AI-related risks and harms

- High-risk AI systems identified
- Number of AI (mis)use cases
- AI system limitations
- AI model cards
- Jailbreak success rate
- Hallucination rate
- Risk identification processes implemented (e.g., security threat modelling, red teaming, etc.)
- Number of attempts and successful bypasses by threat actors
- Effectiveness of mitigations in addressing identified risks

Preventing AI-related harms

- System safety controls and safeguards in place
- Usage of representative datasets
- Diversity in data training teams
- Number of AI audits conducted
- Number of people in the workforce (internal and external) up- and reskilled on digital and AI skills
- Number of employees and managers trained on responsible AI
- Frequency of leading industry forum participation
- Number of other stakeholders engaged with in responsible AI (governments, research institutes, civil society organisations, etc.)

Remediating AI-related harms

- Operational grievance mechanisms in place (internal and external)
- Number of AI-related human rights incidents reported (internal or external)
- Number of AI-related human rights incidents investigated
- Number of AI-related human rights incidents remediated
- Type of remediation

Source: Cisco, 2024, IBM, 2024, Microsoft, 2024, BSR, 2025f, Cisco, 2025, Google, 2025a, Google 2025b, IBM, 2025, Intel, 2025, Longpre et al., 2025, and Microsoft, 2025; created by Greenwheel.

Responding to allegations

Even with the appropriate measures adopted in line with international norms and best practices to address the human rights risks posed by AI, **businesses may still have adverse impacts on rightsholders**.

Businesses may face allegations of human rights violations by civil society, journalists, and/or trade unions. In response, **businesses should carry out an investigation to substantiate the claims**. Depending on the severity of the claims and the internal expertise available, companies may reach out to third-party experts for support. Investigation findings should be publicly available; sensitive data should not be presented. Where allegations are substantiated, businesses should provide remedy to affected rightsholders and the steps taken to improve on policies and processes to guarantee non-repetition.

References

[1] See Greenwheel research on AI opportunities in Emerging Markets.
[2] See Greenwheel research on AI's risks to workers and AI's risks to communities.
[3] UNESCO, 2025.
[4] UN B-Tech and UNOHCHR, 2023.
[5] OECD, 2025 and UNESCO, 2025.
[6] UN, 2024.
[7] Mäntymäki et al, 2022.
[8] Calvo, et al., 2020.
[9] Prunkl, 2024 and Dunning et al., 2024.
[10] Placani, 2024.
[11] Yang et al., 2025.
[12] White and Case, 2024.
[13] AI, Artificial Intelligence, Personal Data, and Privacy, 2024.
[14] Pouget and Zuhdi, 2024.
[15] Pinsent Masons, 2024.
[16] European Commission, 2024.
[17] Zanatta and Rielli, 2024 and Atanasovska and Robeli, 2025.
[18] Commentary provided by civil society representatives at the 13th United Nations Forum on Business and Human Rights in Geneva.
[19] OECD, 2025.
[20] Council of Europe, 2024.
[21] WEF, 2024.
[22] BSR, 2025a.
[23] BSR, 2025a.
[24] BSR, 2025b.
[25] BSR, 2025b.
[26] Best practices shared by businesses and civils society at RightsCon, 2025.
[27] BSR, 2025c.
[28] BSR, 2025d.
[29] Atanasovska and Trombevski, 2024.
[30] BSR, 2025d.
[31] See Greenwheel research on AI's risks to workers and AI's risks to communities.
[32] IPIE, 2024.
[33] Mökander, 2023.
[34] Mökander et al., 2022.
[35] Claruna, 2025.
[36] IPIE, 2024.
[37] Khan and Park, 2024.
[38] WEF, 2025.
[39] B-Tech, 2021.
[40] Chatham House, 2023.
[41] BSR, 2025.
[42] B-Tech, 2023.
[43] OECD, 2024.
[44] OECD, 2023.
[45] A model card is a short document providing relevant information about a machine learning or AI model. It can help promote the explainability of AI systems. IAPP, 2023.
[46] BSR, 2025g.
[47] Stein and Dunlop, 2024 and EU AI Act, 2025.
[48] OHCHR, 2011.
[49] BSR, 2025g.

