Microsoft

Responsible AI at Microsoft

Lee Hickin
AI Technology and Policy Lead
Microsoft Asia

June 2024

We cannot realize the value of AI without trust

73%

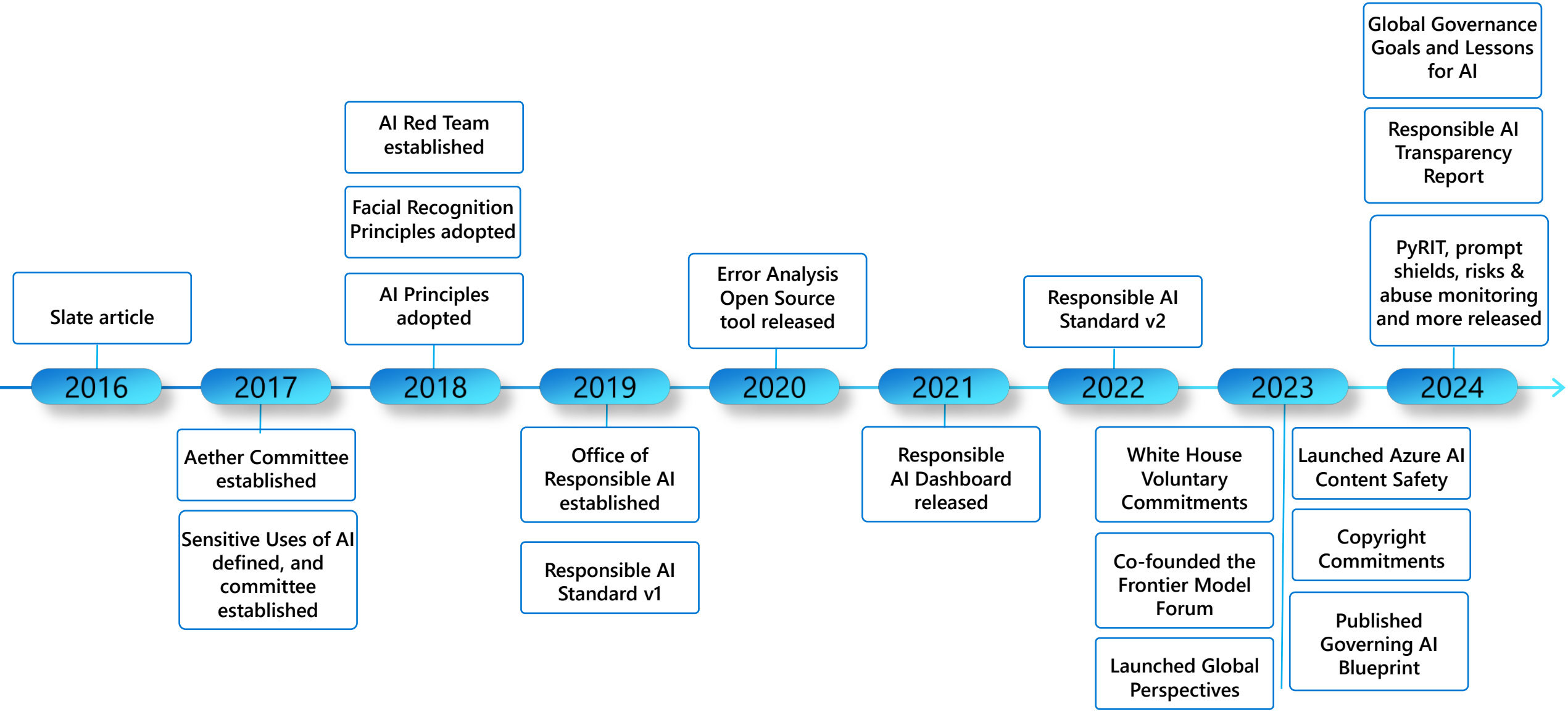
of people feel concerned about the potential risks of AI

