

Frankfurt, 19 September 2023

# Redefining the Horizon: Integrating AI Ethics and Literacy into the EU Data Strategy Framework

**Alessandra Bini**

Senior Counsel, Trust and Compliance Officer IBM  
Europe

Member of the Board of AIGI and AIGI Rep in  
Lombardia and Liguria



# Agenda

**Part I: Understanding the landscape for trustworthy AI systems**

**Part II: AI Ethics**

**Part III: AI Literacy**

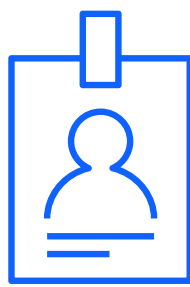
## Part I:

Understanding the  
landscape for trustworthy  
AY Systems

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



Credit



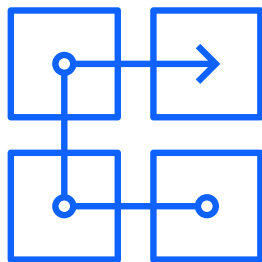
Employment



Admissions



Healthcare



Enterprise  
workflows



Justice

By embedding ethical principles into AI applications and processes, we can build systems that are trustworthy.

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

---

Company values

---

Company reputation

---

Social justice and equity

---

Client & investor inquiries

---

Differentiation

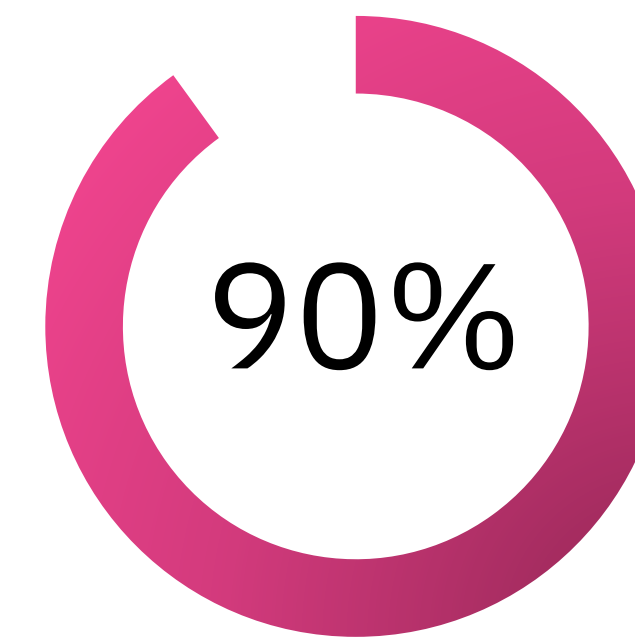
---

Business opportunities

---

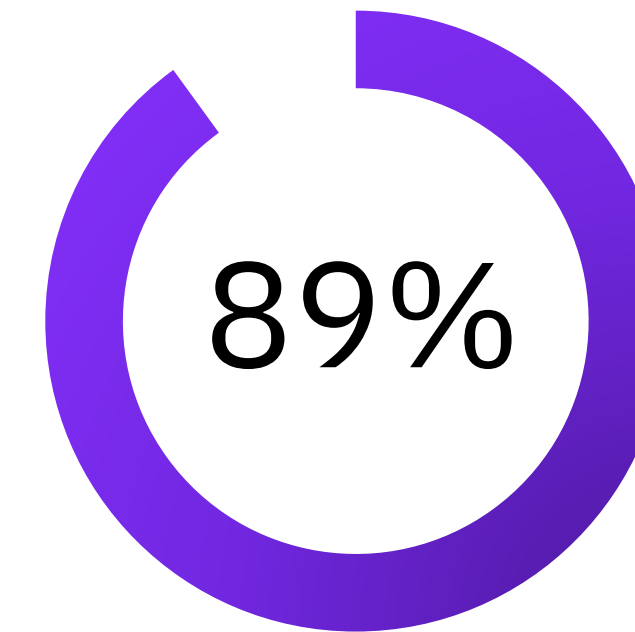
Existing or expected regulations

Brand  
reputation



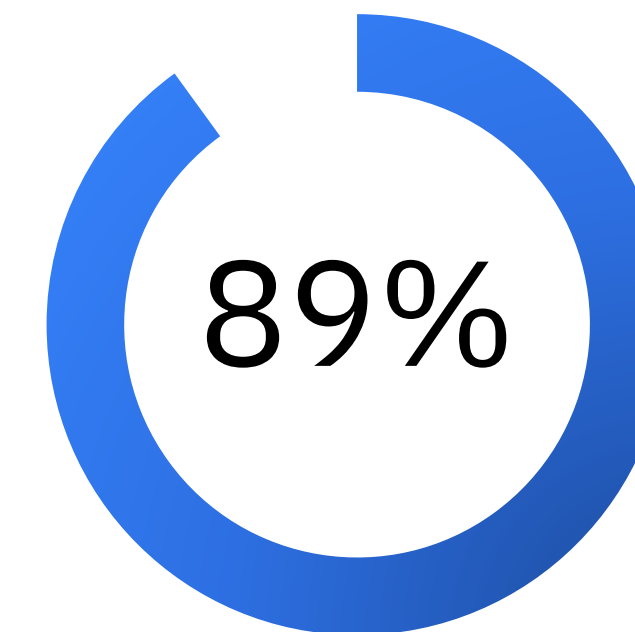
of companies using AI say  
maintaining brand integrity  
and customer trust is a top  
priority

Increased  
regulation



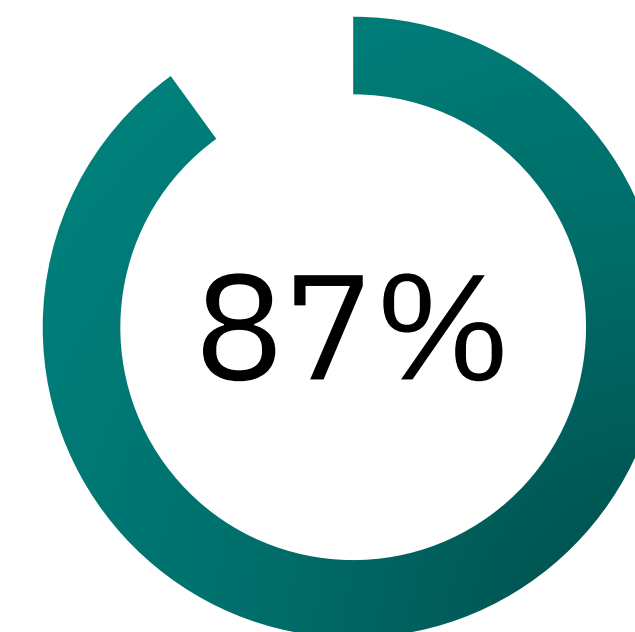
of companies using AI say  
meeting external  
regulatory and compliance  
obligations is a top priority

Complexity of AI  
deployments



of companies using AI say  
ability to monitor data and  
AI across the *entire*  
lifecycle is a top priority

Social  
justice



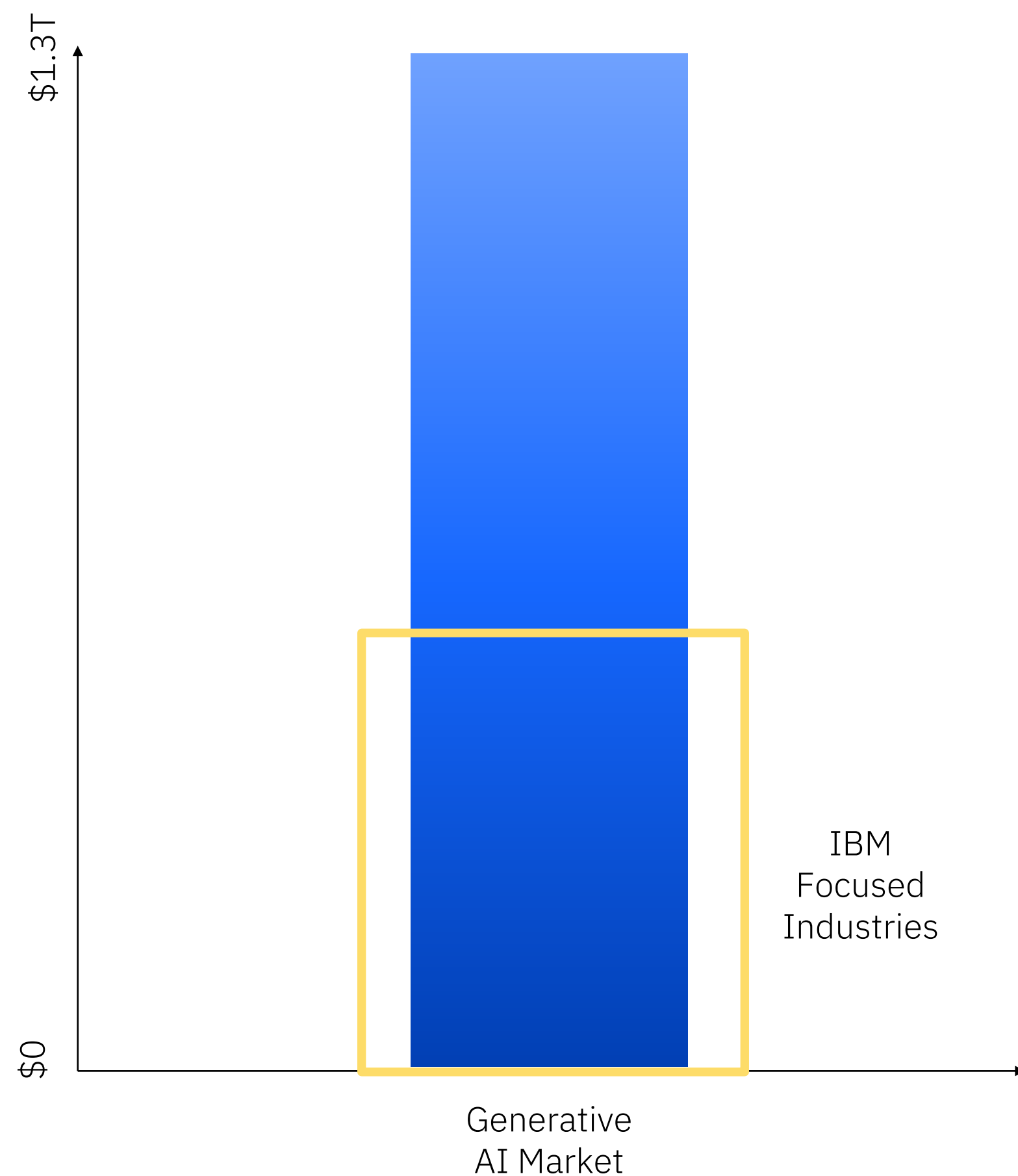
of companies using AI say  
ensuring applications and  
services minimize bias is a  
top priority



