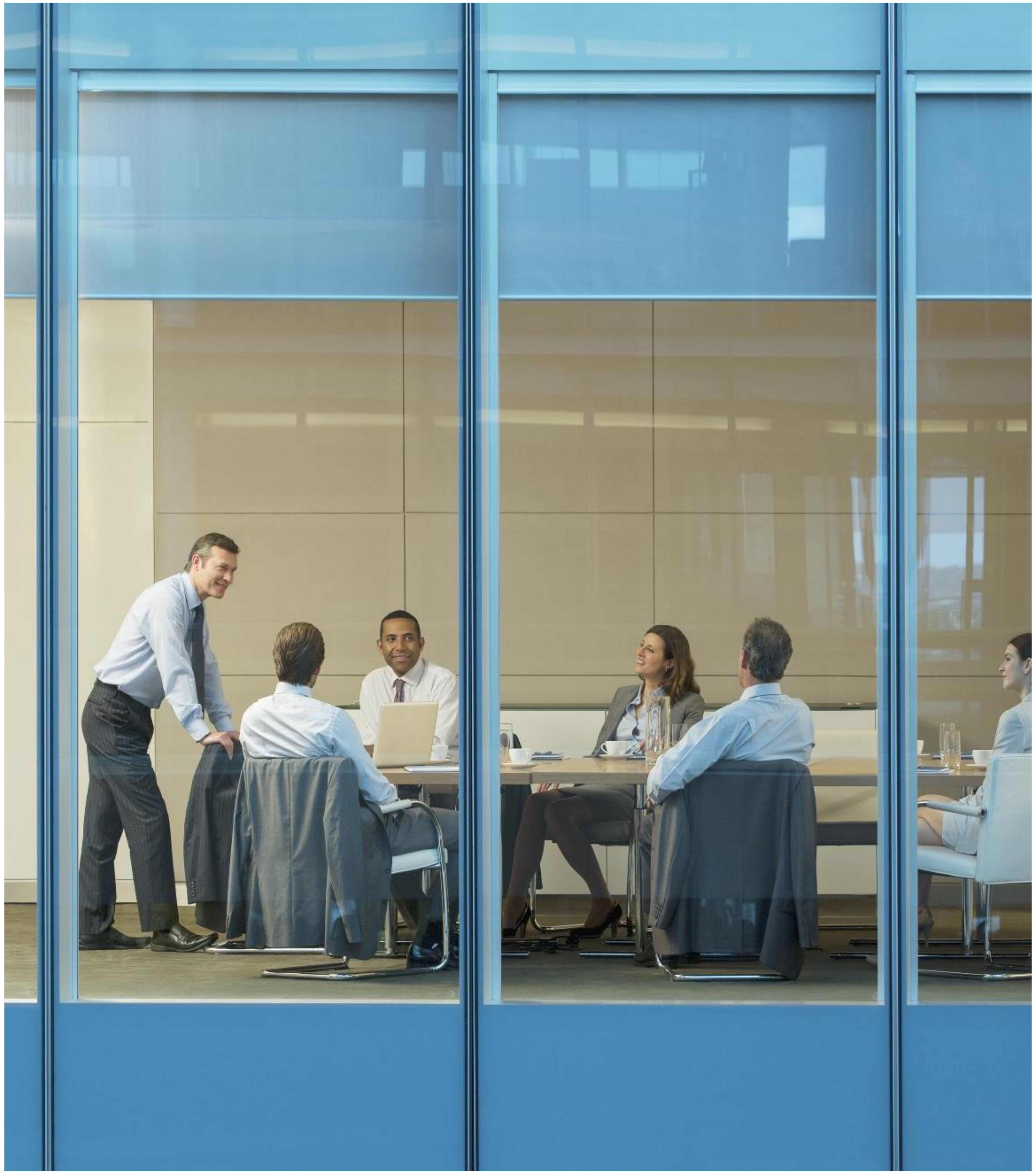
AI Ethics at IBM

—

A multidisciplinary,
multidimensional approach

Client presentation

Rachel Amity Brown
Tech Ethics Education Program Manager
Chief Privacy Office
reamity@us.ibm.com



AI is already used in many higher-stakes decision-making applications

Credit

Employment

Admissions

Healthcare

Enterprise workflows

Justice

It’s only by embedding ethical principles into AI applications and processes, that trustworthy systems can be built.

Why should organizations that build or use AI care about ethics?