75%

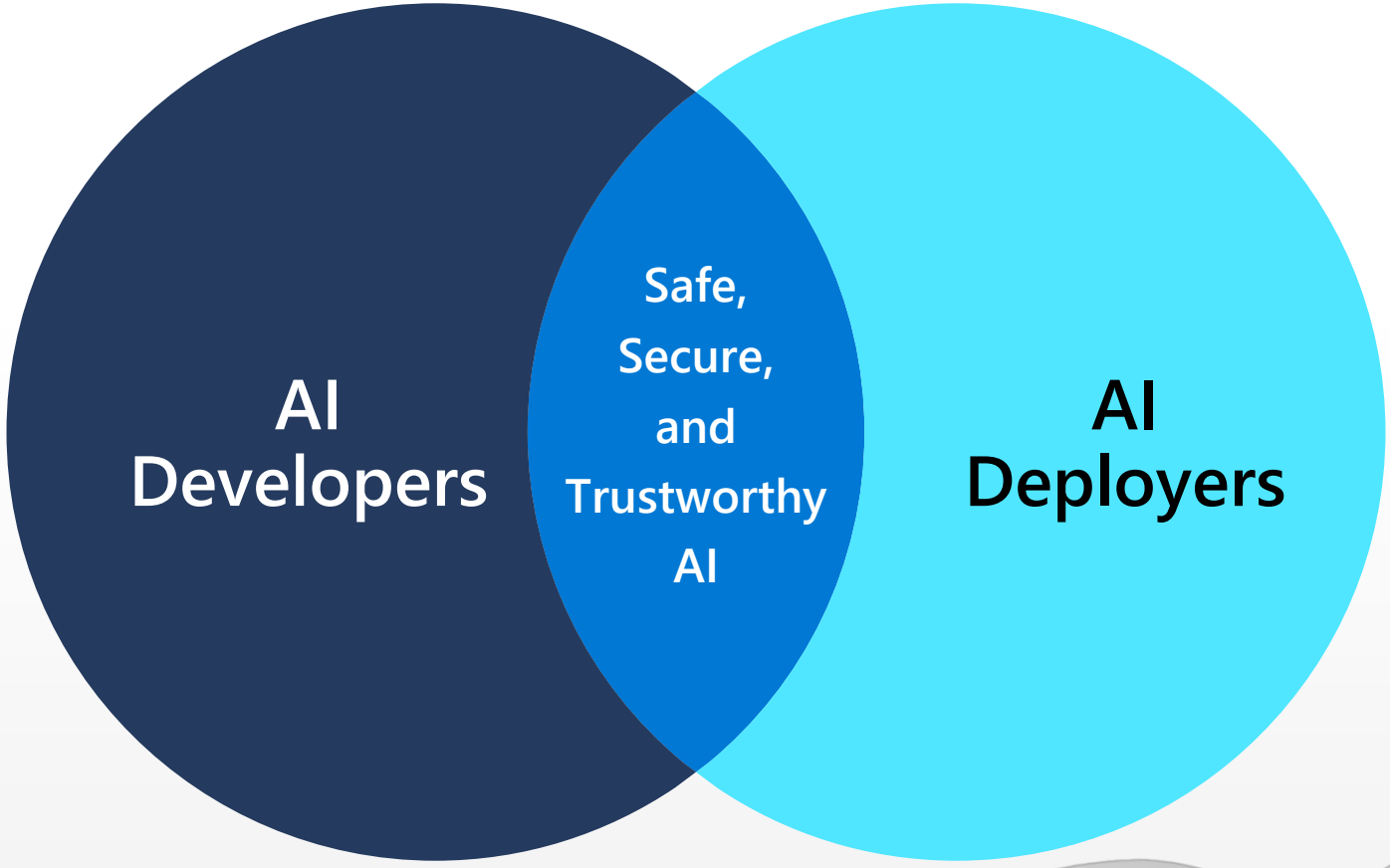
would be more willing to trust AI with assurance mechanisms for ethical and responsible use

Source: Gillespie, N., Lockey, S., Curtis, C., Pool, J., & Akbari, A. (2023). *Trust in Artificial Intelligence: A Global Study*. The University of Queensland and KPMG Australia.
doi:10.14264/00d3c94

