From Principles to Practice: IBM’s Approach to AI Governance

Presentation Before the Colorado AI Taskforce

Ryan Hagemann
Global AI Policy Issue Lead
IBM

Principles for Trust and Transparency

1 The purpose of AI is to augment — not replace — human intelligence

2 Data and insights belong to their creator

3 New technology, including AI systems, must be transparent and explainable

Pillars of Trust

Explainability

An AI system’s ability to provide a human-interpretable explanation for its predictions and insights

Fairness

Equitable treatment of individuals or groups by an AI system — depends on the context in which the AI system is used

Robustness

An AI system’s ability to effectively handle exceptional conditions, such as abnormalities in input

Transparency

An AI system’s ability to include and share information on how it has been designed and developed

Privacy

An AI system’s ability to prioritize and safeguard consumers’ privacy and data rights

Key IBM policy[1]

Regulate AI risk, not AI algorithms

Because each AI application is unique, we strongly believe that regulation must account for the context in which AI is deployed and must ensure that high-risk uses of AI are regulated more closely.

Make AI creators and deployers accountable, not immune to liability

Legislation should consider the different roles of AI creators and deployers and hold them accountable in the context in which they develop or deploy AI.

Support open AI innovation, not an AI licensing regime

An AI licensing regime would be a serious blow to open innovation and risks creating a form of regulatory capture.

[1] As articulated in Chairman and CEO Arvind Krishna's Sept. 2023 regulatory POV on advancing trusted AI and Chief Privacy and Trust Officer Christina Montgomery's May 2023 testimony before the US Senate Judiciary Committee

Key policy and thought leadership

Foundation models: Opportunities, risks, and mitigations

Version 2.0 of IBM's point of view on foundation models expands our taxonomy of risks and outlines examples of risks in practice.

The EU AI Act Is About to Hit the Books: Compliance Steps You Need to Know

The EU AI Act has ushered in a new era for AI governance. IBM welcomes the Act and its risk-based approach to regulating AI.

Why We Must Protect an Open Innovation Ecosystem for AI

The only way to guarantee the transformative changes of AI can be harnessed by all is to ensure that the future of AI is open.

At IBM, we believe that an open innovation ecosystem for AI is critical to ensuring the benefits of AI are distributed broadly throughout society, and that development coexists with safety.

→ Open is safe.

→ Open is innovation.

→ Open is opportunity.

IBM Granite models

In May 2024, IBM IBM released a family of Granite models into open source, inviting clients, developers and global experts to push the boundaries of what AI can achieve in enterprise environments.

InstructLab

In May 2024, IBM and Red Hat launched InstructLab, an open source project for enhancing LLMs through through constant incremental contributions, much like software development has worked in open source for decades.

The AI Alliance

In December 2023, IBM and Meta co-founded the AI Alliance, which has grown from 50 founding members to an active, international community of over 100 leading organizations with a mission to build and support open technology for AI.