Company values

Company reputation

Social justice and equity

Client and investor inquiries

Differentiation

Business opportunities

Existing or expected regulations

The first step toward trustworthiness is AI ethics.

AI ethics:
A multidisciplinary field that examines how to optimize AI's beneficial impact while reducing risks and adverse outcomes for all stakeholders in a way that prioritizes human agency and well-being, as well as environmental flourishing.

IBM has continuously strived for responsible innovation capable of bringing benefits to everyone and not just a few.

IBM applies the same philosophy to AI through its principles for trust and transparency.

The purpose of AI is to augment human intelligence

Data and insights belong to their creator

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

An AI system’s ability to treat individuals or groups equitably, depending 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.

IBM’s principles for trust and transparency are at work throughout the entire business.

IBM is at the forefront of global efforts to hold AI to high ethical standards.

IBM’s AI principles and pillars in practice

Governance

Ethics by design

Foundation models and generative AI

Methods and tools

Partnerships

IBM AI Ethics Board

The IBM AI Ethics Board is at the heart of the ethical decision-making the company applies to AI.

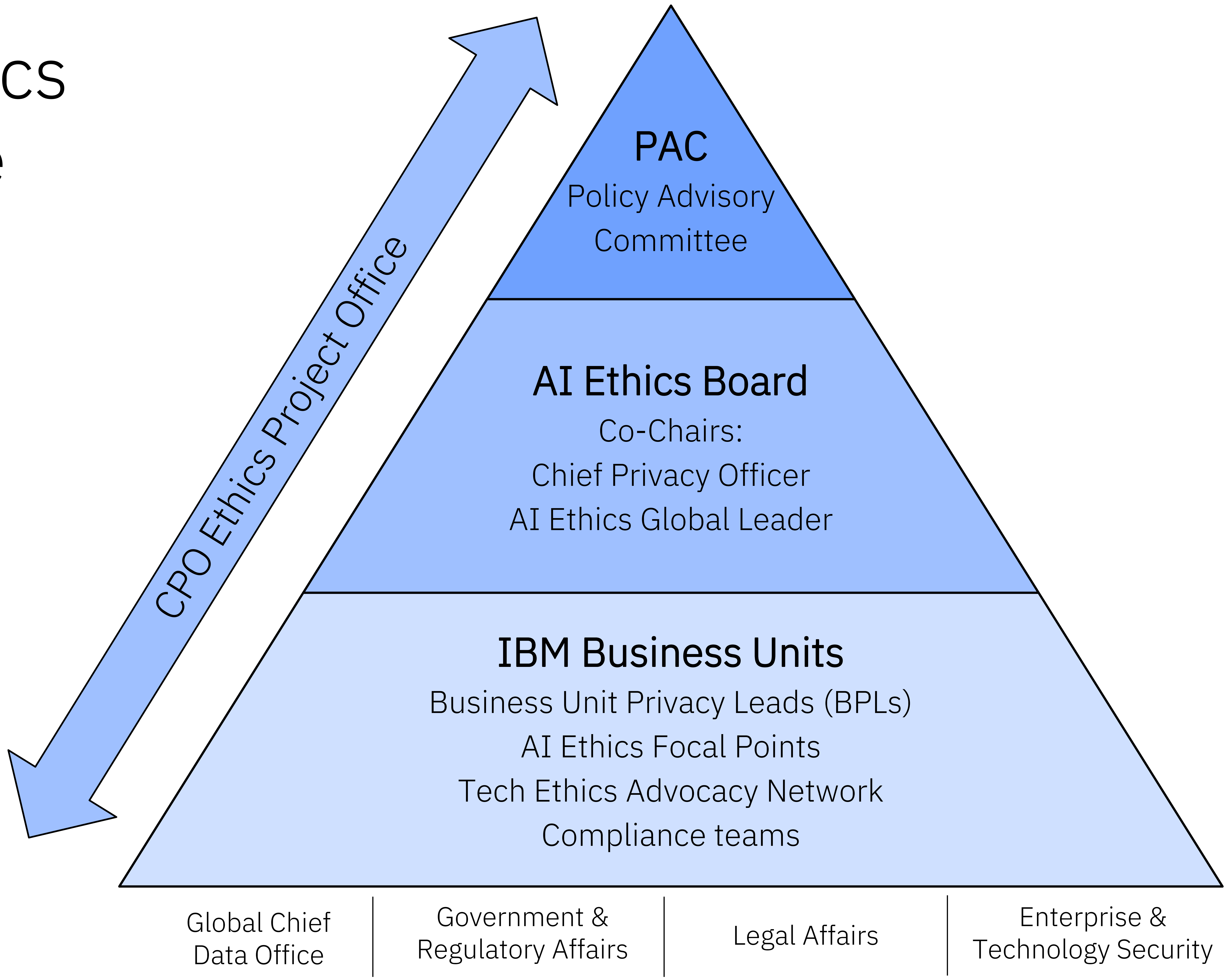
The board's mission is to support a centralized governance, review, and decision-making process for IBM ethics policies, practices, communications, research, products, and services.