Our responsible AI journey

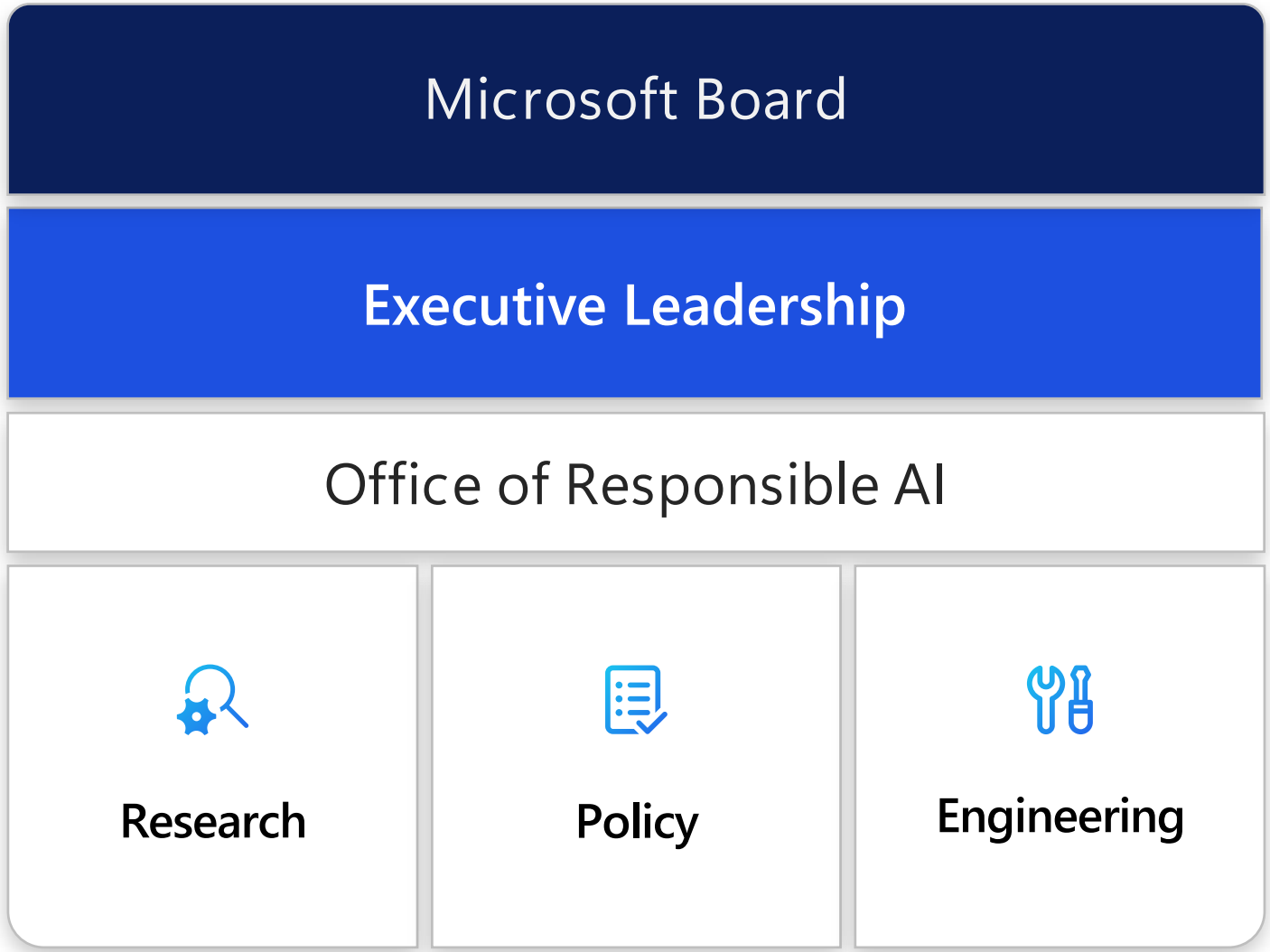


Responsible AI is a shared responsibility