Key Information

No investment strategy or risk management technique can guarantee returns or eliminate risks in any market environment. Past performance is not a guide to future results. The prices of investments and income from them may fall as well as rise and an investor’s investment is subject to potential loss, in whole or in part. Forecasts and estimates are based upon subjective assumptions about circumstances and events that may not yet have taken place and may never do so. The statements and opinions expressed in this article are those of the author as of the date of publication, and do not necessarily represent the view of Redwheel. This article does not constitute investment advice and the information shown is for illustrative purposes only. Whilst updated figures are not available for all sources, we have performed further analysis and believe that this data has not significantly changed and is reflective for 2025.

Disclaimer

Redwheel ® and Ecofin ® are registered trademarks of RWC Partners Limited (“RWC“). The term “Redwheel“ may include any one or more Redwheel branded regulated entities including, RWC Asset Management LLP which is authorised and regulated by the UK Financial Conduct Authority and the US Securities and Exchange Commission (“SEC“); RWC Asset Advisors (US) LLC, which is registered with the SEC; RWC Singapore (Pte) Limited, which is licensed as a Licensed Fund Management Company by the Monetary Authority of Singapore; Redwheel Australia Pty Ltd is an Australian Financial Services Licensee with the Australian Securities and Investment Commission; and Redwheel Europe Fondsmæglerselskab A/S (“Redwheel Europe“) which is regulated by the Danish Financial Supervisory Authority. Redwheel may act as investment manager or adviser, or otherwise provide services, to more than one product pursuing a similar investment strategy or focus to the product detailed in this document. Redwheel and RWC (together “Redwheel Group“) seeks to minimise any conflicts of interest, and endeavours to act at all times in accordance with its legal and regulatory obligations as well as its own policies and codes of conduct. This document is directed only at professional, institutional, wholesale or qualified investors. The services provided by Redwheel are available only to such persons. It is not intended for distribution to and should not be relied on by any person who would qualify as a retail or individual investor in any jurisdiction or for distribution to, or use by, any person or entity in any jurisdiction where such distribution or use would be contrary to local law or regulation. This document has been prepared for general information purposes only and has not been delivered for registration in any jurisdiction nor has its content been reviewed or approved by any regulatory authority in any jurisdiction. The information contained herein does not constitute: (i) a binding legal agreement; (ii) legal, regulatory, tax, accounting or other advice; (iii) an offer, recommendation or solicitation to buy or sell shares in any fund, security, commodity, financial instrument or derivative linked to, or otherwise included in a portfolio managed or advised by Redwheel; or (iv) an offer to enter into any other transaction whatsoever (each a “Transaction“). Redwheel Group bears no responsibility for your investment research and/or investment decisions and you should consult your own lawyer, accountant, tax adviser or other professional adviser before entering into any Transaction. No representations and/or warranties are made that the information contained herein is either up to date and/or accurate and is not intended to be used or relied upon by any counterparty, investor or any other third party. Redwheel Group uses information from third party vendors, such as statistical and other data, that it believes to be reliable. However, the accuracy of this data, which may be used to calculate results or otherwise compile data that finds its way over time into Redwheel Group research data stored on its systems, is not guaranteed. If such information is not accurate, some of the conclusions reached or statements made may be adversely affected. Any opinion expressed herein, which may be subjective in nature, may not be shared by all directors, officers, employees, or representatives of Redwheel Group and may be subject to change without notice. Redwheel Group is not liable for any decisions made or actions or inactions taken by you or others based on the contents of this document and neither Redwheel Group nor any of its directors, officers, employees, or representatives (including affiliates) accepts any liability whatsoever for any errors and/or omissions or for any direct, indirect, special, incidental, or consequential loss, damages, or expenses of any kind howsoever arising from the use of, or reliance on, any information contained herein. Information contained in this