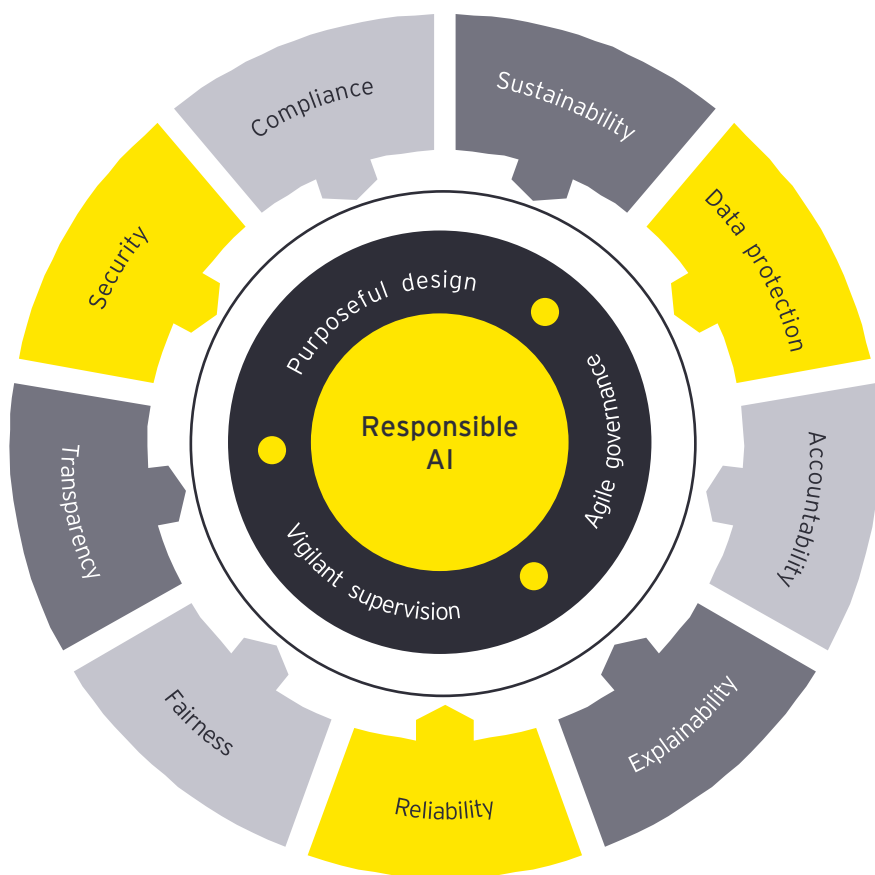


# Responsible AI principles



Shape the future  
with confidence



The detailed definitions for EY responsible AI principles have been crafted and cross-validated with leading ethical frameworks and standards on AI, including those outlined by National Institute of Standards and Technology (NIST), International Organization for Standardization (ISO), the Organization for Economic Cooperation and Development (OECD), the European Union's high-level expert group (HLEG), as well as key AI-related EY activities (e.g., EY.ai Confidence Index, EY Responsible AI Readiness Assessment).

## Definition of AI

Artificial intelligence (AI) is a machine-based system that infers from input and data it receives how to generate outputs, such as predictions, content, recommendations or decisions that can influence physical or virtual environments. While shaped by human objectives and values, AI systems vary in their autonomy and adaptability after deployment.



# Accountability

There is unambiguous ownership over AI systems, their impacts and resulting outputs across the AI lifecycle.

## Importance to EY

As the development, deployment and use of AI systems continues to play a larger role in our personal and working lives, applying clear lines of accountability is essential. It provides a “human face” to AI that provides clarity to end users on who to contact to learn more, challenge outcomes and seek redress if they are adversely impacted by AI systems and their outputs, thereby building trust, increasing engagement and helping optimize ultimate return on investment.

Accountability underpins all of EY responsible AI principles and aligns with our broader values. Moreover, accountability is vital for EY teams to comply with evolving AI laws and regulations, and effectively manage our reputational risk.

Accountability entails holding individuals, organizations and/or entities responsible for the development, deployment and outputs of AI systems and their proper functioning throughout their lifecycle and, as such, applies to a variety of stakeholders including designers, deployers and users themselves.

Accountability should be **unambiguous, transparent and documented** across the lifecycle of an AI system from development through to operation and retirement. For example, if a decision has been made on the trade-off between performance and explainability, the rationale for this design decision should be documented along with who has made the decision and when.

Accountability can be **reactive and proactive**.