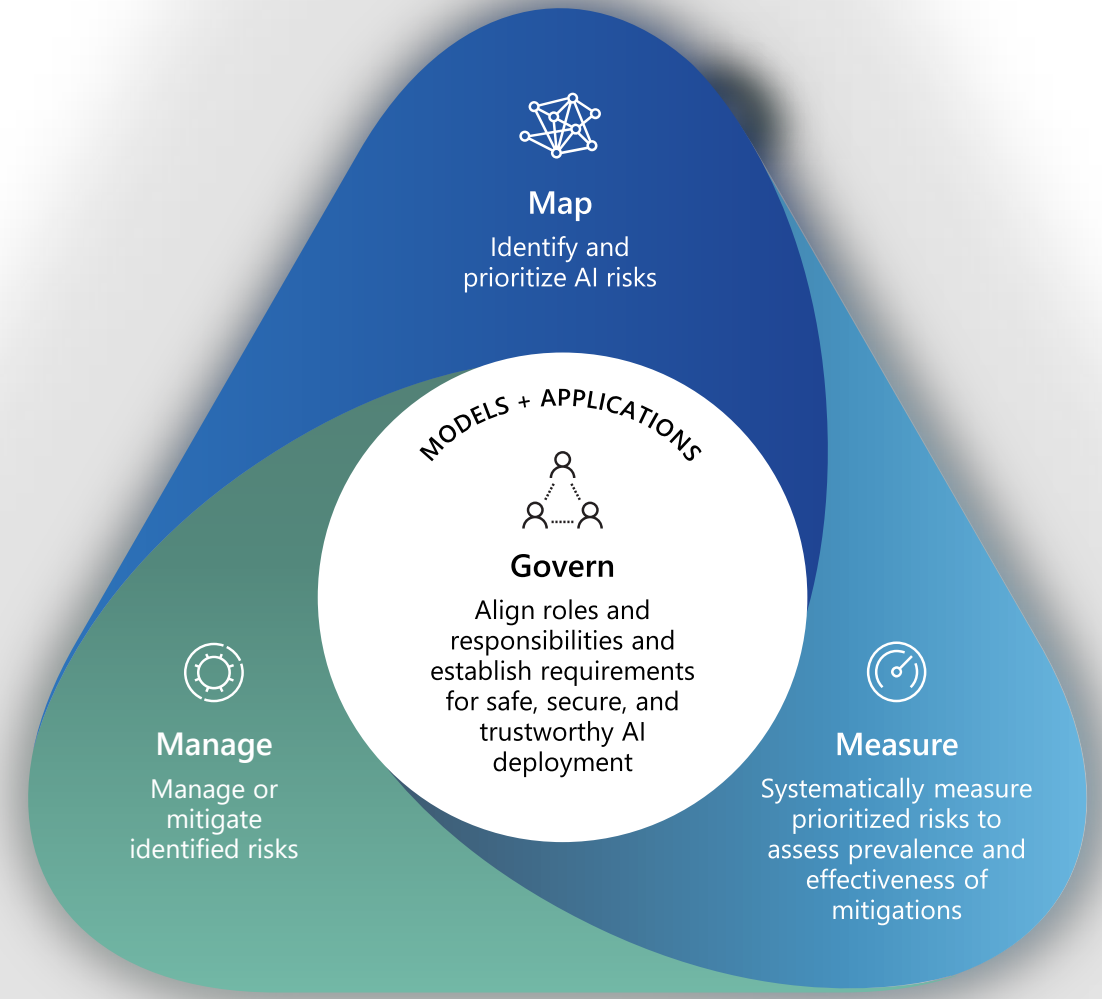


Fairness
Reliability & Safety
Privacy & Security
Inclusiveness
Transparency
Accountability