IBM AI Ethics Board Co-Chairs

Francesca Rossi
IBM Fellow and
AI Ethics Global
Leader Research

Christina Montgomery
IBM Vice President,
Chief Privacy &
Trust Officer

IBM AI Ethics governance structure



CPO Ethics Project Office

PAC

Policy Advisory Committee

AI Ethics Board

Co-Chairs:

Chief Privacy Officer

AI Ethics Global Leader

IBM Business Units

Business Unit Privacy Leads (BPLs)

AI Ethics Focal Points

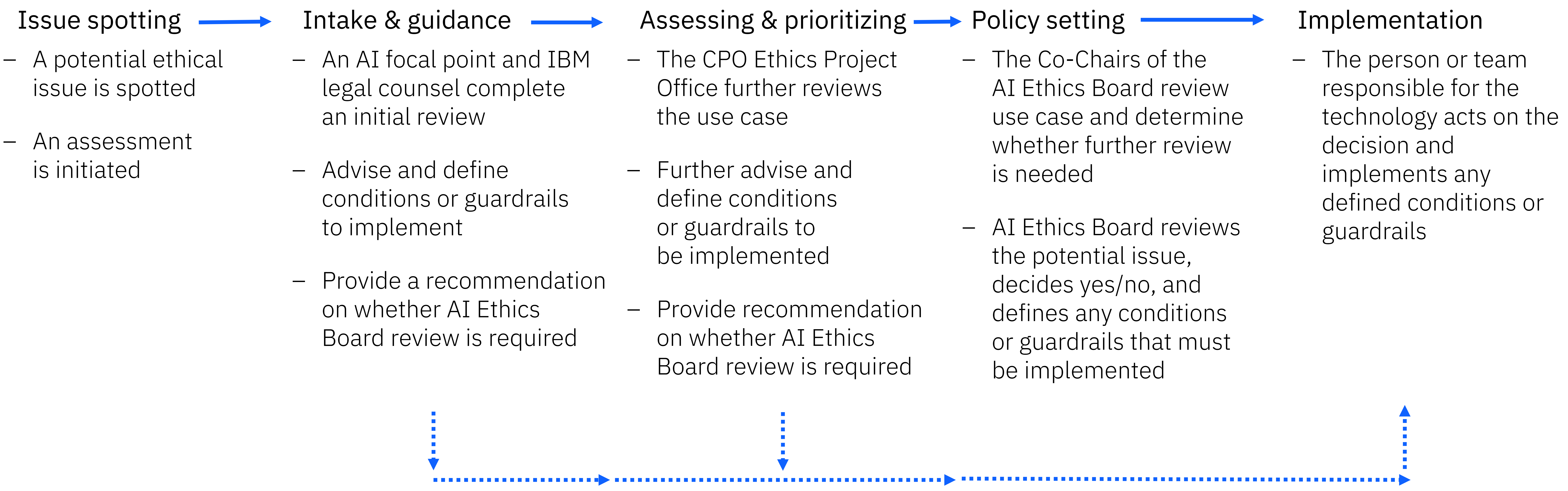
Tech Ethics Advocacy Network

Compliance teams

Global Chief Data Office | Government & Regulatory Affairs | Legal Affairs | Enterprise & Technology Security

Use case assessment process

IBM AI Ethics Board reviews use cases to ensure they are consistent with IBM’s principles and core values.



