THE ETHICS OF AI:

Tackling Bias, Privacy, Security, Amplification, and Inclusiveness

MANOJ TAVARAJOO

August 2025

The Ethics of AI: Tackling Bias, Privacy, Security, Amplification, and Inclusiveness

Manoj Tavarajoo | AI Essentials for Leaders Series – Article 11

Let's connect on LinkedIn

AI is transforming industries and creating new opportunities for growth, but it also amplifies risks at scale. Unlike traditional technologies, AI systems operate through algorithms that learn from data and adapt in real time. This makes them powerful but also prone to bias, privacy violations, security vulnerabilities, and unintended amplification of harmful content.

For leaders, addressing AI ethics is not optional. Responsible use of AI is now a strategic imperative that affects trust, brand reputation, regulatory compliance, and ultimately competitive advantage.

1. Algorithmic Bias: When Data Embeds Inequality

AI models learn from historical data, and if that data reflects human biases, the models replicate and amplify them.

- **Case example – Amazon's hiring algorithm:** Amazon scrapped an internal AI recruitment tool after it was found to downgrade CVs containing words like "women's" because it was trained on data skewed toward male applicants.
- **Case example – Healthcare algorithms:** A widely used algorithm in the US healthcare system underestimated health risks for Black patients because it used healthcare spending as a proxy for need, ignoring systemic inequalities in access to care.

Bias often emerges because training data is incomplete, unrepresentative, or reflects systemic inequality. When deployed at scale, these biases can affect millions of people in ways that are unfair and damaging.

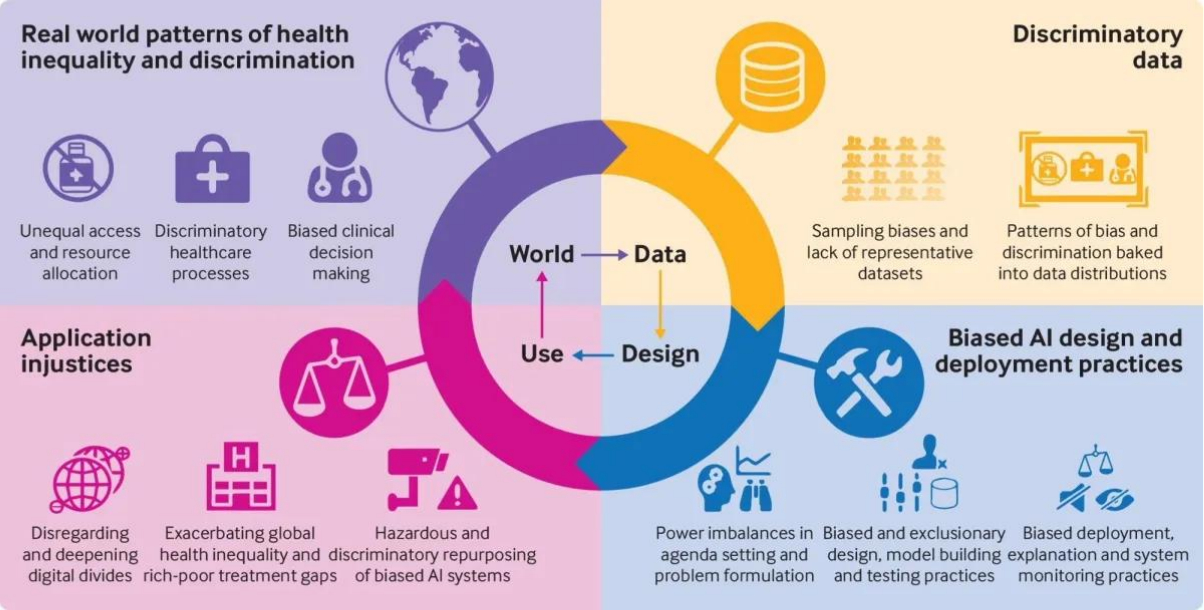

Real world patterns of health inequality and discrimination
Unequal access and resource allocation
Discriminatory healthcare processes
Biased clinical decision making
Discriminatory data
Sampling biases and lack of representative datasets
Patterns of bias and discrimination baked into data distributions
World
Data
Design
Use
Application injustices
Disregarding and deepening digital divides
Exacerbating global health inequality and rich-poor treatment gaps
Hazardous and discriminatory repurposing of biased AI systems
Biased AI design and deployment practices
Power imbalances in agenda setting and problem formulation
Biased and exclusionary design, model building and testing practices
Biased deployment, explanation and system monitoring practices