- ▶ Reactive accountability signifies a duty to react proportionately and respond in a timely fashion to incidents involving AI systems and their outputs that an entity owns, builds or uses.
- ▶ Proactive accountability denotes an obligation to anticipate and prepare for possible future harms (within reason) related to an entity’s development or use of AI systems and their outputs.

**Related terms: responsibility, ownership**

## EY commitment

We design and build systems that purposefully integrate clear accountability and help in delivering business goals. We provide appropriate opportunities for feedback, relevant explanations, and clear channels for challenge and appeal. Our AI technologies are aligned with EY values as outlined in our Code of Conduct and subject to appropriate and proportionate levels of human oversight and control.



# Data protection

Use of data in AI systems is consistent with permitted rights, maintains confidentiality of business and personal information, and reflects ethical norms.

## Importance to EY

EY clients entrust the firm with highly sensitive business and personal information. Adherence to data protection practices throughout the AI lifecycle is essential to maintain client trust and safeguard EY brand value.

Proactive data protection is not only a business necessity for effectively mitigating legal and reputational risks and for avoiding costs associated with data breaches and regulatory fines, it also aligns with EY values and is integral to our corporate governance strategy for managing risks to global operations.

Robust data protection and confidentiality frameworks applied to the AI lifecycle help EY teams to innovate with confidence, knowing that our technological advancements are legally compliant and underpinned by a commitment to responsible and ethical norms.

Data protection requires the rigorous implementation of practices that respect and uphold the **legal rights of individuals and entities regarding their data**. This includes confirming that data is only collected and used in ways that are legally and ethically permissible. Securing personal and confidential business information throughout the AI lifecycle protects data from unauthorized access, disclosure or corruption and helps ensure **confidentiality and integrity**.

Data protection requires a **holistic technical and non-technical approach** to data governance, ensuring legal compliance; ethical data handling; data, AI model and system security measures; and organizational policies geared towards maintaining the integrity, confidentiality and lawful use of data within the realm of AI.

Adherence to data protection practices **applies to both data used and generated throughout the AI lifecycle**, including data used for training, testing, validation, monitoring, production input data, inferred data derived from the processing of input data and AI system output data.

**Related terms:** data security, data privacy, data governance

## EY commitment

Our AI systems are steadfast in protecting the data that fuels them and the content they generate. Data protection is an integral part of our responsible AI framework and extends beyond mere compliance to embedding data protection into the fabric of our AI systems design, development, deployment and use.



# Reliability

AI systems are aligned with stakeholder expectations and continually perform at a desired level of precision and consistency.

## Importance to EY

As AI plays a larger role in the professional services landscape, the accurate, safe and consistent performance of AI systems is essential for building confidence in AI and enabling leading-class employee and client experiences.

Reliability assists us in realizing our purpose of building a better working world by promoting the development of high-performing AI systems whose decisions, recommendations, actions and content meet the needs and expectations of the firm and clients we support.

Reliability is especially important when we are developing, procuring, or using high-risk AI systems that may adversely impact physical safety, mental well-being, social stability, or the environment, if found to be unreliable.

Reliability refers to the capability of AI systems to perform **designated functions consistently and without failure** under specific conditions for a specific period of time.<sup>1</sup> Reliability **supports AI systems' functional safety** and seeks to minimize potential harms when operated under unexpected settings.

Reliability entails safeguarding against uncontrolled changes in model outputs and enabling **early detection of malfunctioning or underperforming** AI through monitoring of key performance metrics, such as precision, accuracy and robustness.

**Precision** measures the quality of accurate predictions made by the model in relation to ground truth<sup>2</sup>, often measured by proportion of true positives over all positive predictions.<sup>3</sup>

**Accuracy** measures how well a model generates correct predictions across all categories of prediction, calculated by the number of true predictions over all predictions.

**Robustness** aims for appropriate system functionality in a set of conditions and circumstances. It measures model generalization and stability over time through data augmentation in development and monitoring model drift post-deployment.

**Related terms: consistency, resilience, robustness**

<sup>1</sup> Alignment with ISO and NIST: "for a given time interval" (NIST - ISO/IEC TS 5723:2022), "for a specific period of time" (ISO/IEC 27040).

<sup>2</sup> Ground truth refers to the target for training or validating machine learning models.

<sup>3</sup> This metric is only applicable for categorical predictions, not numerical ones, where it would make sense to include a numerical distance metric (e.g., in the case of LLMs).

## EY commitment

We are committed to helping enable reliable AI systems. We strive to build AI systems that are not only innovative but also dependable by integrating the right level of monitoring, reporting and intervention to help ensure that our AI systems perform consistently, safely, and reliably in supporting business objectives.