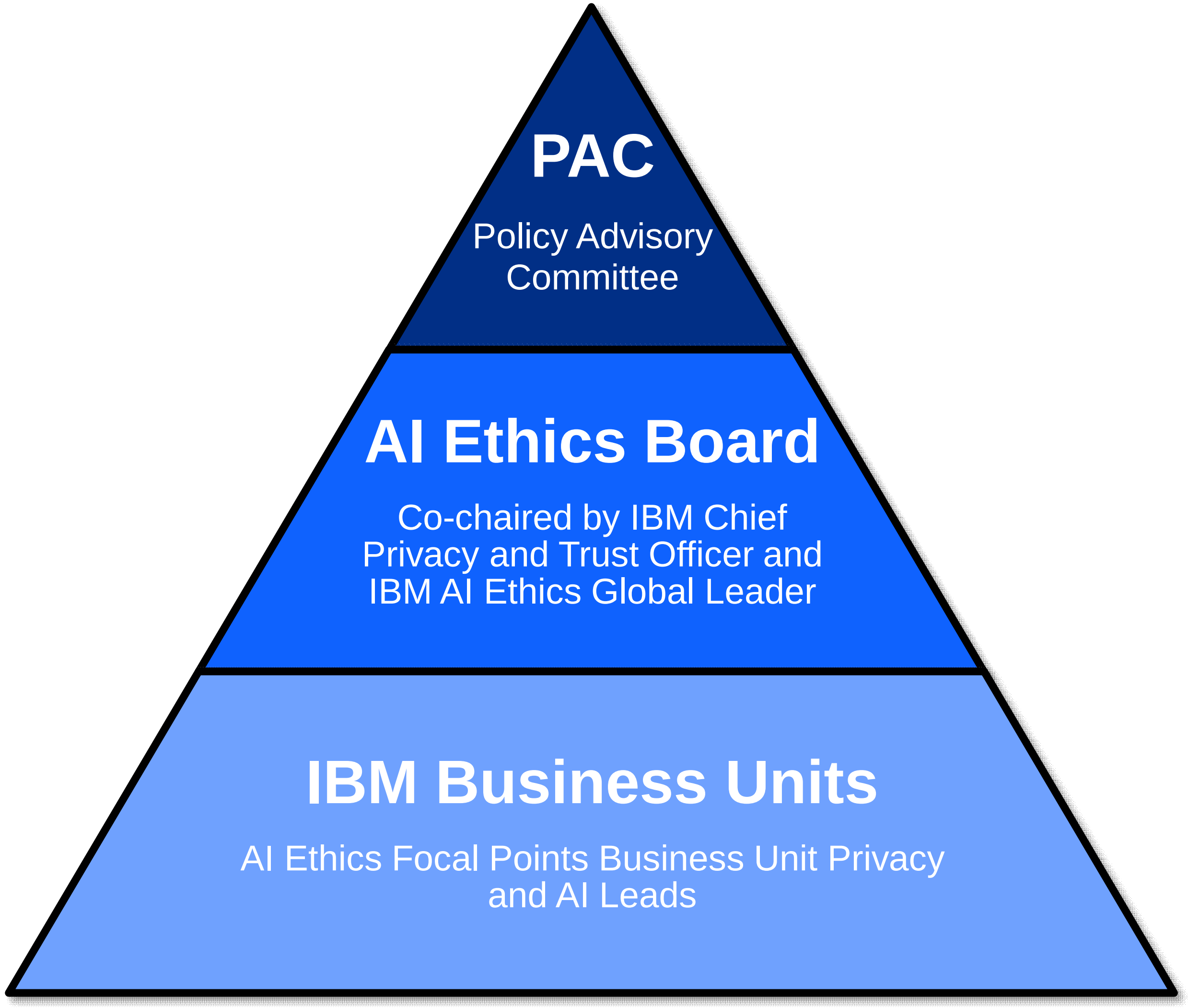
Trustworthy AI toolkits

Since 2018, IBM Research has developed and donated several trustworthy AI toolkits to the open source community so that anyone, anywhere in the world can use trusted tools to mitigate potential AI risks.

IBM’s Approach to AI Governance

The AI Ethics Board is at the center of IBM’s commitment to trust

- Comprised of a diverse set of stakeholders from across the company
- Steered by Policy Advisory Committee and supported by AI Ethics Focal Points and a strong Advocacy Network
- Instills a culture of trustworthy AI through a centralized governance, review, and decision-making process
- Deploys role-based educational programs to raise awareness and foster accountability across Business Units
- Embeds ethical principles into practices with an Ethics by Design approach



PAC

Policy Advisory Committee

AI Ethics Board

Co-chaired by IBM Chief Privacy and Trust Officer and IBM AI Ethics Global Leader