Issue spotting

- A potential ethical issue is spotted
- An assessment is initiated

Intake & guidance

- An AI focal point and IBM legal counsel complete an initial review
- Advise and define conditions or guardrails to implement
- Provide a recommendation on whether AI Ethics Board review is required

Assessing & prioritizing

- The CPO Ethics Project Office further reviews the use case
- Further advise and define conditions or guardrails to be implemented
- Provide recommendation on whether AI Ethics Board review is required

Policy setting

- The Co-Chairs of the AI Ethics Board review use case and determine whether further review is needed
- AI Ethics Board reviews the potential issue, decides yes/no, and defines any conditions or guardrails that must be implemented

Implementation

- The person or team responsible for the technology acts on the decision and implements any defined conditions or guardrails

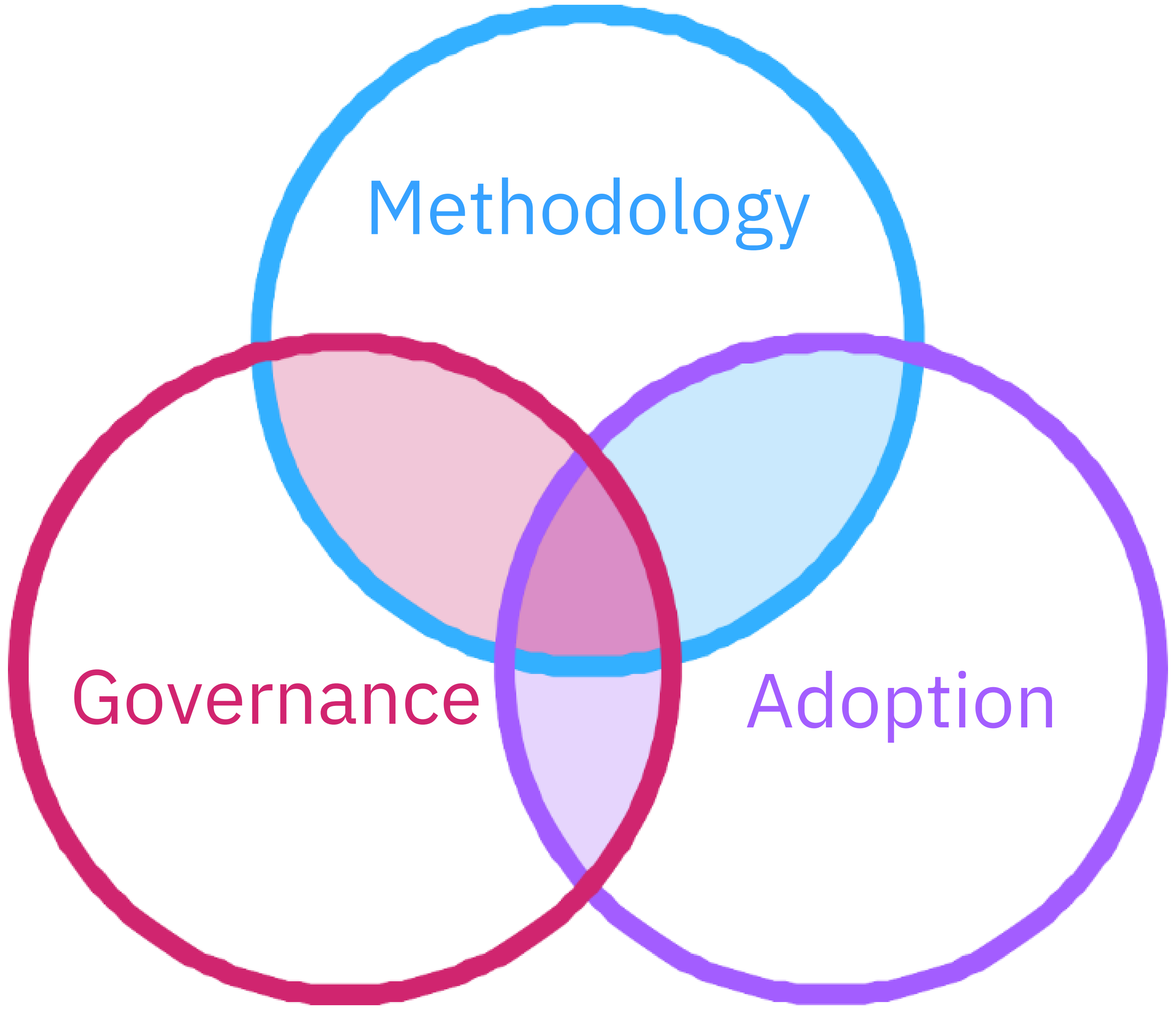
For issues deemed low-risk, or projects that align with previous board review guidance, an AI Ethics Board review is not required.