Figure 1: How Bias Enters AI Systems: Inequality and discrimination in the design and use of AI in healthcare applications (Source: British Medical Journal)

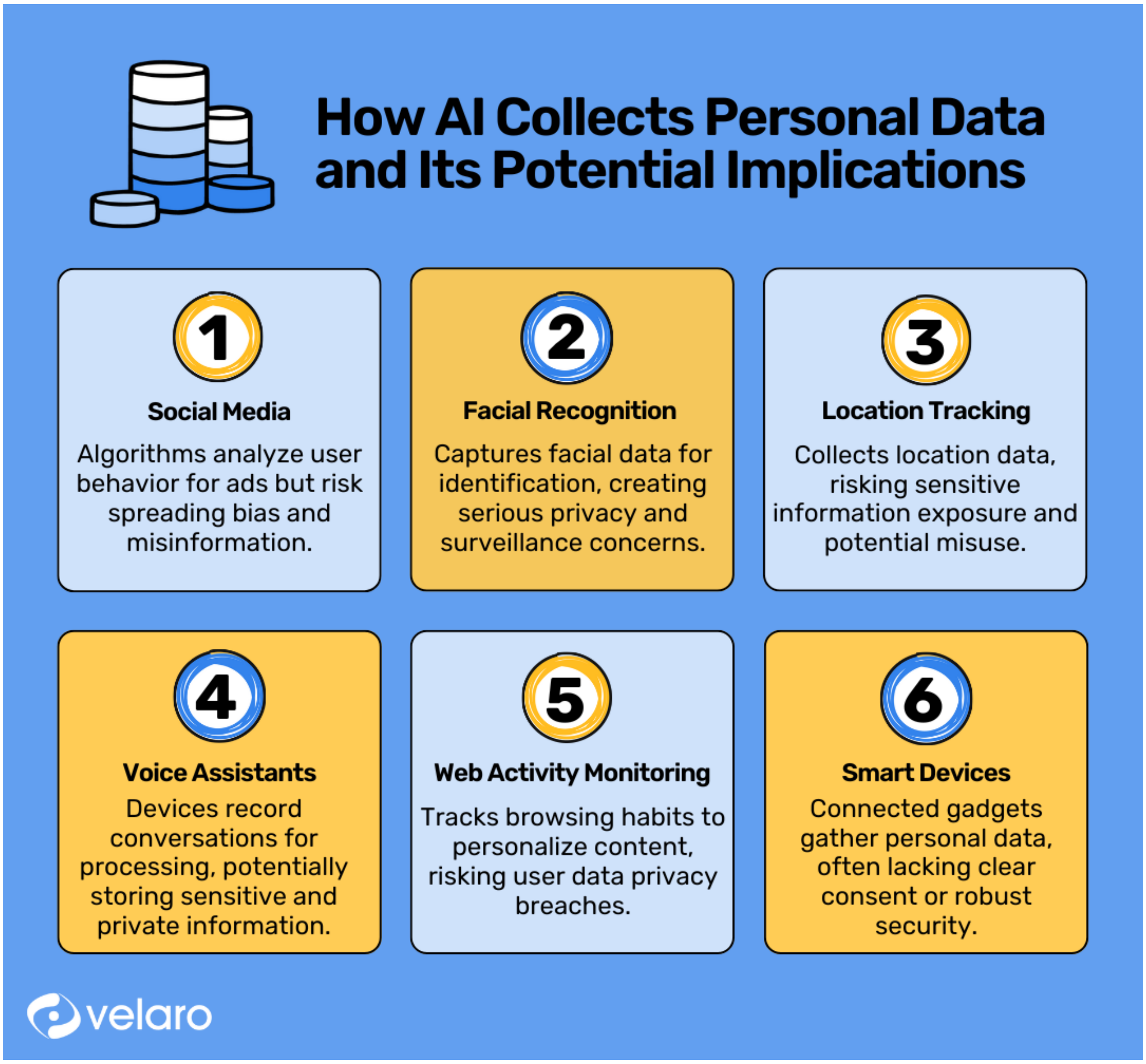
Leadership imperative: Establish governance to monitor training data, audit algorithms, and include diverse voices in model design and testing.