# Generative AI and Foundation Models

## - Opportunities

+AI → AI+



- ✓ Flexible & Scalable to accelerate with AI adoption
- ✓ It impacts how people interact with technology
- ✓ Business is placing it at their strategic core



# Generative AI and Foundation Models - Risks

## 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 <i>Risks associated with the content provided to foundation models</i>	Output <i>Risks associated with the content generated by foundation models</i>	Other challenges <i>Risks associated with how foundation models are used</i>
<b>Traditional</b> <i>Risks known from earlier forms of AI</i>	Data laws Privacy Robustness	Fairness	Transparency
<b>Amplified</b> <i>Known risks intensified by foundation models</i>	Fairness Intellectual property Privacy Transparency	Explainability Misuse	Accountability Environment Human agency Human dignity Impact on jobs Legal uncertainty
<b>New</b> <i>Emerging risks intrinsic to the generative capabilities of foundation models</i>	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

# Risks are not just theoretical

## Artists Are Suing Artificial Intelligence Companies and the Lawsuit Could Upend Legal Precedents Around Art

By [Shanti Escalante-De Mattei](#)  May 5, 2023 10:37am

## Thomson Reuters, ROSS Intelligence duel over fair use of legal headnotes for training AI

2023 IPDBRF 0029 • By Patrick H.J. Hughes

WESTLAW Intellectual Property Daily Briefing • February 28, 2023



## ChatGPT maker OpenAI faces a lawsuit over how it used people’s data

A California law firm says the company’s use of scraped data from the web violates the rights of millions of internet users



By [Gerrit De Vynck](#)

Updated June 28, 2023 at 3:01 p.m. EDT | Published June 28, 2023 at 1:01 p.m. EDT

Google accused of 'stealing everything ever created and shared on the internet' to create Bard

12-07-2023 Marisa Woutersen

## FTC investigates OpenAI over data leak and ChatGPT’s inaccuracy

Litigation | Copyright | Intellectual Property | Litigation | Class Actions & Multi-District Litigation

## OpenAI, Microsoft want court to toss lawsuit accusing them of abusing open-source code

By [Blake Brittain](#)

January 27, 2023 10:59 PM GMT+1 · Updated 7 months ago

Technology

## Getty asks London court to stop UK sales of Stability AI system

By [Sam Tobin](#)

June 1, 2023 8:14 PM GMT+2 · Updated 3 months ago

LEADERSHIP INSIGHTS

## In generative AI legal Wild West, the courtroom battles are just getting started

PUBLISHED MON, APR 3 2023•10:56 AM EDT | UPDATED MON, APR 3 2023•11:29 AM EDT



[Ellen Sheng](#)  
[@ELLENSHENG](#)

SHARE [f](#) [t](#) [in](#) [✉](#)