Ethics by design

Ethics by design is a structured framework with a goal of integrating tech ethics in the technology development pipeline, including but not limited to AI systems.

Its mission is to enable AI and other technology as a force for good by embedding technology ethics principles throughout IBM’s products and services, and in IBM’s broader operations across all business units and geographies.

Ethics by design focus areas



Methodology

Defining the recommended tools and best practices for practitioners to follow

Adoption

Putting the methodology into practice

Governance

Defining roles, responsibilities, and control points to promote and evaluate methodology adoption

Foundation models

What are they?

Foundation models are AI models that can be adapted to a wide range of downstream tasks.

They are typically large-scale (e.g., billions of parameters) generative AI models trained on unlabeled data using self-supervision.

What are their benefits?

- Performing complex tasks
- Increase in productivity
- Shorter time to value
- Diverse data modalities
- Amortized expenses

Foundation model risks

IBM's point of view on foundation model opportunities, risks, and mitigations outlines three categories of risk to help clarify potential risks and mitigation mechanisms.

	Input *Risks associated with the content provided to foundation models*	Output *Risks associated with the content generated by foundation models*	Other challenges *Risks associated with how foundation models are used*
Traditional *Risks known from earlier forms of AI*	Data laws Privacy Robustness	Fairness	Transparency
Amplified *Known risks intensified by foundation models*	Fairness Intellectual property Privacy Transparency	Explainability Misuse	Accountability Environment Human agency Human dignity Impact on jobs Legal uncertainty
New *Emerging risks intrinsic to the generative capabilities of foundation models*	Intellectual property Value alignment Privacy Robustness	Fairness Harmful code generation Intellectual property Misuse Privacy Traceability Value alignment	Diversity and inclusion Impact on education Intellectual property

IBM’s Principles for Trust and Transparency and Pillars of Trust support the responsible development, use, and governance of foundation models.

Foundation model guardrails and mitigation: Tools

watsonx

An enterprise-ready AI and data platform designed to multiply the impact of AI across your business. The platform comprises three powerful components:

- the watsonx.ai studio for new foundation models, generative AI and machine learning
- the watsonx.data fit-for-purpose store for the flexibility of a data lake and the performance of a data warehouse
- the watsonx.governance toolkit to enable AI workflows that are built with responsibility, transparency and explainability

Watson OpenScale

Tracks and measures outcomes from AI models through their lifecycle and helps organizations monitor fairness, explainability, resiliency, alignment with business outcomes, and compliance.

Trustworthy AI Toolkits

IBM-developed open-source toolkits that help you make AI more explainable, fair, robust, private, and transparent.

Foundation model guardrails and mitigation: Practices

Transparency reporting

Use standardized factsheet templates to accurately log details of the data and model, purpose, and potential use and harms.

Filtering undesirable data

Use curated, high-quality data.

Domain adaptation

Can help clients minimize the scope of risk the models can give rise to.

Human oversight and human-in-the-loop

Can help clients identify and correct errors and biases in the generated output and ensure that the generated content is accurate, relevant, of high quality, not drifting, and aligned.

Team diversity

Can help clients ensure that a variety of perspectives and experiences are considered.

IBM Enterprise Design Thinking

Can help clients define ethical behaviors throughout the AI design and development process.

Ethics review

Can help clients ensure the responsible development and use of the technology.

Ethics by Design

Can help clients enable AI and other technologies as a force for good by embedding tech ethics principles throughout products, services, and broader operations.

Consulting engagement

IBM Consulting is dedicated to help clients with the safe and responsible use of AI irrespective of the preferred tech stack.

IBM’s trustworthy AI tools

Tool kits

AI Explainability 360
Comprehensive open-source toolkit for explaining ML models & data.

AI Fairness 360
Comprehensive open-source toolkit for detecting & mitigating bias in ML models.

Adversarial Robustness 360
Comprehensive open-source toolkit for defending AI from attacks.

AI FactSheets 360
A research effort to foster trust in AI by increasing transparency and enabling governance.

AI Privacy 360
Toolbox to support the assessment of privacy risks of AI-based solutions, and to help them adhere to any relevant privacy requirements.

Uncertainty Quantification 360
Comprehensive open-source toolkit for computing and communicating meaningful limitations of ML predictions.

Product offerings

IBM Watson Studio
Empowers data scientists, developers, and analysts to build, run and manage AI models, and optimize decisions anywhere on IBM Cloud Pak for Data.