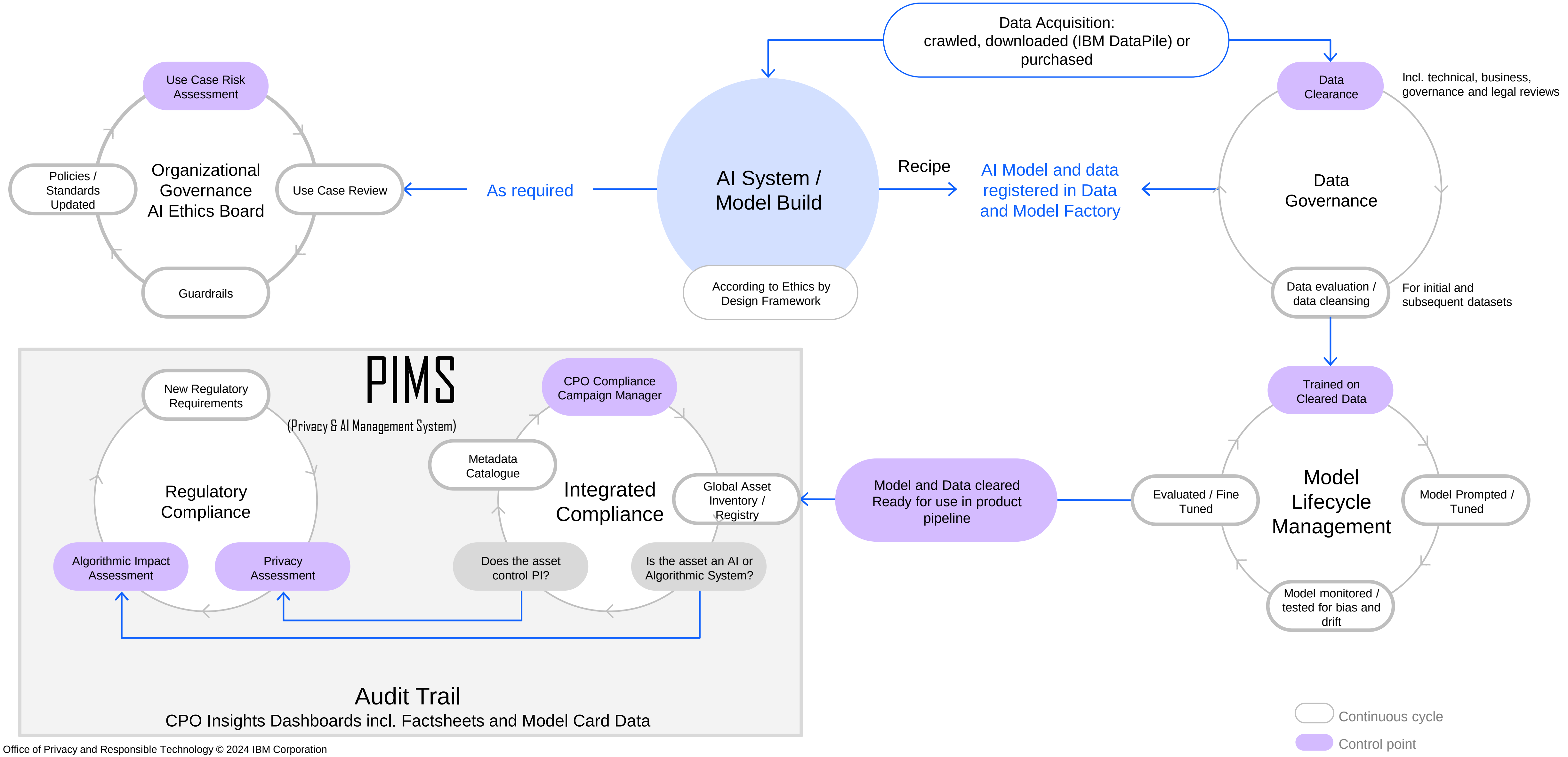
IBM Business Units

AI Ethics Focal Points Business Unit Privacy and AI Leads

CISO | Product teams | Advocacy Network | Enterprise Data Management | Government & Regulatory Affairs | Legal

Integrated Governance Program – Target Operating Model

watsonx.ai watsonx.data watsonx.governance



Organizational Governance AI Ethics Board

Use Case Risk Assessment

Use Case Review

Guardrails

Policies / Standards Updated

As required

AI System / Model Build

According to Ethics by Design Framework

Data Acquisition: crawled, downloaded (IBM DataPile) or purchased

Recipe

AI Model and data registered in Data and Model Factory

Data Governance

Data Clearance

Incl. technical, business, governance and legal reviews

Data evaluation / data cleansing

For initial and subsequent datasets

Model Lifecycle Management

Trained on Cleared Data

Model Prompted / Tuned

Model monitored / tested for bias and drift

Evaluated / Fine Tuned

Model and Data cleared Ready for use in product pipeline

PIMS

(Privacy & AI Management System)

Regulatory Compliance

New Regulatory Requirements

Privacy Assessment

Algorithmic Impact Assessment

Integrated Compliance

CPO Compliance Campaign Manager

Global Asset Inventory / Registry

Is the asset an AI or Algorithmic System?

Does the asset control PI?