NEWSLETTERS • DATA SHEET

## ‘Prone to hallucinations and bias’: A Texas judge puts A.I. in its place



# Privacy and AI Laws

Two thirds of countries across the world have some form of Privacy law, and enforcement continues to increase

AI regulations are already here



## Privacy Laws and Regulations

GDPR includes fines of up to €20million or 4% of global annual revenue

The Personal Information Protection Law (PIPL) includes fines of up to approximately 7.7 million or 5% of annual revenue

## AI Laws and Regulations

NYC Council AI Law (Effective 2023)

The proposed EU AI Regulation includes fines of up to 6% of a company’s annual revenues for noncompliance

Proposed AI regulation by the Cyberspace Administration of China (CAC)

Sources  
[China - Data Protection Overview](#)  
[New AI Regulations Are Coming. Is Your Organization Ready?](#)



# The Global AI Regulatory Landscape



## United States

The US has no existing national AI regulations

Various state and local laws apply to specific uses of AI:

- New York City (Employment Decisions)
- Maryland (Facial Recognition in Interview Process)
- Colorado (Insurance Discrimination)

Existing consumer protection and employment laws are applied regardless of whether AI is used in decision-making



## European Union

The EU has established resolutions around liability for AI, but currently has no comprehensive laws that apply to AI systems

Proposed AI Act (Draft AI Act) is first of its kind that aims to curb potential harms by regulating entire sector:

- Risk based approach
- Updated 2022 Draft would include "general purpose" AI systems and foundation models
- Currently in legislative process ("Trialogue")



## China

China has the most comprehensive rules for use of AI systems including both national and provincial regulations and draft measures for generative AI

National regulations include provisions:

- Requiring companies to provide consumers notice of any automated decision making and the ability to opt out
- Mandating that AI services be moral, ethical, accountable, transparent, and "spread positive energy" and embody "core socialist values"
- Prohibiting the use of algorithms leveraging personal data to offer different prices to consumers.
- Requiring platform providers and end users to include watermarks for AI generated images, texts, audio and videos.



# Proposed AI ACT

## Principles

Ensure that AI systems are safe and respect existing law on fundamental rights and the EU values

Ensure legal certainty to facilitate investment and innovation in AI

Enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems

Facilitate the development of a single market for safe and trustworthy AI and preventing fragmentation

## Approach

AI definition is technologically neutral

Risk based approach

- Unacceptable – prohibited use (Title II)
- High risk – permitted on the European market subject to compliance with certain mandatory requirements and an ex-ante conformity assessment (Title III)
- Non high risk – subject to certain transparency requirement to enable users to take informed decisions

## Governance

Governance systems at Union and national level

Establishment of an EU-wide database for stand-alone high-risk AI systems operated by the Commission

Registration obligations for providers of high-risk AI systems

Monitoring and reporting obligations for providers of AI systems with regard to post-market

Monitoring and reporting and investigating on AI-related incidents and malfunctioning.

# Part II: AI Ethics

By embedding ethical principles into AI applications and processes, we can build systems that are trustworthy.

The first step  
toward  
trustworthiness  
is AI ethics.

AI ethics:

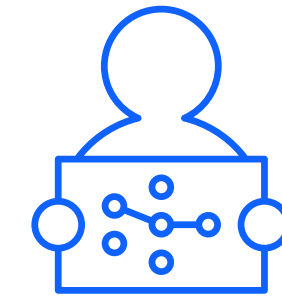
*A multidisciplinary field that examines how to **optimize AI's beneficial impact while reducing risks and adverse outcomes** for all stakeholders in a way that prioritizes human agency and well-being, as well as environmental flourishing.*



## Principles for Trust and Transparency

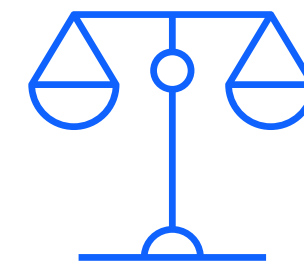
- 1 The purpose of AI is to augment — not replace — human intelligence
- 2 Data and insights belong to their creator
- 3 New technology, including AI systems, must be transparent and explainable