IBM Watson OpenScale
Tracks and measures outcomes from AI throughout its lifecycle, adapts and governs AI in changing business situations, and monitor AI models for bias, fairness, and trust.

IBM Cloud Pak for Data
A data and AI platform with a data fabric that makes data available for AI and analytics, on any cloud, and supports the creation of AI Factsheets.

IBM AI Governance
New, one-stop solution built on IBM Cloud Pak for Data that includes what’s needed to develop a model management process by capturing model development time, metadata, post-deployment model monitoring, and customized workflows.

Global leadership and collaboration

U.S. National AI Advisory Committee (NAIAC)

Chief Privacy Officer Christina Montgomery named to NAIAC and U.S. Chamber of Commerce Commission on Competition, Inclusion, and Innovation

Partnership on AI

Brings together diverse global voices to define best practices for beneficial AI; IBM is a founding member

World Economic Forum’s Global AI Action Alliance

Guides the responsible development of AI; co-chaired by Arvind Krishna, IBM Chairman and CEO

MIT-IBM Watson AI Lab

Research focused on healthcare, security and finance using the IBM Cloud, AI platform, blockchain and quantum

European Commission Expert Group on AI

Defined the ethics guidelines for trustworthy AI

IEEE Global Initiative on AI Ethics

Supports development of AI that prioritizes ethical considerations

ITU AI for Good Global Summit

Global and inclusive United Nations platform on using AI to achieve the UN Sustainable Development Goals

Data & Trust Alliance

Develops new practices and tools to advance the responsible use of data and AI across industries and disciplines

Policy

As a global company, IBM not only applies our principles throughout IBM, but also advocates for policies to promote AI ethics.

Policy

Precision Regulation for AI

"That is why today we are calling for precision regulation of AI. We support targeted policies that would increase the responsibilities for companies to develop and operate trustworthy AI. Given the ubiquity of AI [...] there will be no one-size-fits-all rules that can properly accommodate the many unique characteristics of every industry making use of this technology and its impact on individuals."

—

IBM Policy Lab
January 21, 2020

Rome Call for AI Ethics and Notre Dame – IBM Tech Ethics Lab

The Rome Call for AI Ethics seeks to create a shared sense of responsibility among businesses, governments, universities, and other organizations to create a future in which humankind is at the center of all technological development.

"IBM firmly opposes and will not condone uses of any technology, including facial recognition technology offered by other vendors, for mass surveillance, racial profiling, violations of basic human rights and freedoms, or any purpose which is not consistent with our values and Principles of Trust and Transparency."