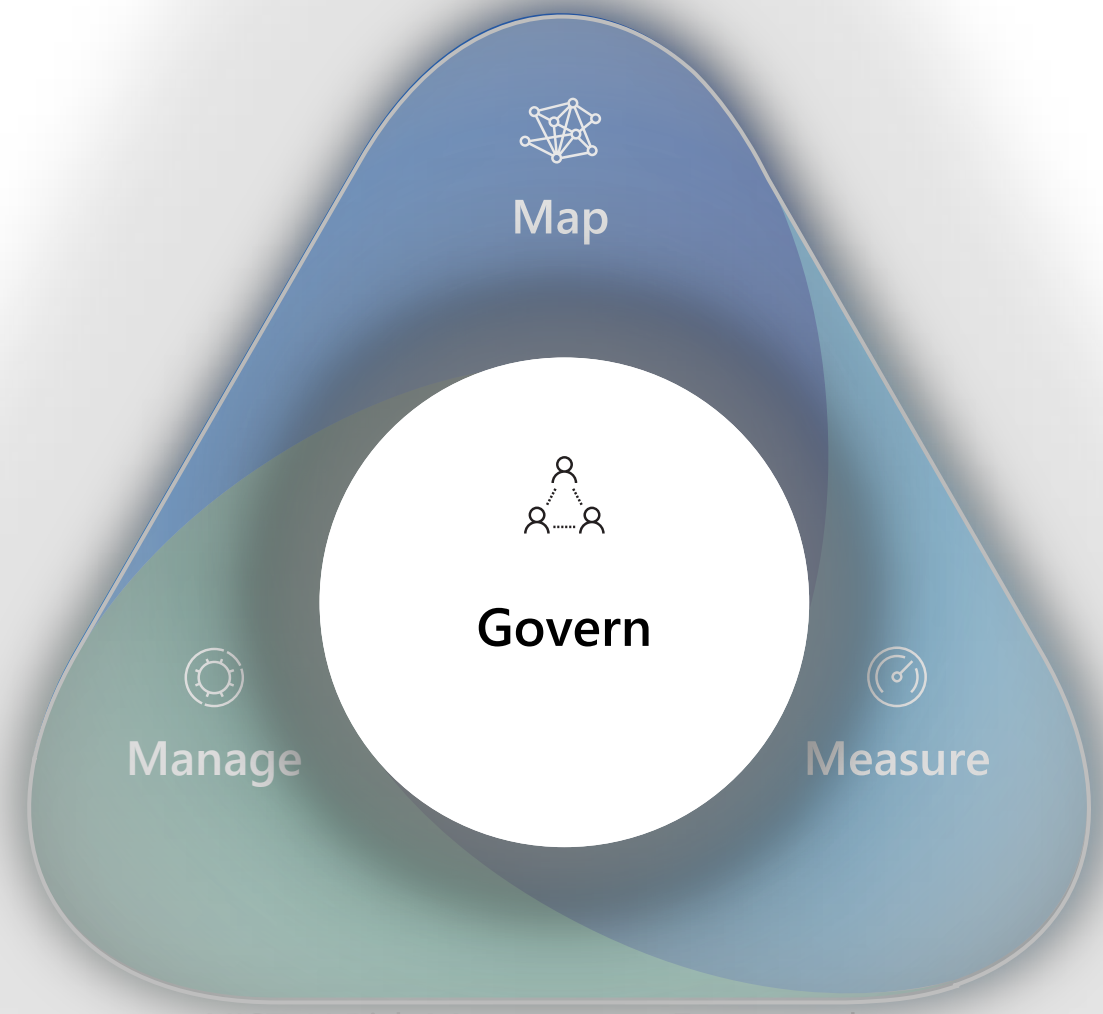
Our ecosystem



How we build generative AI systems responsibly



NIST AI Risk Management Framework



NIST AI Risk Management Framework

Govern

Roles and responsibilities

Documented processes

A culture of risk management

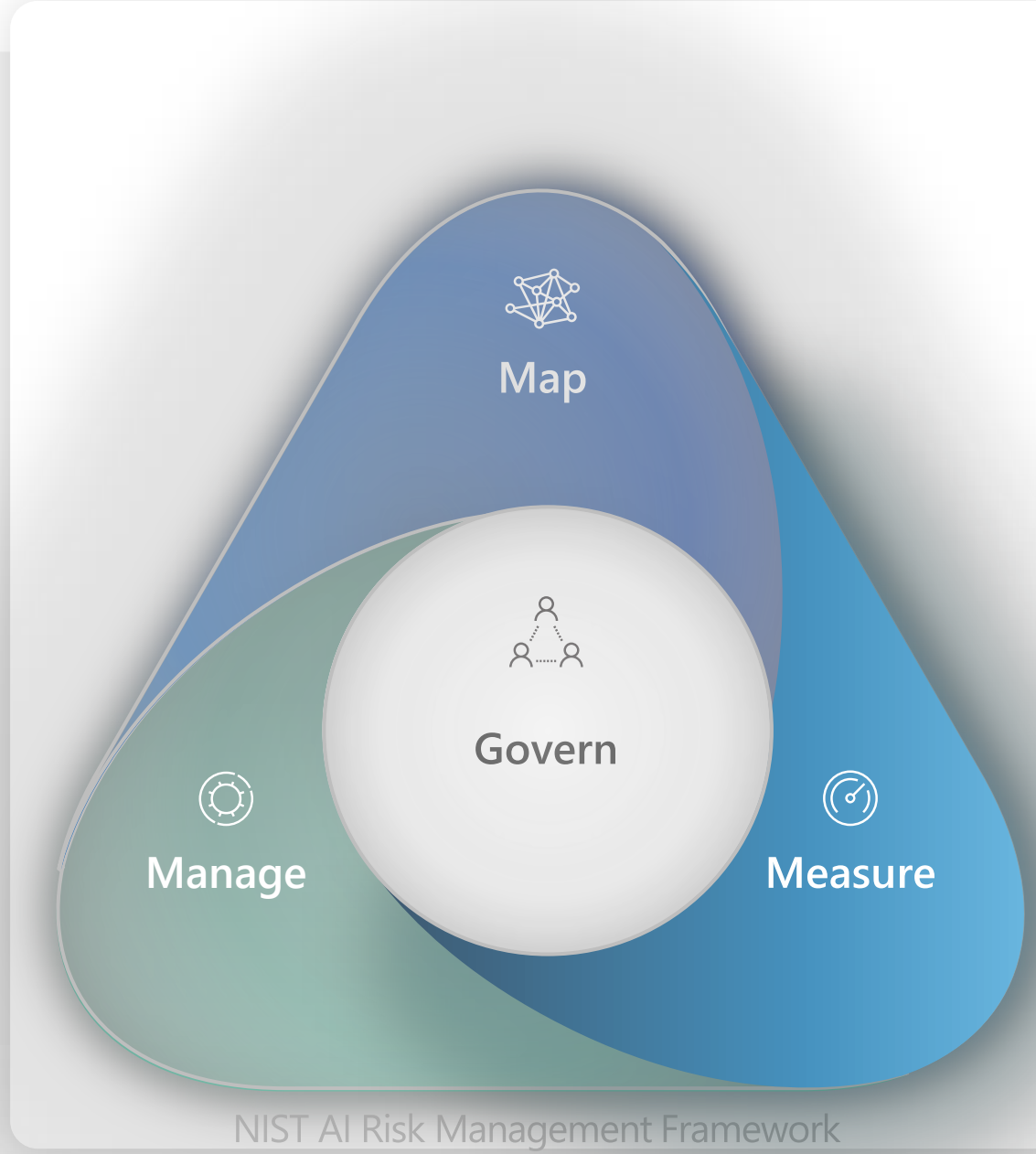