## Pillars of Trust



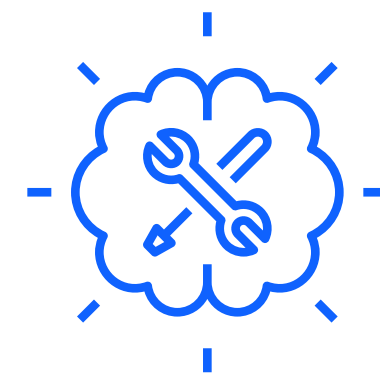
### Explainability

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



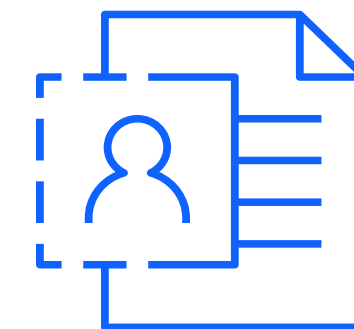
### Fairness

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



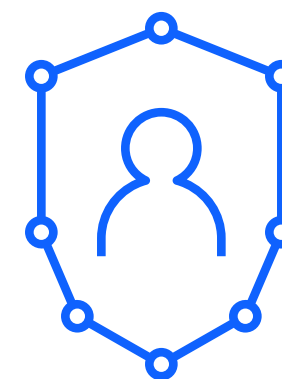
### Robustness

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

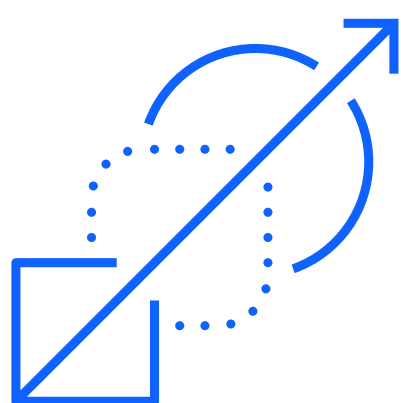
# Our principles and pillars in practice