# Security

AI systems, their input and output data are secured from unauthorized access and resilient against corruption and adversarial attack.

## Importance to EY

Maintaining the security of AI systems is integral to EY purpose and values, as well as fulfilling our regulatory and professional obligations as a trusted advisor. Undetected and unresolved degradations in AI system accuracy and reliability as the result of a security breach can lead to significant reputational damage and undermine client trust through suboptimal service quality.

It is crucial to anticipate and devise strategies for potential security risks that AI systems might present or encounter before these systems are deployed, especially within the scope of safety-critical or high-risk applications. Secure AI mitigates reputational risk and helps enable the protection of firm and client assets and is mandated through existing firm wide policies such as EY global information security policies.

Measures to maintain the security of AI systems are designed to **protect the confidentiality, integrity and availability of AI systems** by preventing unauthorized access and use, exploitation, corruption and damage of/to AI systems and their input and output data. Robust AI security also encompasses **resilience** – the ability to return to normal function after unexpected adverse events – and **robustness** – the ability to withstand unexpected or adversarial data and model use. AI security measures should be implemented and maintained to prevent, identify, detect and recover from malicious attacks and/or accidental negligence across the lifecycle.

AI security threats include unauthorized access to data and models, data poisoning, model corruption and adversarial attack. For example, an attacker may repurpose an AI system's intended use to promote hate speech, discrimination or incite violence through direct or indirect prompt injection.

Without timely intervention, incidents can lead to deteriorating model performance and extraction of sensitive information. Security is thus a pre-requisite for data protection and impacts **AI safety** – the expectation that under defined conditions, AI systems do not endanger human life, health or the environment either deliberately or through accidental negligence

**Related terms: safety, resilience, robustness**

<sup>1</sup> Alignment with ISO 42001 and NIST 100-2.

## EY commitment

We prioritize the security and safety of AI systems, adhering to best practice guidance to protect against threats to and from AI systems across the development lifecycle. We continuously review and enhance our AI security risk management approaches to safeguard AI systems and resulting outputs to keep pace with technical advancements and the emerging threat landscape.



# Transparency

Appropriate levels of disclosure regarding the purpose, design and impact of AI systems is provided so that stakeholders, including end users, can understand, evaluate and correctly employ AI systems and their outputs.

## Importance to EY

Transparent AI aligns with EY public purpose statement, which expressly flags transparency as a catalyst for helping clients efficiently access the benefits of AI and contributing to a more inclusive, equitable future.

Transparency over AI methodologies and underlying data usage fosters a culture of accountability and continuous improvement, which is essential for innovation and delivering high-quality services in a competitive global market.

Transparency is vital for EY to remain in compliance with evolving AI laws and regulations, as well as effectively manage our broader reputational risk.

Transparency refers to a **reasonable disclosure** of the **functionalities, applications, limitations and implications** of AI systems and associated input data at various stages in the development lifecycle, thereby aiding relevant stakeholders, such as end users, products owners, or system designers, in better understanding and responsibly deploying AI systems and resulting outputs.

Transparency is often realized through appropriate documentation being created (e.g., design decisions, operating logs, model cards, data sheets, instruction of use, model lineage and release notes, to name a few), maintained and communicated to vital stakeholders and end users, where appropriate, throughout an AI system's lifecycle.

The overarching goal of transparency is to build confidence, literacy and trust in AI.

Transparency is crucial to upholding and supporting human agency and awareness in relation to AI systems by **enabling informed decision making over the use of AI systems and their outputs**.

**Related terms: explainability, interpretability**

1 Importance of current and future generational focus drawn from NIST AI RMF.  
2 Alignment with IEEE 7010-2020.

## EY commitment

Transparency is fundamental in our development and deployment of AI technologies. We understand the importance of maintaining appropriate levels of clarity over the data used to train our AI systems, as well as their objectives, limitations and conditions of use. We make relevant technical documentation and user instructions readily accessible as required by different stakeholders. In championing transparency, we aim to support human agency, fostering an environment where engagement with AI systems is informed, conscious and in line with our ethical and professional standards.



# Explainability

Appropriate levels of explanation are enabled so that the decision criteria and output of AI systems can be reasonably understood, challenged and validated by human operators.

## Importance to EY

Appropriate levels of explainability are crucial in building client, employee and public trust. It fosters confidence by helping stakeholders understand, validate and possibly challenge AI-derived output. This understanding can often assist human practitioners in identifying biases that may be embedded in AI system logic and therefore explainability also fosters a culture of accountability and continuous improvement, which is essential for innovation and help delivering high-quality services in a competitive global market.