2. Privacy: Protecting Data in a Surveillance Age

AI thrives on data. The more it collects, the better it performs. But the hunger for data raises serious concerns about privacy and surveillance.

- **Case example – Cambridge Analytica:** Data from millions of Facebook users was harvested without consent and used to influence elections, sparking global debate on data rights and corporate responsibility.
- **Case example – Apple's privacy-first stance:** Apple has positioned itself as a defender of user privacy, restricting app tracking and promoting features like "App Tracking Transparency." This has both strengthened its brand and reshaped digital advertising markets.

As AI becomes more embedded in daily life, the line between convenience and intrusion blurs. Leaders must ensure their organisations respect user consent, minimise unnecessary data collection, and comply with evolving data protection regulations.



How AI Collects Personal Data and Its Potential Implications

1 **Social Media**
Algorithms analyze user behavior for ads but risk spreading bias and misinformation.

2 **Facial Recognition**
Captures facial data for identification, creating serious privacy and surveillance concerns.

3 **Location Tracking**
Collects location data, risking sensitive information exposure and potential misuse.

4 **Voice Assistants**
Devices record conversations for processing, potentially storing sensitive and private information.

5 **Web Activity Monitoring**
Tracks browsing habits to personalize content, risking user data privacy breaches.

6 **Smart Devices**
Connected gadgets gather personal data, often lacking clear consent or robust security.

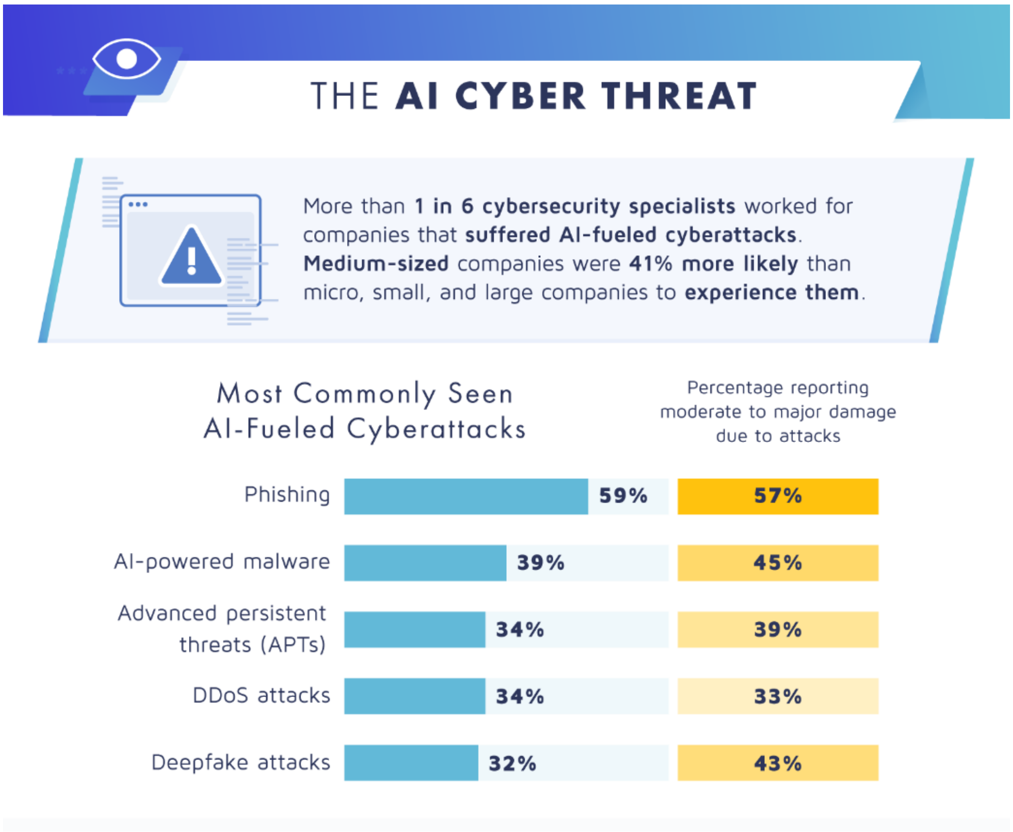
Figure 2: Data Privacy Risks in AI: How AI Collects Personal Data and Its Potential Implications (Source: Velaro)

Leadership imperative: Build privacy by design, implement data minimisation principles, and ensure transparency with customers about how their data is used.