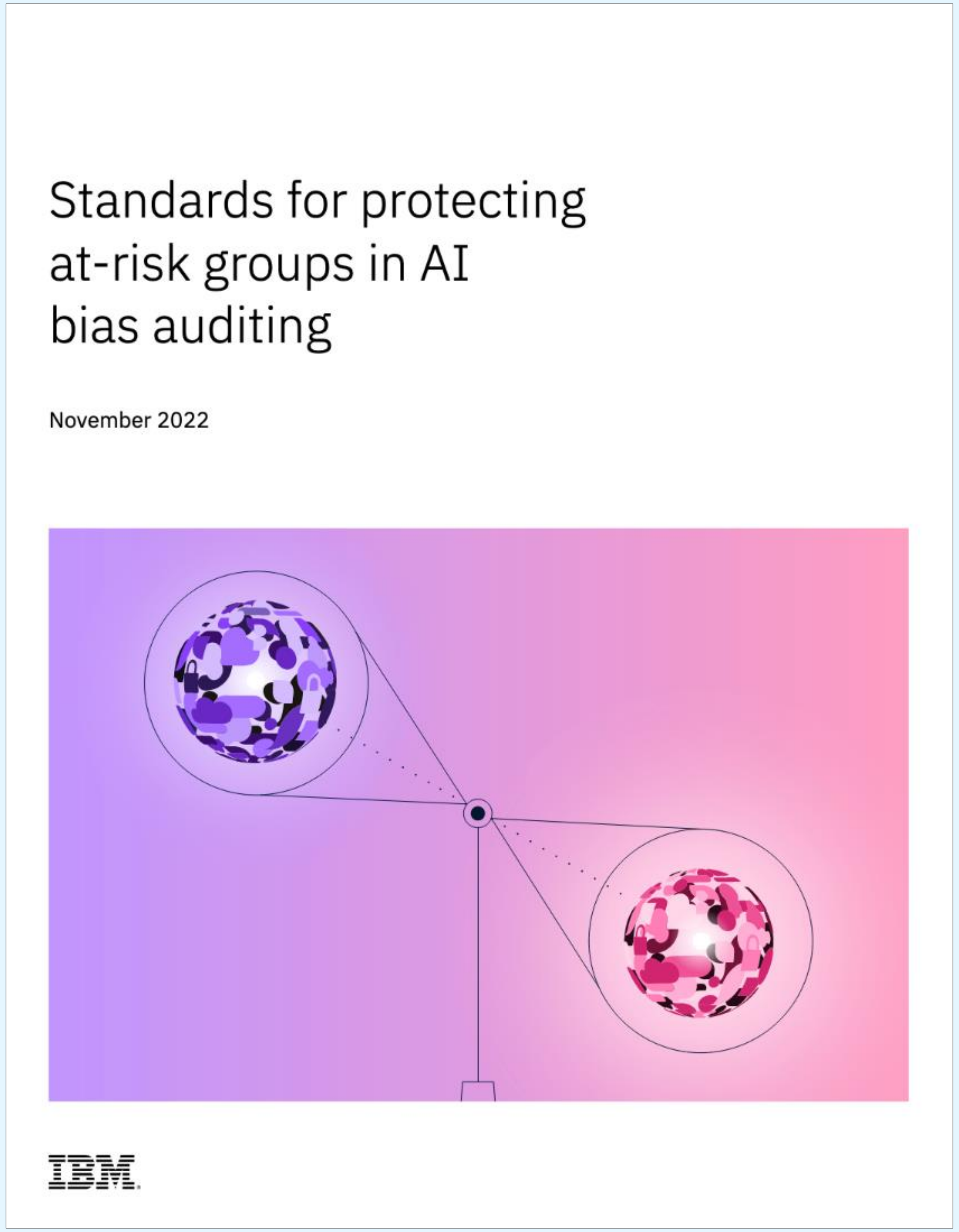
—

Arvind Krishna
IBM Chairman and CEO

From the letter to the US Congress
June 8, 2020

Standards for protecting at-risk groups in AI bias auditing

> “It should be standard practice in bias audit reporting to articulate the assumptions used for determining the relevant protected characteristics and associated classes used in the bias audit.”
>
> —
>
> *Standards for protecting at-risk groups in AI bias auditing*
> November 2022