Metadata Catalogue

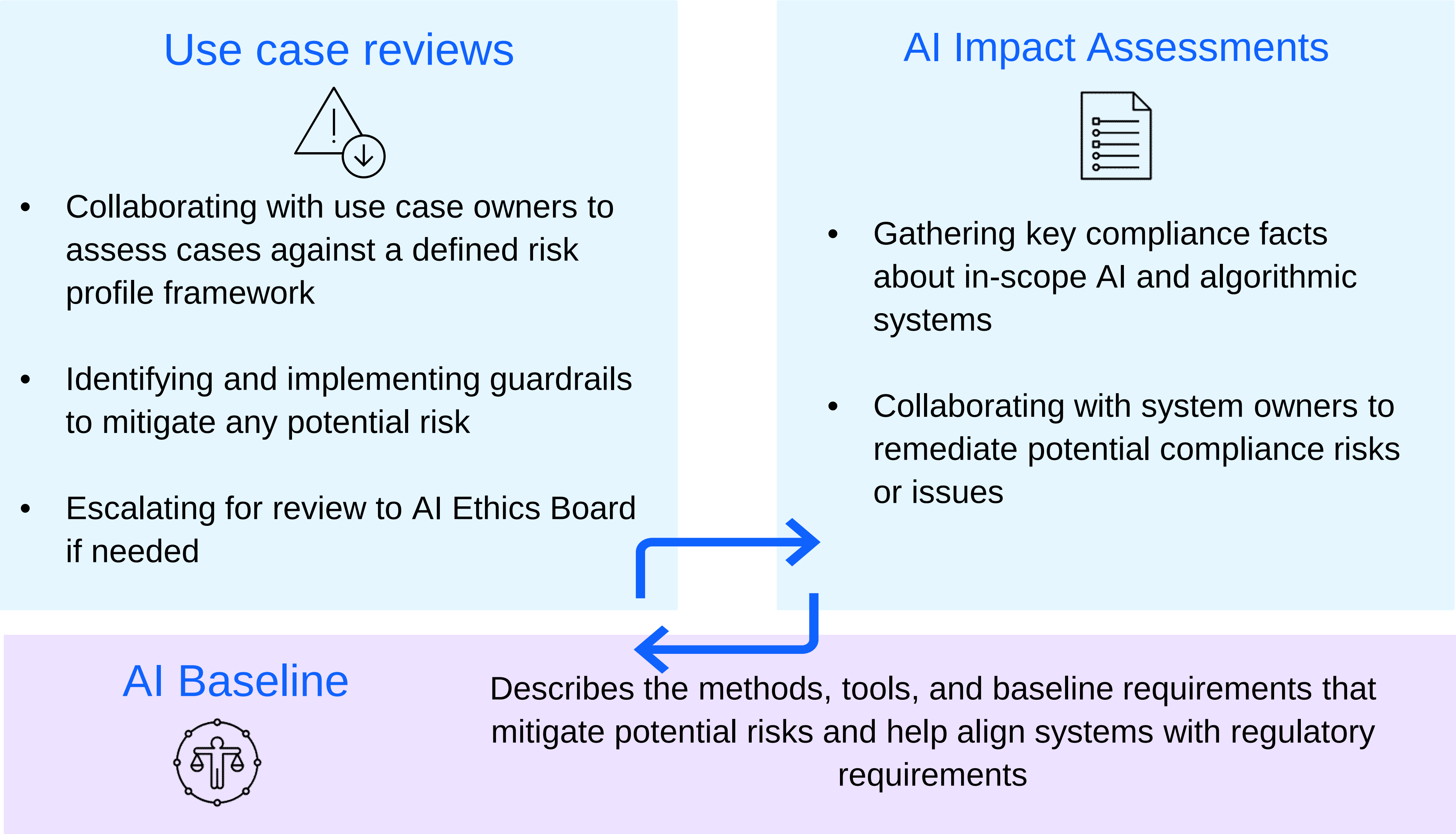
Audit Trail

CPO Insights Dashboards incl. Factsheets and Model Card Data

○ Continuous cycle

● Control point

Assessing use cases and cataloging systems



Use case reviews

- Collaborating with use case owners to assess cases against a defined risk profile framework
- Identifying and implementing guardrails to mitigate any potential risk
- Escalating for review to AI Ethics Board if needed

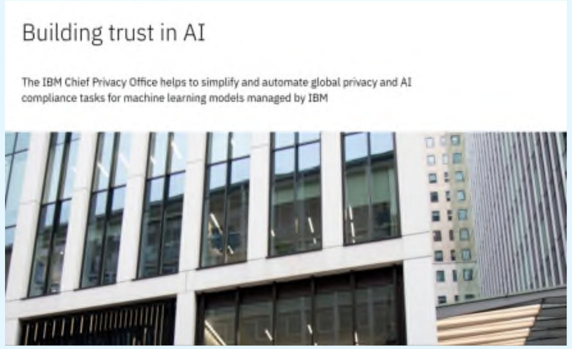
AI Impact Assessments

- Gathering key compliance facts about in-scope AI and algorithmic systems
- Collaborating with system owners to remediate potential compliance risks or issues

AI Baseline

Describes the methods, tools, and baseline requirements that mitigate potential risks and help align systems with regulatory requirements