3. Cybersecurity: Defending Against AI-Enabled Threats

AI not only introduces new vulnerabilities, it also creates entirely new attack vectors. Malicious actors are now using AI to probe systems, automate attacks, and generate convincing disinformation.

- **Case example – Deepfakes:** AI-generated synthetic media used for fraud, reputational attacks, and misinformation campaigns.
- **Case example – Adversarial attacks:** Small manipulations to input data (such as a few pixels in an image) can fool AI systems into making dangerous errors.
- **Case example – Critical infrastructure:** Power grids, transport networks, and healthcare systems face heightened cyber risk when AI is embedded without adequate safeguards.



More than **1 in 6 cybersecurity specialists** worked for companies that **suffered AI-fueled cyberattacks**. **Medium-sized** companies were **41% more likely** than micro, small, and large companies to **experience them**.

Figure 3: Most Commonly Seen AI-Fuelled Cyberattacks

(Source: Digital Information World)

Leadership imperative: Invest in resilient AI architectures, continuous monitoring, and red-teaming to stress test AI systems against attack.

4. Digital Amplification: AI at Scale

One of the unique risks of AI is **amplification**. AI systems can magnify both the positive and negative effects of content, decisions, or processes.

- **Case example – YouTube recommendations:** YouTube's recommendation engine has been criticised for pushing users toward increasingly extreme content because its optimisation favoured engagement above all else.
- **Case example – Twitter misinformation:** During the COVID-19 pandemic, AI-driven content ranking amplified misinformation, leading to global health consequences and regulatory scrutiny.

When amplification goes wrong, the consequences are rapid and far-reaching. What might have been a local issue in a pre-digital world can now escalate globally in seconds.