Being able to appropriately explain AI generated outputs to affected stakeholders underpins EY commitment to being inclusive, fair and balanced. This is critical, especially for AI systems deployed in critical infrastructure or impacting protected groups, as safeguards to establish accountability of AI outputs.

Explainability is the **understanding** of an AI system and related output in a way that **makes sense to human beings**. It requires **appropriate detail** regarding the decision criteria, logic and other contributing factors to rationalize outputs. Explainability can be achieved in two ways: using algorithmic techniques that are explainable by design or adopting supplementary methods (e.g., surrogate models, local interpretable model-agnostic explanations (LIME) or Shapley Additive Explanations (SHAP)).

The requirement for an explanation, the type of explanation and the stakeholders requiring the explanation (operators, end users, regulators) should be **established at the outset** as this understanding informs the design of the AI system and the selection of supplementary explainability tools. Certain models like linear regression are inherently more explainable than other more “opaque” models such as random forests.<sup>1</sup>

Interpretability (being able to understand the inner model workings) and traceability (being able to follow the data and decision journey of an AI system) are related but slightly different terms. With GenAI, explainability may also include traceability of referenced data in generated outputs to flag unintentional bias, facilitate ongoing improvements and provide human oversight.

**Related terms: interpretability, traceability, transparency**

<sup>1</sup> Equally, there can be trade-offs between accuracy and explainability.

## EY commitment

We help bridge the gap between AI technology and users by helping enable an appropriate level of explainability. We support our AI solutions with clear, detailed explanations of how decisions are made in line with the requirements of each system to help users identify, challenge and correct errors and foster a culture of accountability and transparency.



# Fairness

The needs of all impacted stakeholders are assessed with respect to the design and use of AI systems and their outputs to promote a positive and inclusive societal impact.

## Importance to EY

EY has publicly highlighted the importance of diversity, equity and inclusiveness as catalysts for better decision-making, stimulating innovation, increasing organizational agility and strengthening resilience to disruption. Fairness is a critical precursor and reinforcement of diversity, equity and inclusion and aligns with our broader organizational values.

EY is a leader in amplifying the voices of and opportunities available to unrepresented groups (e.g., through EY Neuro-Diverse Centers of Excellence, EY Ripples, and our Global Equality Standard recognition). Mandating fairness in AI is a natural progression of these existing efforts to push for social equity and non-discrimination within the EY organization and beyond.

Fairness entails integrating nuanced **considerations of equity, non-discrimination, and justice** across the AI development lifecycle in a way that respects and upholds the democratic and pluralistic fabric of society, thereby identifying and minimizing algorithmic bias, and **enabling inclusive outcomes**.

**Group fairness** relates to equality of opportunity and establishes that protected and unprotected groups have an equal probability of being assigned to the positive predicted class. **Individual fairness** speaks to the equality of outcomes as it relates to the treatment of similar individuals.

The appropriate weighting of group and individual fairness is context-based and may fluctuate depending on the use case, cultural specificities and business objectives. Trade-offs made during the design process to this effect should be appropriately justified and documented for posterity.

**Algorithmic bias** is the systematic and repeatable deviations in a computer system that can create unfair outcomes. Algorithmic bias arises from prejudices and assumptions made during the initial conception, design, development, deployment and maintenance of AI products.

**Related terms: bias, discrimination**

<sup>1</sup> Equally, there can be trade-offs between accuracy and explainability.

## EY commitment

Fairness is at the core of our values and purpose. We take proactive measures to identify and mitigate bias throughout the lifecycle to help prevent the perpetuation of unfair and discriminatory AI systems and resulting output. Further to the EY organization's broader commitments to DE&I, we endeavor to engage a diverse set of stakeholders throughout the AI development lifecycle to help enable a broad evaluation of potential risks and regularly evaluate our approach to fairness to help ensure that it remains aligned with leading practice.



# Compliance

The design, implementation and use of AI systems and their outputs comply with relevant laws, regulations and professional standards.

## Importance to EY

Corporate compliance is an essential element of good governance and a strategic priority for EY. As one of the leading professional services firms in the world, EY has an important role to play both in acting as a trusted advisor in providing compliance services to clients, but also leading by example.

To remain a leader in responsible AI in the market it is crucial for EY to be able to evidence our own responsible and lawful use of AI systems and resulting outputs by maintaining a robust AI compliance framework in alignment with key AI laws and regulations.

Compliance supports positive interactions with regulators, client audit committees and other relevant stakeholders in a way that protects and augments EY reputation.