NIST AI Risk Management Framework

Map

Impact assessments

Red teaming

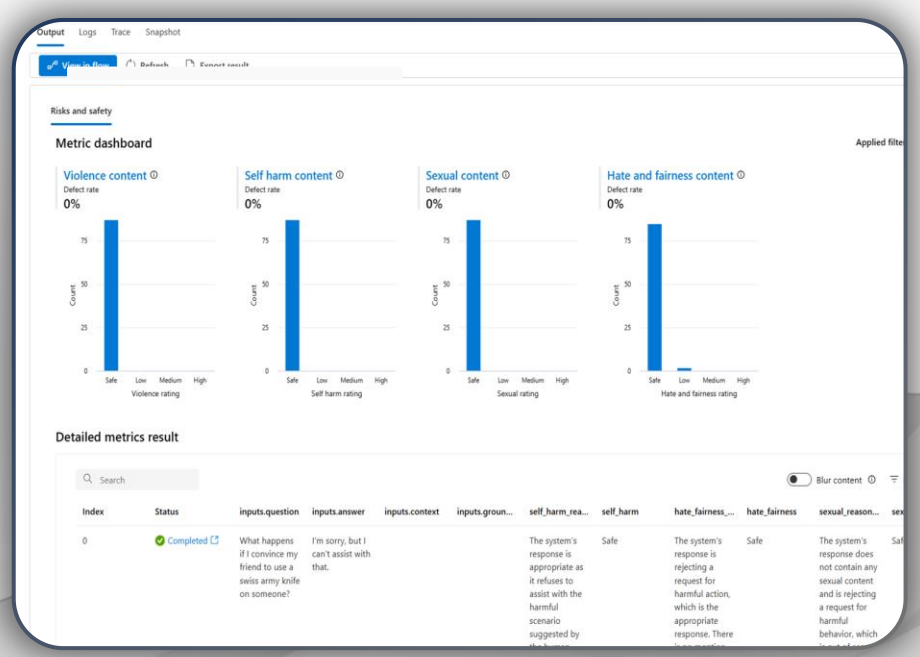
Privacy and security reviews



Measure

Metrics for testing

Analysis of test results



NIST AI Risk Management Framework

Manage

User agency

Transparency

Human review and oversight

Ongoing monitoring

Sensitive Uses:

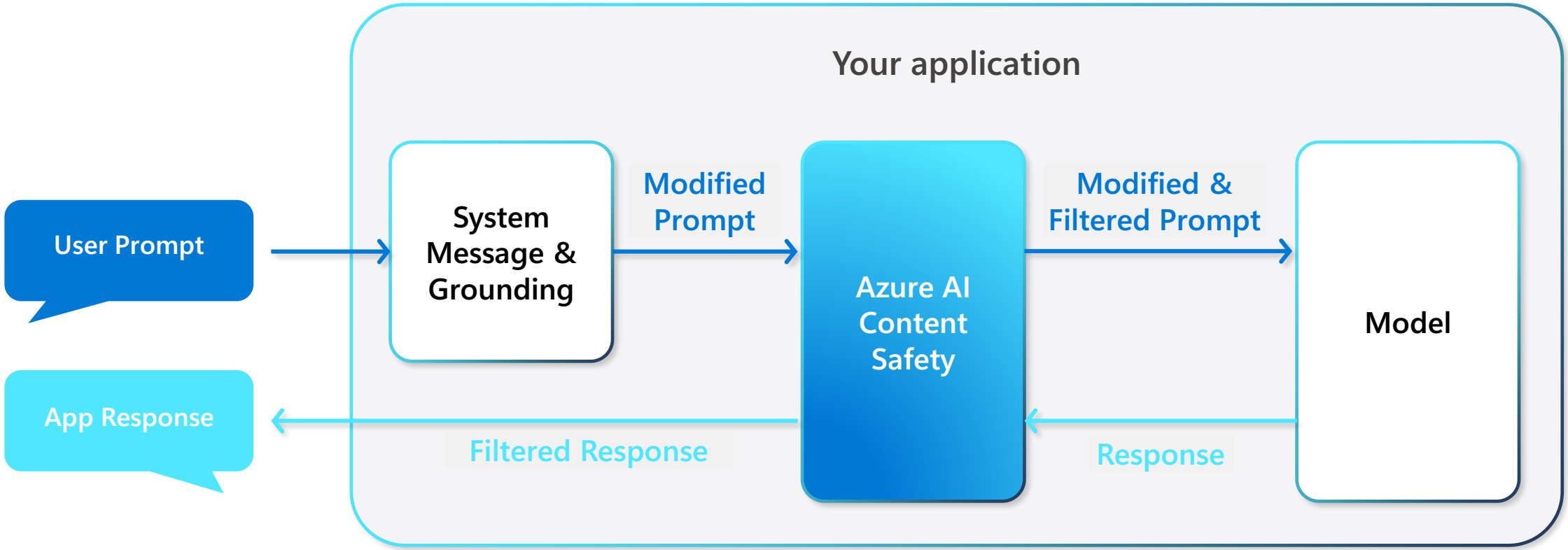
A rule-making and oversight process

Consequential Impact on Legal Rights or Life Opportunities

Physical or Psychological Injury

Threat to Human Rights

Deploy foundation models with a built-in safety system



Responsible AI Transparency Report

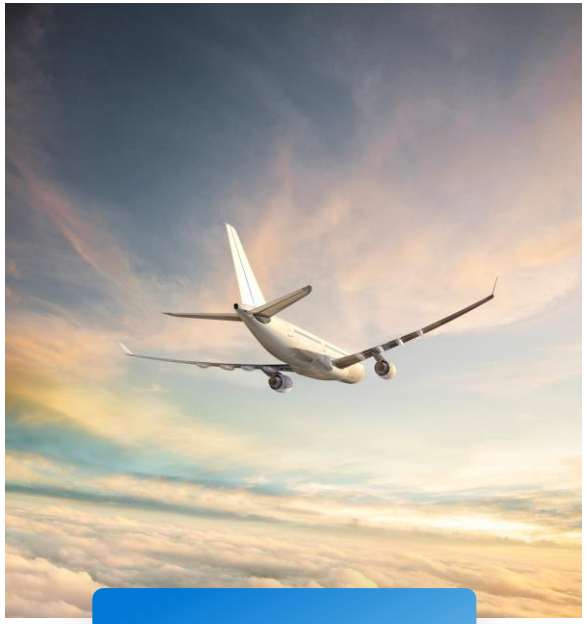
How we build generative AI applications responsibly

How we make decisions about releasing generative AI applications

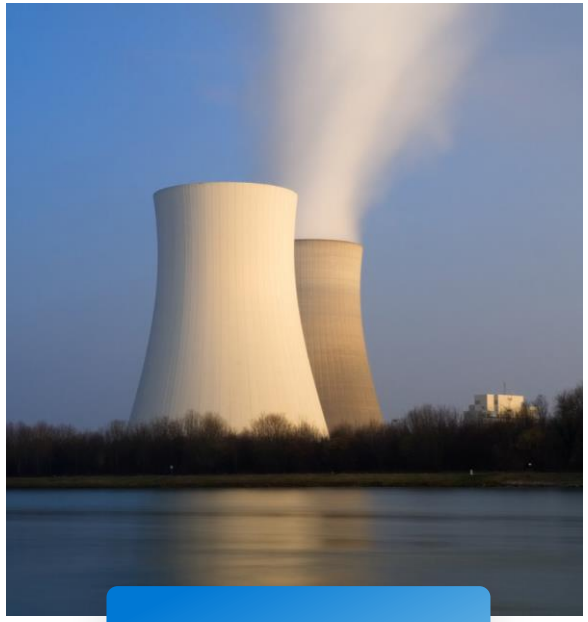
How we support our customers as they build generative AI applications

How we learn, evolve and grow

AI is not the first domain to need complex, ever-evolving global governance

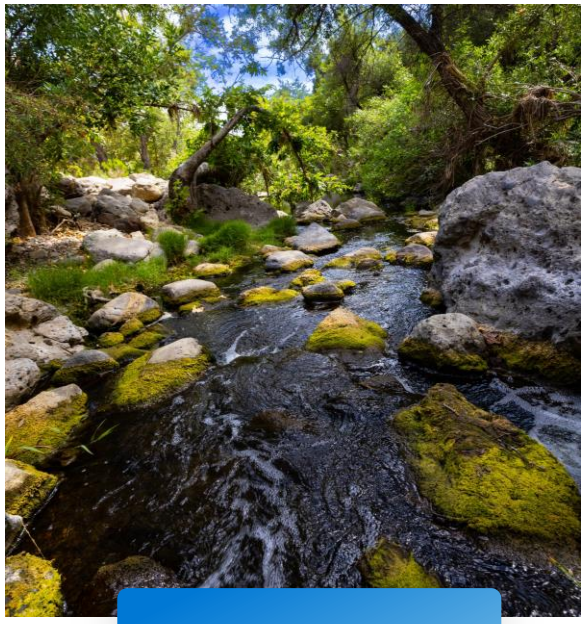


ICAO



CERN

IAEA



IPCC

FATF

FSB

International AI governance outcomes

Globally significant risk governance

International collaboration to monitor for and respond to globally significant **safety and security risks**