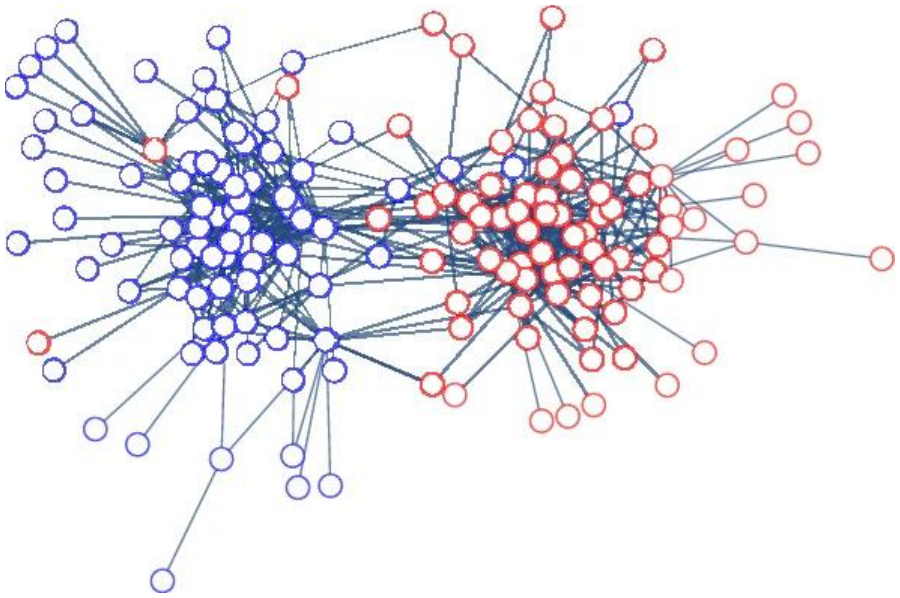


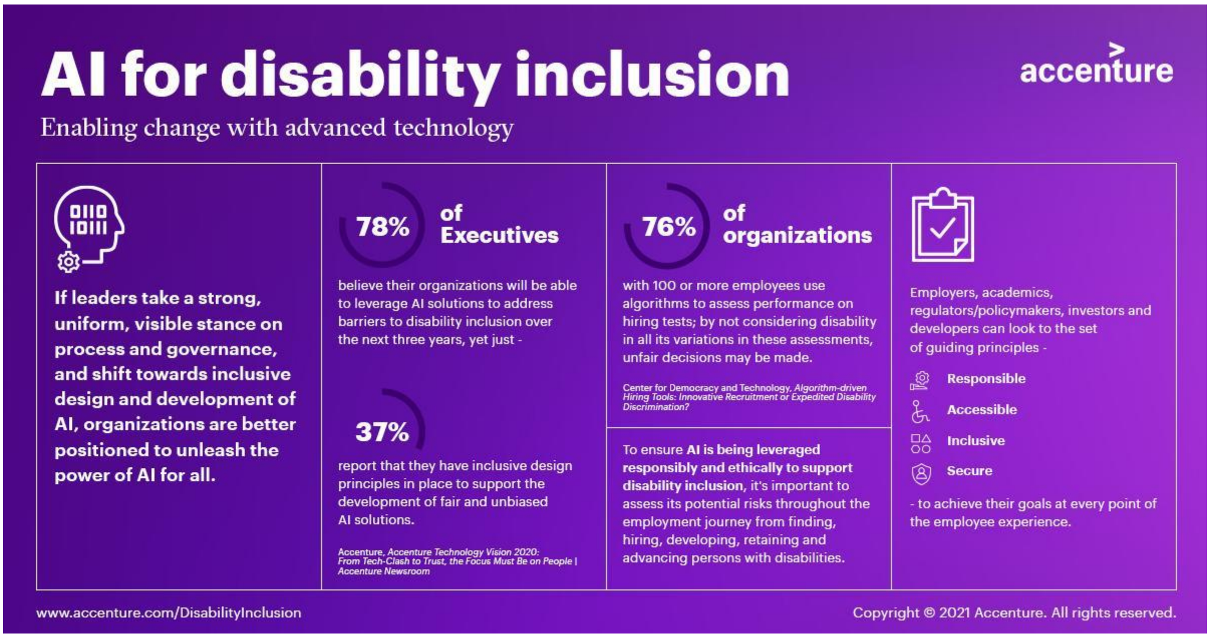
Figure 4: Digital Amplification Cycle: Social media echo chambers under passive regulation. AI recommendations create feedback loops that amplify both positive and negative effects.

Leadership imperative: Balance scale with responsibility. Put guardrails in place to prevent misuse and design systems that optimise not just for engagement or efficiency, but for long-term trust and societal good.

5. Inclusiveness: Designing for All

AI systems risk leaving people behind if inclusiveness is not deliberately embedded. Models trained on narrow populations or designed without diverse input can marginalise whole groups.

- **Case example – Voice assistants:** Early systems struggled to recognise female voices or accents outside American English.
- **Case example – Accessibility gaps:** Many AI-driven apps are not designed with accessibility features for people with disabilities.
- **Case example – Financial inclusion:** Fintech AI can expand access to credit in underserved regions, but only if models are trained on diverse data.



AI for disability inclusion

Enabling change with advanced technology

accenture

If leaders take a strong, uniform, visible stance on process and governance, and shift towards inclusive design and development of AI, organizations are better positioned to unleash the power of AI for all.

78% of Executives

believe their organizations will be able to leverage AI solutions to address barriers to disability inclusion over the next three years, yet just -

37%

report that they have inclusive design principles in place to support the development of fair and unbiased AI solutions.

Accenture, *Accenture Technology Vision 2020: From Tech-Clash to Trust, the Focus Must Be on People | Accenture Newsroom*

76% of organizations

with 100 or more employees use algorithms to assess performance on hiring tests; by not considering disability in all its variations in these assessments, unfair decisions may be made.

Center for Democracy and Technology, *Algorithm-driven Hiring Tools: Innovative Recruitment or Expedited Disability Discrimination?*

To ensure **AI is being leveraged responsibly and ethically to support disability inclusion**, it's important to assess its potential risks throughout the employment journey from finding, hiring, developing, retaining and advancing persons with disabilities.

Employers, academics, regulators/policymakers, investors and developers can look to the set of guiding principles -

- **Responsible**
- **Accessible**
- **Inclusive**
- **Secure**

- to achieve their goals at every point of the employee experience.

www.accenture.com/DisabilityInclusion

Copyright © 2021 Accenture. All rights reserved.

Figure 5: AI for Disability Inclusion (Source: Accenture)

Leadership imperative: Embed inclusiveness by involving diverse stakeholders, ensuring datasets represent broad populations, and testing outcomes across demographics.