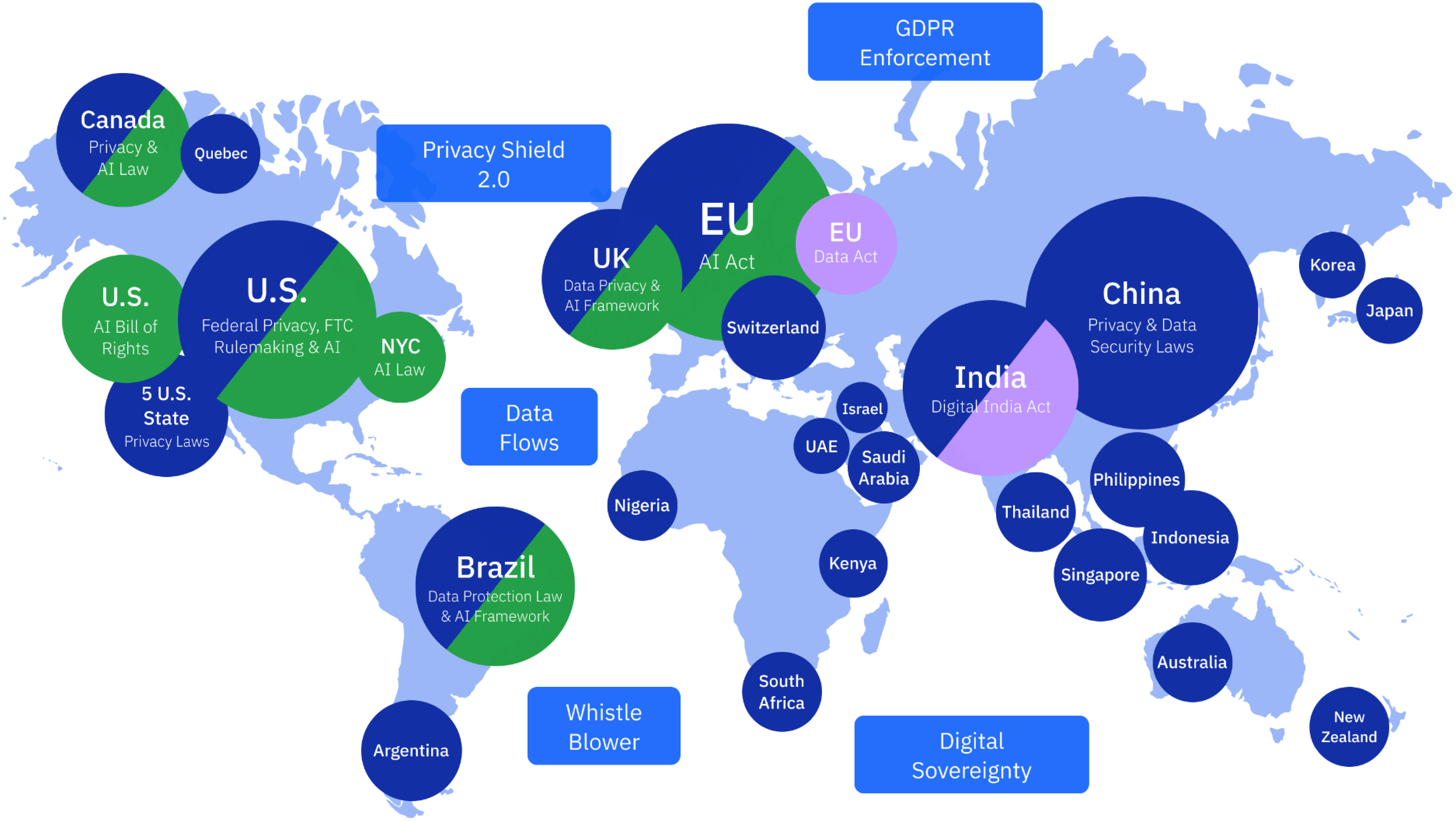


Standards for protecting at-risk groups in AI bias auditing

November 2022

IBM

Privacy and AI laws



● Privacy

● AI

● Data Governance

● Macro-environment

Two-thirds of countries across the world have some type of privacy laws and data governance regulations.

Extending IBM AI Ethics framework

IBM's AI Ethics principles, pillars, practices, and policies form a strong foundation for guiding the responsible development and use of other rapidly evolving tech, including: neurotechnology and quantum computing.

Neurotechnology

Potential issues around mental privacy, human agency, and identity

Quantum computing

Potential issues around the responsible use of such huge computer power and newly-possible capabilities

Key takeaways

Ethical considerations are at the heart of how IBM brings technology to the world.

As a values-based company, IBM is uniquely positioned to design and build AI systems that are underpinned by a strong ethical backbone – which is what clients need and want.

IBM’s commitments to data privacy and data governance are embedded into every AI system the company designs and builds.

Resources

IBM AI Ethics Webpage

IBM's Principles for Trust and Transparency

IBM's Pillars of Trust

Foundation models: Opportunities, risks, and mitigations
IBM's perspective on building and using foundation models in alignment with ethical expectations

A Policymaker's Guide to Foundation Models
IBM's guide to benefits and risks of foundation models, as well as recommendations for policymakers

Don't Pause AI Development; Prioritize Ethics Instead
Blog by IBM Chief Privacy Officer Christina Montgomery and IBM Global AI Ethics Leader Francesca Rossi about putting ethics at the forefront in the age of generative AI

How to Make AI More Ethical, Transparent, and Useful for Everyone
US Chamber of Commerce interview with IBM Chief Privacy Officer, Christina Montgomery

How our commitment to ethics, trust and transparency is differentiating IBM
An overview of some of the ways IBM is working to have a lasting, positive ethical impact

Thank you

This document is current as of the initial date of publication and may be changed by IBM at any time.

Statements regarding IBM’s future direction and intent are subject to change or withdrawal without notice and represent goals and objectives only.

THIS DOCUMENT IS DISTRIBUTED “AS IS” WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IN NO EVENT, SHALL IBM BE LIABLE FOR ANY DAMAGE ARISING FROM THE USE OF THIS INFORMATION, INCLUDING BUT NOT LIMITED TO, LOSS OF DATA, BUSINESS INTERRUPTION, LOSS OF PROFIT OR LOSS OF OPPORTUNITY.

Client examples are presented as illustrations of how those clients have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

Not all offerings are available in every country in which IBM operates.

It is the user’s responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

IBM