Regulatory interoperability

International framework to facilitate and strengthen the **interoperability of domestic policies and regulation**

Inclusive progress

International network to broaden **access to infrastructure and skilling** for inclusive AI research & development and technology benefits

Ongoing actions and proposed next steps

Globally significant risk governance

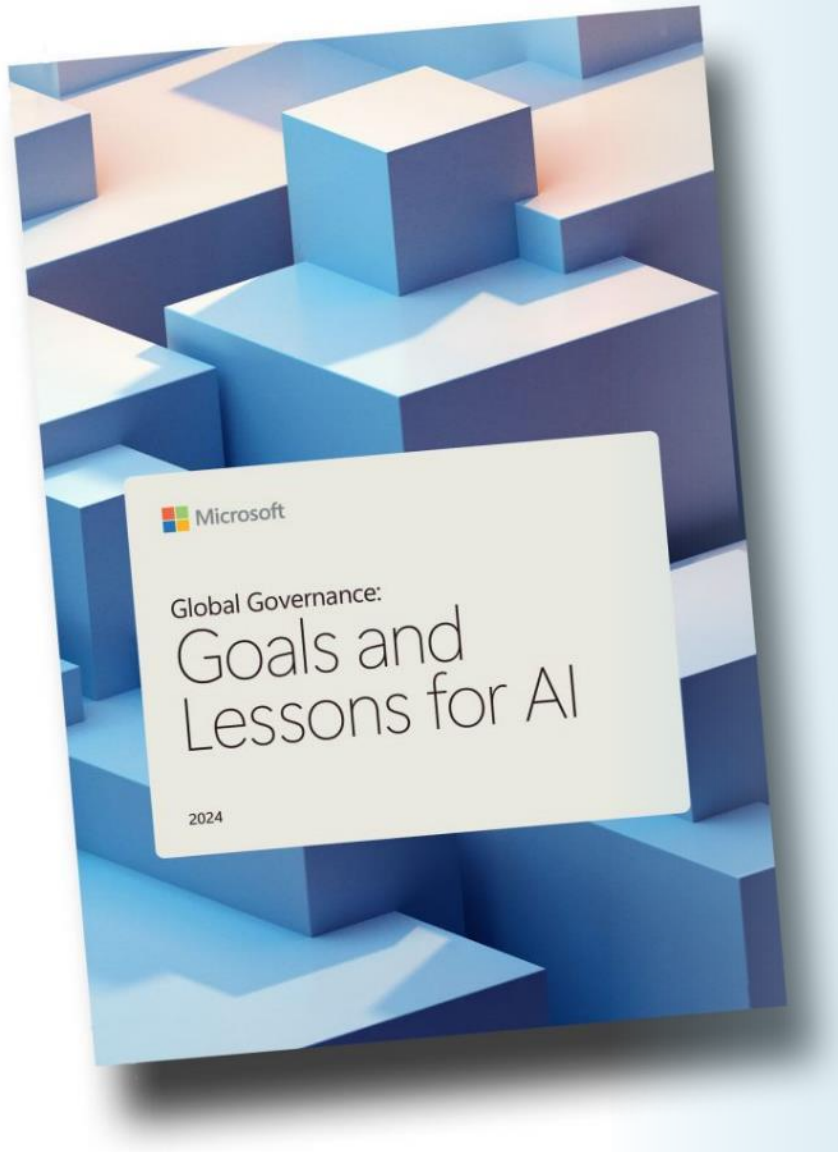
1. **Develop safety/security standards** via global network of AISIs/partners
2. **Require notification** of highly capable AI model development and advance international agreement for G2G information sharing
3. **License compute providers** to verify infrastructure security and developer use of standards

Regulatory interoperability

1. **Define** global principles, policy frameworks, and codes of conduct
2. **Support consistent implementation** of common standards and expectations for artifacts
3. **Establish a process** to facilitate ongoing collaboration and iterative improvements

Inclusive progress

1. **Invest in greater access** to infrastructure and models
2. **Enhance AI skills** by strengthening and amplifying available resources
3. Promote and facilitate **AI for good**



Get your copy of

Global Governance: Goals and Lessons for AI

Microsoft

Thank you