document should not be viewed as indicative of future results. Past performance of any Transaction is not indicative of future results. The value of investments can go down as well as up. Certain assumptions and forward looking statements may have been made either for modelling purposes, to simplify the presentation and/or calculation of any projections or estimates contained herein and Redwheel Group does not represent that that any such assumptions or statements will reflect actual future events or that all assumptions have been considered or stated. There can be no assurance that estimated returns or projections will be realised or that actual returns or performance results will not materially differ from those estimated herein. Some of the information contained in this document may be aggregated data of Transactions executed by Redwheel that has been compiled so as not to identify the underlying Transactions of any particular customer. No representations or warranties of any kind are intended or should be inferred with respect to the economic return from, or the tax consequences of, an investment in a Redwheel-managed fund. 22 This document expresses no views as to the suitability or appropriateness of the fund or any other investments described herein to the individual circumstances of any recipient. The information transmitted is intended only for the person or entity to which it has been given and may contain confidential and/or privileged material. In accepting receipt of the information transmitted you agree that you and/or your affiliates, partners, directors, officers and employees, as applicable, will keep all information strictly confidential. Any review, retransmission, dissemination or other use of, or taking of any action in reliance upon, this information is prohibited. Any distribution or reproduction of this document is not authorised and is prohibited without the express written consent of Redwheel Group. The risks of investment are detailed in the Prospectus and should be considered in conjunction with your investment adviser. Please refer to the Prospectus, Key Investor Information Document (UCITS KIID), Key Information Document (PRIIPS KID), Summary of Investor Rights and other legal documents as well as annual and semi-annual reports before making investment decisions; these documents are available free of charge from RWC or on RWC's website: https://www.redwheel.com/ and available in local languages where required. RWC as the global distributor has the right to terminate the arrangements made for marketing Redwheel Funds in certain jurisdictions and to certain investors. Redwheel Europe is the sub-distributor of shares in Redwheel Funds in the European Economic Area ("EEA") and is regulated by the Danish Financial Supervisory Authority. This document is not a solicitation or an offer to buy or sell any fund or other investment and is issued in the UK by RWC and in the EEA by RW Europe. This document does not constitute investment, legal or tax advice and expresses no views as to the suitability or appropriateness of any investment and is provided for information purposes only. The views expressed in the commentary are those of the investment team. Funds managed by Redwheel are not, and will not be, registered under the Securities Act of 1933 (the "Securities Act") and are not available for purchase by US persons (as defined in Regulation S under the Securities Act) except to persons who are "qualified purchasers" (as defined in the Investment Company Act of 1940) and "accredited investors" (as defined in Rule 501(a) under the Securities Act). This document does not constitute an offer to sell, purchase, subscribe for or otherwise invest in units or shares of any fund managed by Redwheel. Any offering is made only pursuant to the relevant offering document and the relevant subscription application. Prospective investors should review the offering memorandum in its entirety, including the risk factors in the offering memorandum, before making a decision to invest.

This document does not constitute an offer to sell, purchase, subscribe for or otherwise invest in units or shares of any fund managed by Redwheel. Any offering is made only pursuant to the relevant offering document and the relevant subscription application. Prospective investors should review the offering memorandum in its entirety, including the risk factors in the offering memorandum, before making a decision to invest.

CONTACT US
Please contact us if you have any questions or
would like to discuss any of our strategies.
invest@redwheel.com | www.redwheel.com

Redwheel London
Verde
10 Bressenden Place
London SW1E 5DH
+4420 72276000

Redwheel Europe
Fondsmæglerselskab A/S,
Havnegade 39, 1058
København K, Denmark

Redwheel Miami
2640 South Bayshore Drive
Suite 201
Miami
Florida 33133
+1 305 6029501

Redwheel Singapore
80 Raffles Place
#22-23
UOB Plaza 2
Singapore 048624
+65 68129540