6. Building Responsible AI

Tackling bias, privacy, cybersecurity, amplification, and inclusiveness is not just a compliance exercise. It requires a proactive approach embedded into the organisation's DNA. Leaders must:

1. **Establish governance:** Create oversight structures with clear accountability for AI outcomes.
2. **Embed ethics into design:** Ensure teams include diverse perspectives and test for unintended consequences.
3. **Adopt transparency:** Communicate how AI models work and provide explainability for decisions.
4. **Collaborate externally:** Work with regulators, civil society, and industry peers to set standards for responsible AI.

- **Case example – Microsoft's Responsible AI Standard:** Microsoft has published detailed guidelines and governance structures for building AI responsibly, including fairness, reliability, privacy, inclusiveness, and accountability.
- **Case example – Google's AI Principles:** After facing criticism for biased and opaque systems, Google formalised a set of principles to guide AI development, including commitments not to pursue technologies that cause harm.

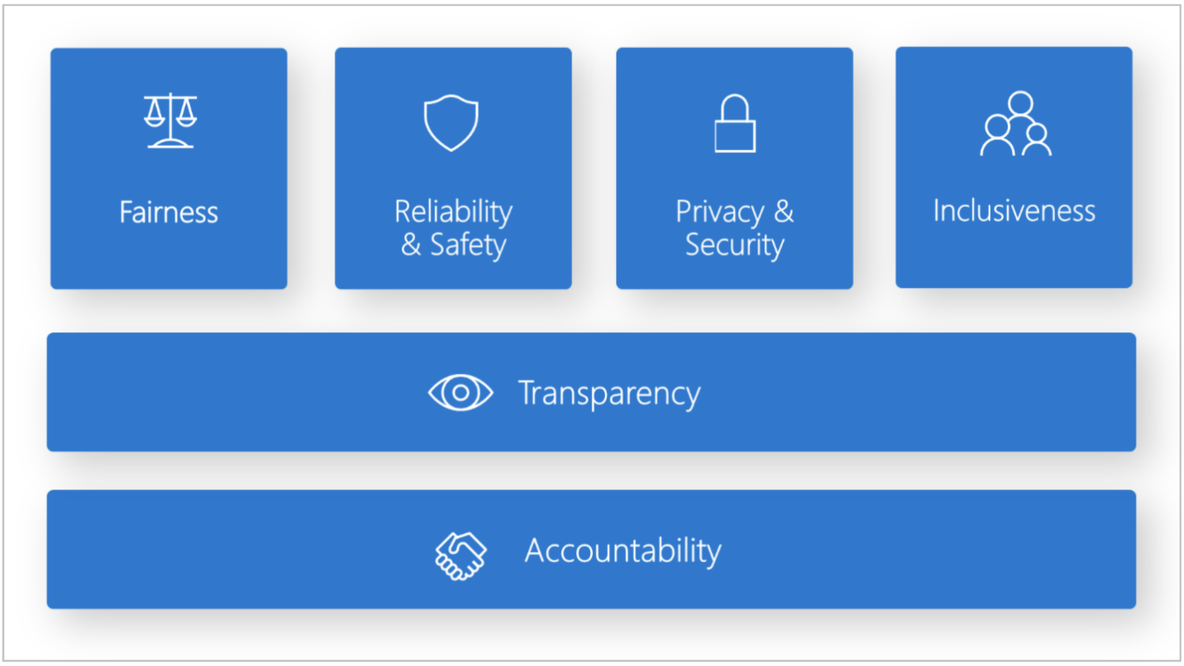


Figure 6: Microsoft's Responsible AI Framework (Source: Microsoft)

The Takeaway for Leaders

AI's promise and perils are inseparable. Leaders who treat ethics as an afterthought will face reputational damage, regulatory backlash, and loss of trust. Those who make ethics central to their AI strategy will earn trust, strengthen relationships with customers and regulators, and create long-term advantage.

Leader questions:

- Are we actively monitoring and mitigating bias in our AI systems?
- Do we embed privacy protection into every AI initiative?
- Are we safeguarding against AI-enabled cybersecurity risks?
- How are we ensuring our AI systems do not amplify harm at scale?
- Are our AI models inclusive and accessible to all users?

Up next:

Becoming an AI-First Organisation: Frameworks for Enterprise Transformation

MyConsultancy

Accelerating Digital Transformation and AI Maturity

www.myconsultancy.com.au