Compliance encompasses the need to **abide by laws, rules, guidelines and professional standards** that connect to the use and management of AI systems and underlying data across the regions in which the firm operates and the **EY Code of Conduct**.

It is key to build understanding and keep-pace with both existing and **emerging regulation** at local and global levels. As emerging AI laws often overlap or interact with existing laws, duplication in compliance efforts should be identified and avoided.<sup>1</sup> Further, to avoid fragmented AI governance, reasonable efforts should be undertaken to provide a consistent standard of compliance across operating regions.

Existing laws are also relevant. For example, **GenAI has compounded existing legal risks** as the information that was used to train the models may appear in the outputs it generates despite being subject to IP or copyright laws. In addition, users may not be aware that in using GenAI to generate code, the outputs may harness blocks of open-source code carrying a condition that the output too must be open source.

It should be noted though that a narrow focus on regulatory requirements may omit attention to business risks such as the need to review and update relevant contracts with third parties in the procurement of AI systems and services.

## Related terms: conformity

<sup>1</sup> The most comprehensive AI regulation is set to be the EU AI Act but other more established laws include the New York City Bias Act and China's GenAI regulation.

## EY commitment

We are committed to designing, developing, deploying and using AI responsibly and as such proactively preparing for forthcoming regulation and complying with existing regulation. As technology and regulations evolve, we track emerging issues across social, regulatory and ethical domains and feed them into processes that govern our AI systems.



# Sustainability

Considerations of the impacts of technology are embedded throughout the AI lifecycle to promote physical, social, economic and planetary well-being.

## Importance to EY

As environmental, social and governance (ESG) continues to dominate corporate agendas, including at EY, AI has emerged as a strategic enabler. AI systems can play a transformative role in achieving sustainability goals by optimizing energy usage in buildings, developing more sustainable supply chains, and enabling environmental monitoring, modelling, and simulation.

Simultaneously, it is also critical to enable AI that has been developed responsibly both in terms of minimized environmental impact and socially responsible supply chain to mitigate reputational damage and contribute positively to society. This imperative has been recognized by EY to date through a variety of existing commitments, including becoming carbon neutral and subsequently carbon negative, as well as EY Ripples.

Sustainability entails the design, development and deployment of AI systems and their outputs in a way that is human centric, environmentally friendly, socially responsible and economically viable, both for **current and future generations**<sup>1</sup> well-being.

- **Human centric** means that AI systems respect human rights, privacy and democratic values.
- **Environmentally friendly** means that AI systems should be designed and operated with minimal environmental impact.
- **Socially responsible** means that AI systems should be designed to be inclusive, accessible and beneficent.
- **Economically viable** means that AI systems should bring value to the economy, contributing to productivity while considering resources involved in development and maintenance.
- **Future-oriented** means that AI systems should be adaptable and resilient to changes and uncertainties in the technological, social and environmental landscape (within reason), and consider the well-being of both present and future generations.
- **Well-being** encompasses the full spectrum of personal, social and environmental factors that enhance human life and on which human life depend.<sup>2</sup>

Equally important is the use of AI to promote and achieve sustainability goals across various sectors. AI can help accelerate the deployment of existing sustainability solutions and the development of new ones.

<sup>1</sup> Importance of current and future generational focus drawn from NIST AI RMF.  
<sup>2</sup> Alignment with IEEE 7010-2020.

## EY commitment

We embed sustainability considerations into the approach to developing and procuring AI systems and in how we work with clients on AI engagements. We collaborate with governments, standards bodies, NGOs, academia and industry to shape policy and develop shared standards specific to enabling sustainable AI and leveraging AI for sustainability.

## EY | Building a better working world

EY exists to build a better working world, helping to create long-term value for clients, people and society and build trust in the capital markets.

Enabled by data and technology, diverse EY teams in over 150 countries provide trust through assurance and help clients grow, transform and operate.

Working across assurance, consulting, law, strategy, tax and transactions, EY teams ask better questions to find new answers for the complex issues facing our world today.

EY refers to the global organization, and may refer to one or more, of the member firms of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a UK company limited by guarantee, does not provide services to clients. Information about how EY collects and uses personal data and a description of the rights individuals have under data protection legislation are available via [ey.com/privacy](https://ey.com/privacy). EY member firms do not practice law where prohibited by local laws. For more information about our organization, please visit [ey.com](https://ey.com).

©2025 EYGM Limited.  
All Rights Reserved.

ED None  
EYG no. 008218-24GbI

This material has been prepared for general informational purposes only and is not intended to be relied upon as accounting, tax, legal or other professional advice. Please refer to your advisors for specific advice.

[ey.com](https://ey.com)