Learn more



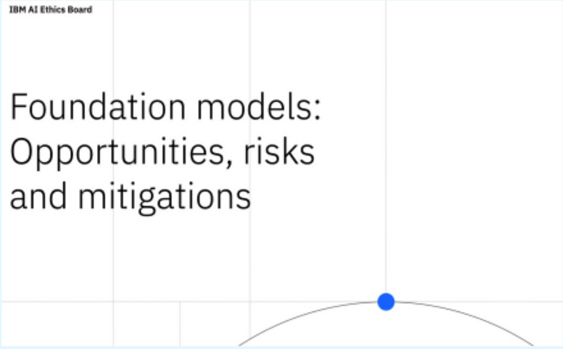
Building trust in AI

The IBM Chief Privacy Office helps to simplify and automate global privacy and AI compliance tasks for machine learning models managed by IBM

Learn more about how IBM's Office of Privacy and Responsible Technology simplified and automated global privacy and AI compliance tasks for machine learning models managed by IBM.
Read the case study →

A look into IBM's AI ethics governance framework

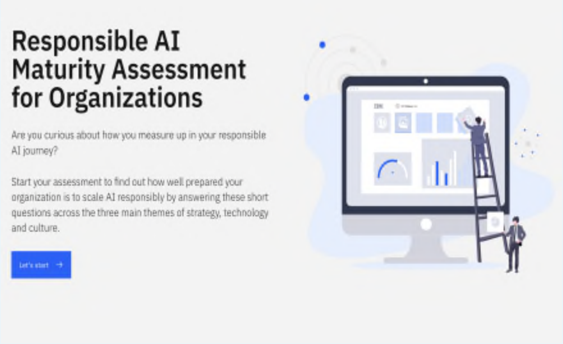
Read Gartner's insights on how to establish an AI governance framework, including a deep-dive into how IBM established its own governance framework.
Read the case study →



IBM AI Ethics Board

Foundation models: Opportunities, risks and mitigations

Read IBM's point of view on how generative AI capabilities can support unprecedented opportunities to benefit business and society alike when they are ethically designed and responsibly brought to market.
Read the blog →.



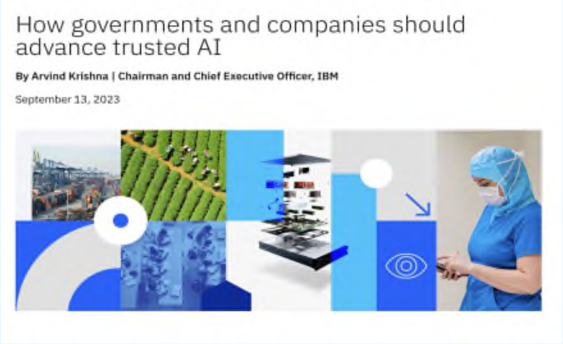
Responsible AI Maturity Assessment for Organizations

Are you curious about how you measure up in your responsible AI journey?

Start your assessment to find out how well prepared your organization is to scale AI responsibly by answering these short questions across the three main themes of strategy, technology and culture.

Find out how well prepared your organization is to scale AI responsibly across strategy, technology, and culture.

Take the assessment →



How governments and companies should advance trusted AI

By Arvind Krishna | Chairman and Chief Executive Officer, IBM

September 13, 2023

Learn more about IBM's position on AI regulation, based on three core tenets.

Read the POV →

Experts from IBM and the University of Notre Dame explore the tangible and intangible ROI of AI Ethics and Governance investments.

Read the article →

Other resources

IBM AI Ethics homepage →

IBM watsonx.governance →

IBM AI Academy video: How responsible AI can prepare you for regulations →

IBM AI Academy video: Trust, transparency, and governance in AI →

Recommendations to the Task Force

IBM Regulatory Point of View

Regulate AI Risk; Not AI Algorithms

Support an approach that regulates use of AI in *high-risk* applications; not all AI risk is the same

Hold creators & deployers responsible; not immune from liability

Legislation should not exempt from legal liability those who create and deploy AI

Support Open AI innovation; not a licensing regime

Risks creating a form of regulatory capture, benefitting incumbents

Would increase costs, hinder innovation, disadvantage smaller players and open-source developers

Opportunities to Improve SB 205

Simplify and clarify critical definitions

Improve "consequential decision" and "substantial factor"

"Consequential decision": We recommend aligning with the definition provided in a recent Data and Trust Alliance paper, which would replace the use of AI as a "substantial factor" in such a decision to a "controlling factor", which better aligns with the involvement of human oversight in AI decision-making. *

Reintroduce Open Exemption

The original open source AI exemption was deleted when the section on general purpose AI was eliminated; we recommend reintroducing that exemption

* For more details, see this Data and Trust Alliance paper on *Framing "Consequential Decisions":* https://dataandtrustalliance.org/work/framing-consequential-decisions