Governance



Ethics by Design



Foundation models  
and generative AI

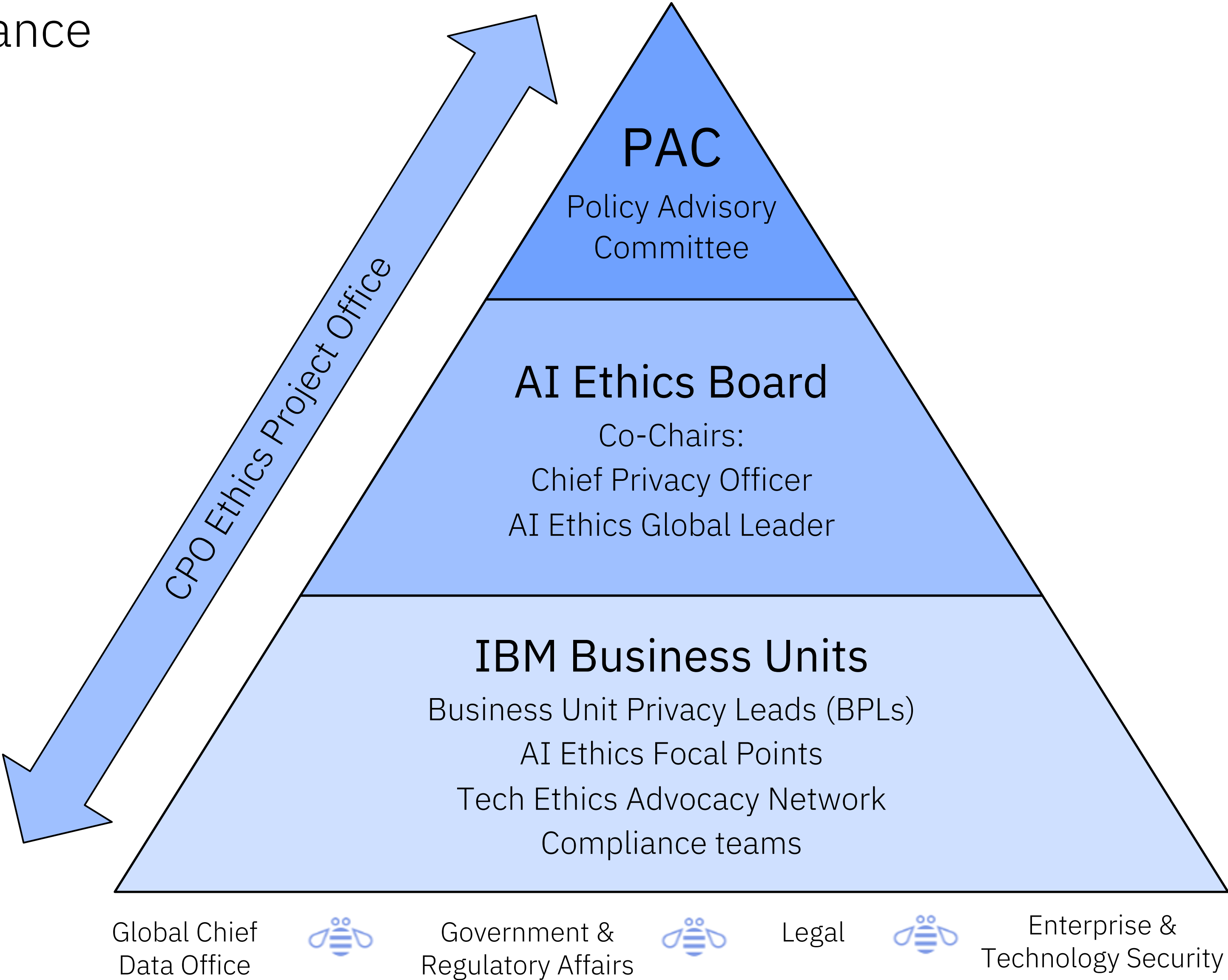


Methods and tools



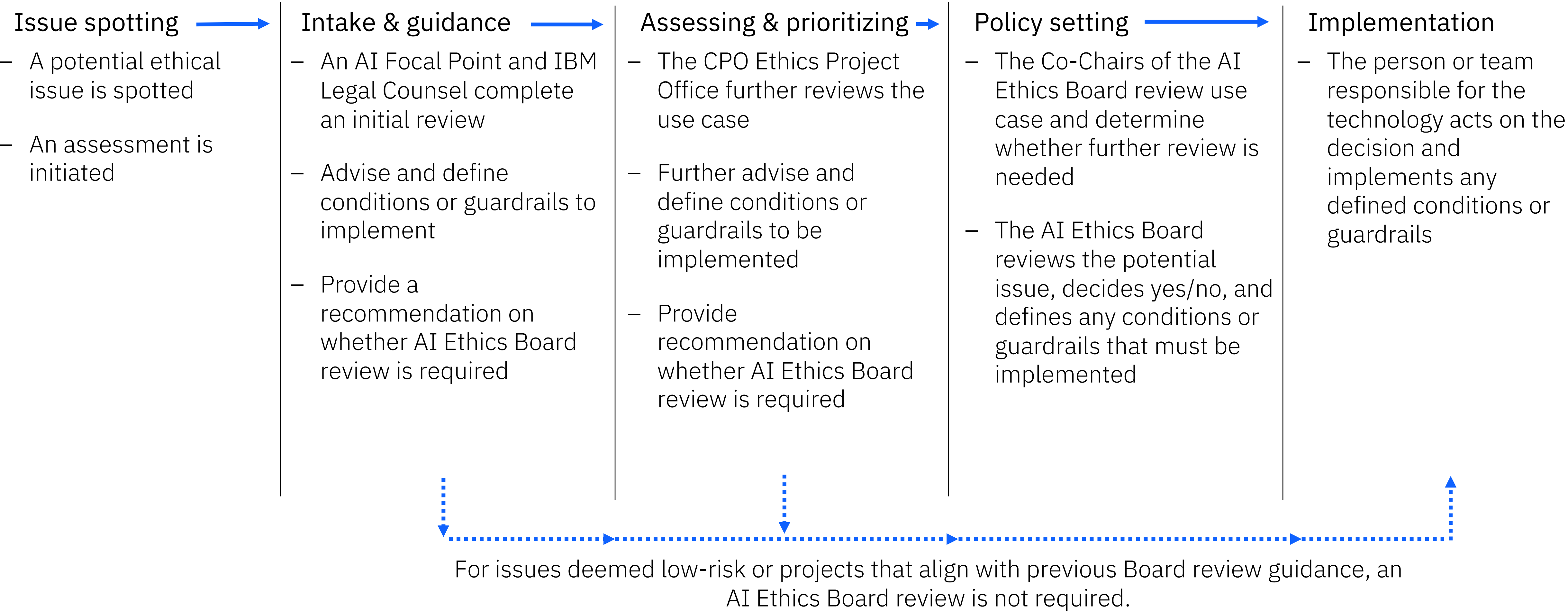
Partnerships

# AI Ethics governance structure



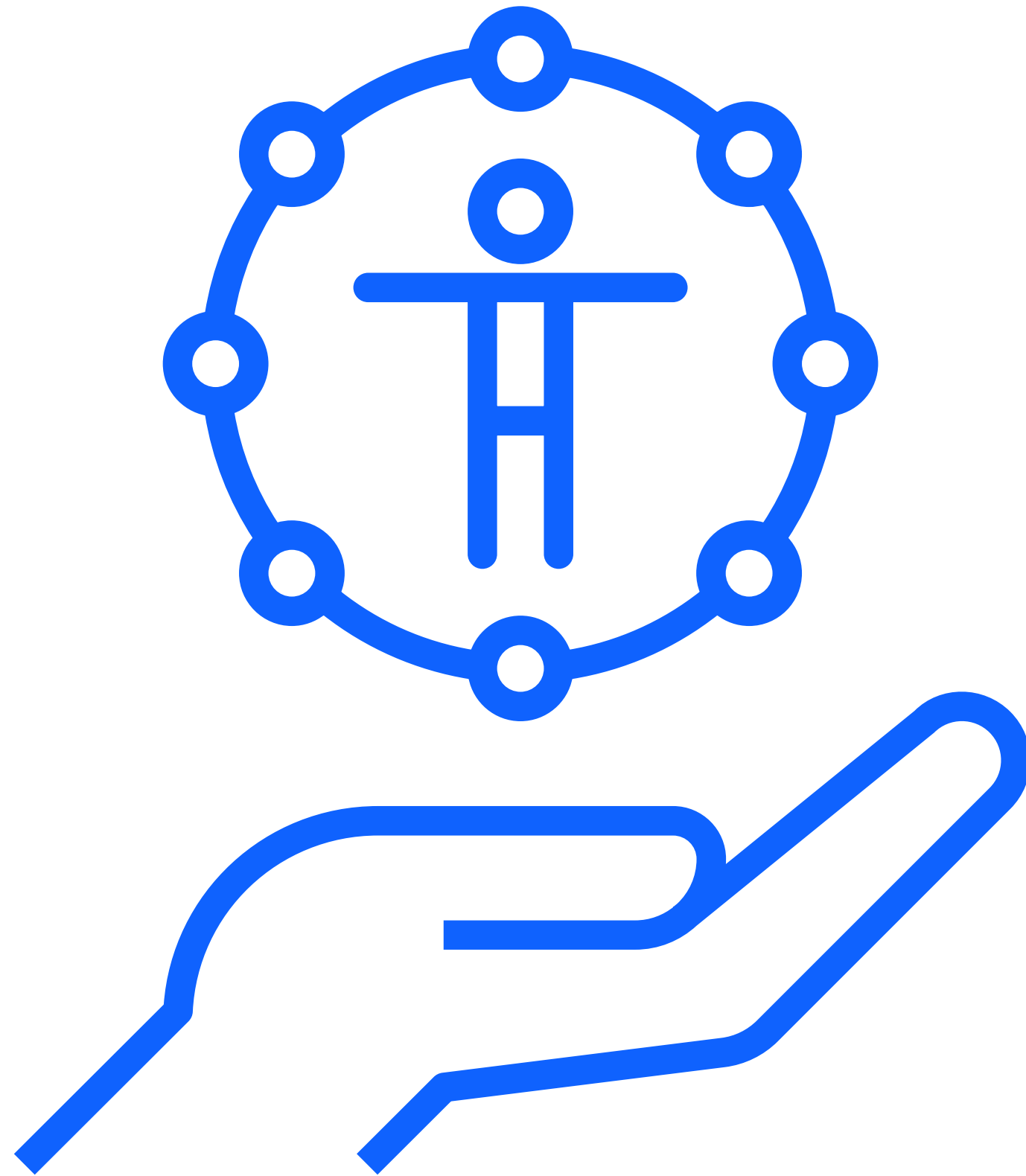
# Use case assessment process

The AI Ethics Board reviews use cases and other proposals to see they are consistent with IBM’s principles and core values.





# 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 technologies as a [force for good](#) by embedding tech ethics principles throughout IBM's products & services, and in IBM's broader operations across all business units and geographies.

# IBM's trustworthy AI tools

## TOOLKITS

### [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.



# Global leadership & 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

## Notre Dame-IBM Tech Ethics Lab

Produces and supports tangible, applied, and interdisciplinary research projects that address core tech ethics questions

## 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

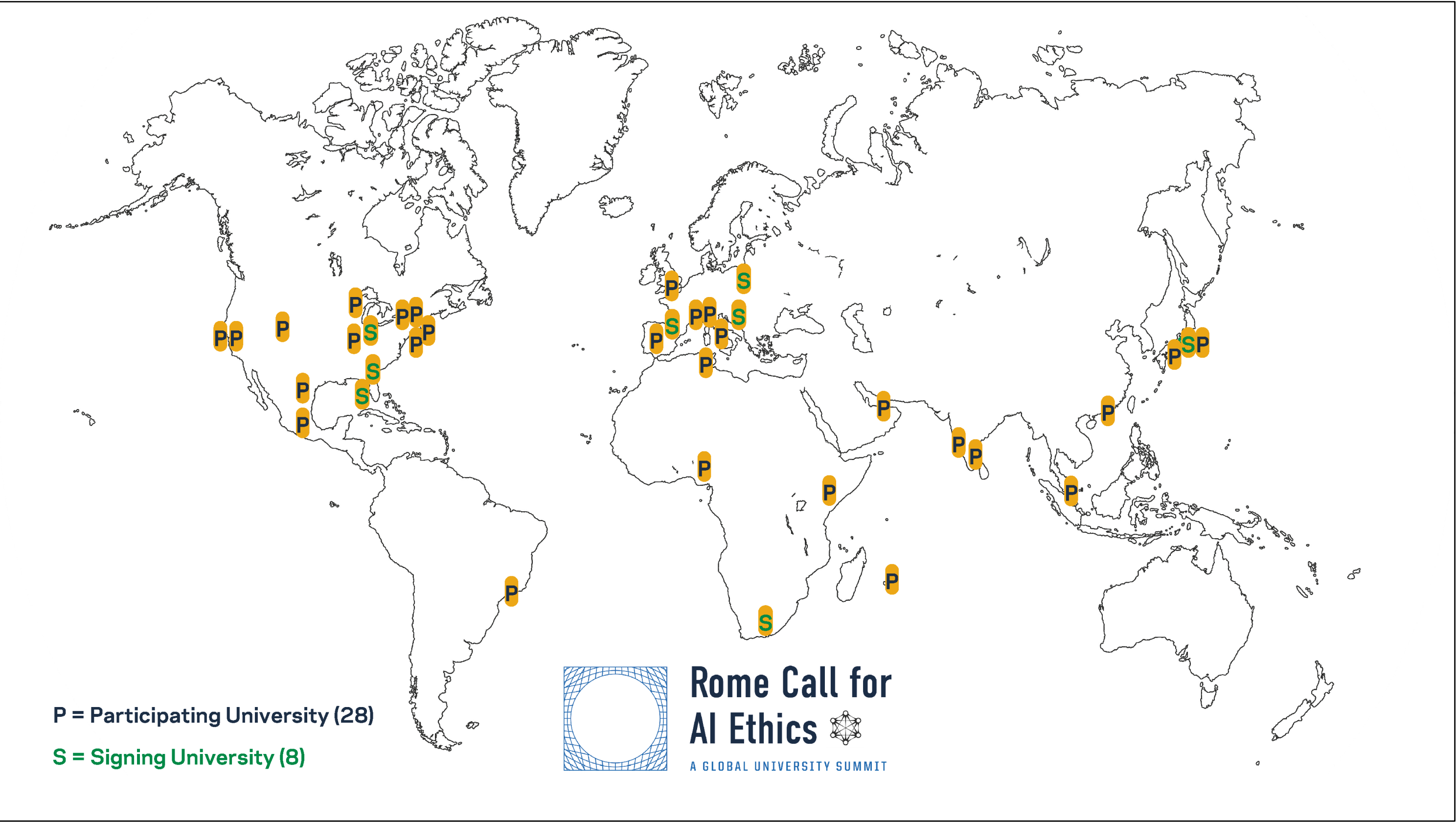
Global and inclusive United Nations platform on using AI to achieve the UN Sustainable Development Goals

## Data & Trust Alliance

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



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



2020 | IBM is one of only two companies invited by the Vatican to be the first signatories to the [Rome Call for AI Ethics](#), which advocates for a human-centered approach to AI

IBM and the University of Notre Dame co-found the [Notre Dame-IBM Technology Ethics Lab](#) to guide the development of new and emerging technologies including artificial intelligence

2022 | 36 universities attend the Rome Call for AI Ethics [Global University Summit](#) hosted by the Notre Dame–IBM Tech Ethics Lab

8 new universities sign the Rome Call at the Summit



# Part III: AI Literacy

Paving the way to AI systems  
that are **trustworthy**, passes  
through the concept of **AI**  
**literacy**



## EU 2030 Digital Compass

Skills Objectives:  
ICT Specialists: 20 million + gender convergence  
Basic Digital Skills: min 80% of population



## OECD Going Digital Toolkit

Skills Objectives:  
“we must all work together strengthen trust in digital environments, for example by raising awareness and empowering people and organisations to better manage digital risk”

Advocacy  
Global Leadership  
& Collaboration

Key Collaborations  
Impactful Events  
Thought & Policy leadership  
Contributions to spreading AI literacy (internally and externally)

Thank you.

© 2023 International Business Machines Corporation

IBM and the IBM logo are trademarks of IBM Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on [ibm.com/trademark](https://ibm.com/trademark).

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.

