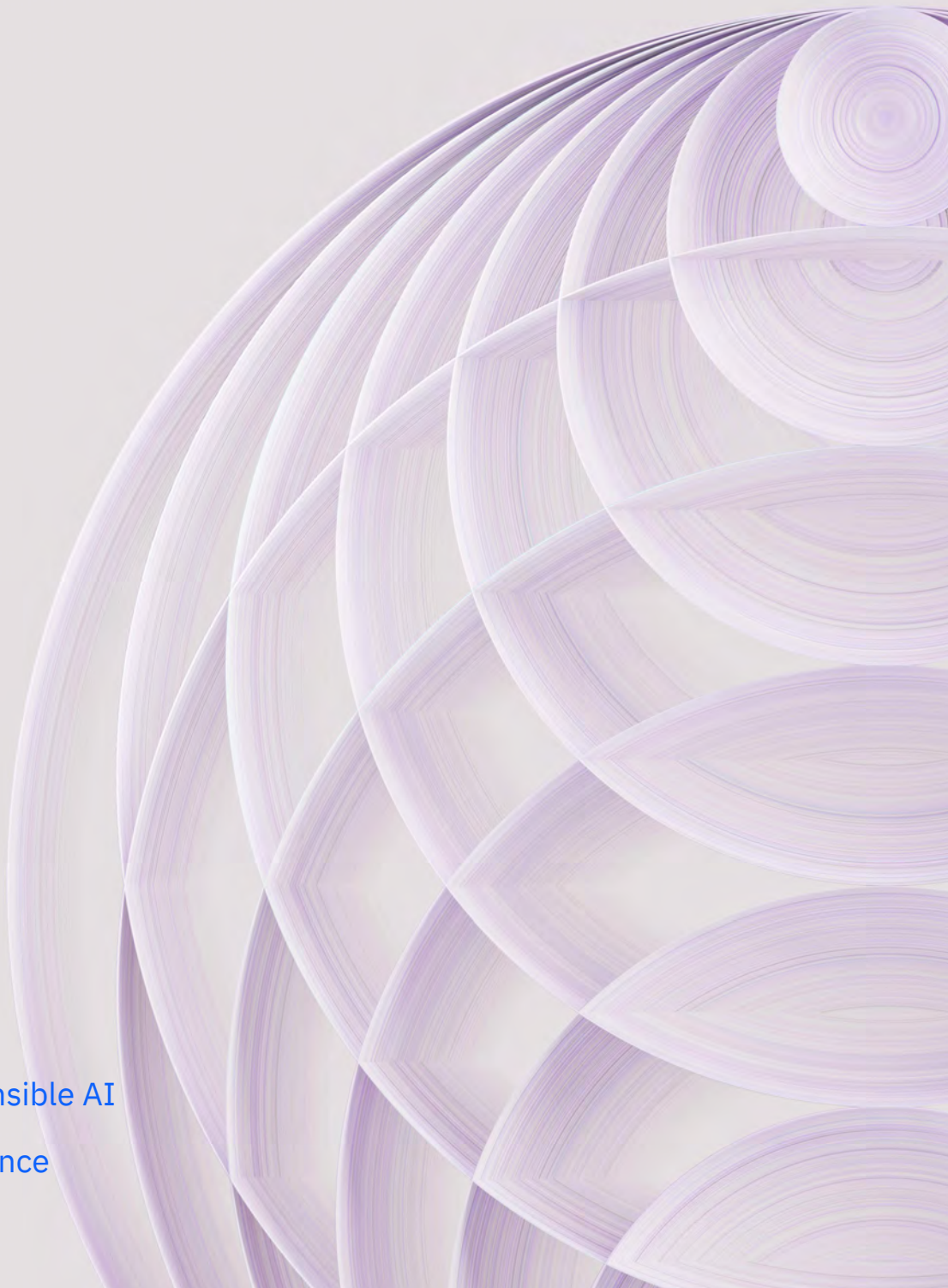


# Getting started with Responsible AI

All you need to know about Responsible AI

Know more about [Responsible AI](#)

Explore [watsonx.governance](#)



IBM's AI Ethics Board was established as a central, cross-disciplinary body to support a culture of ethical, responsible and trustworthy AI throughout IBM.

# A

**AI Ethics Board**

It is essential to rigorously examine training data to prevent embedding real-world biases into AI algorithms, ensuring fair and unbiased decisions.

# B

**Bias control**

Who oversees responsible AI governance? It does not rest with a single individual or department, but is a collective responsibility where everyone must prioritize accountability.

# C

**Collective Ownership**

Responsible AI accepts and promotes the principle that data and insights belong to the creator. Clients' data is their data, and their insights are their insights.

# D

**Data Responsibility**

Empathy is a principle that warrants organizations to understand the societal implications of AI and not just the technological aspects and financial benefits.

# E

**Empathy**

A key expectation from responsible AI is to ensure equitable treatment of individuals or groups by an AI system.

# F

**Fairness**

Governance refers to the guardrails that ensure AI tools and systems are and remain safe and ethical. AI governance helps build responsible AI workflows.

# G

**Governance**

Expressions of hatred with an intent to hurt, humiliate, or insult; rude or hurtful language; toxic words. AI guardrails should remove potentially harmful content.

# H

**HAP: Hate, Abuse and Profanity**

An AI system's ability to be inclusive means it includes data inputs from various demographic groups to avoid underrepresentation or bias.

# I

**Inclusivity**

Human judgment plays a role throughout a seemingly objective system of logical decisions.

# J

**Judgment**

Legal obligations and regulatory requirements that govern the development and deployment of AI technologies.

# L

**Legal Compliance**

A structured approach to developing mitigation strategies that help reduce bias, risks and other potential harmful effects of AI is critical to ensure responsible AI.

# M

**Mitigation**

AI that is based on open technologies that provide a variety of models to cover enterprise use cases and compliance requirements.

# O

**Openess**

This is a key component of how successful the use of AI is in everyday operation, and it contributes immensely to a model's explainability.

# P

**Prediction Accuracy**

Ensuring the datasets used by an organization to train its AI models meet the criteria for accuracy, completeness, validity and fitness for purpose.

# Q

**Quality of Data**

One of the IBM's pillars of trust, robustness is an AI system's ability to effectively handle exceptional conditions, such as abnormalities in input.

# R

**Robustness**

One of the IBM's pillars of trust, AI systems must prioritize and safeguard consumers' privacy and data rights.

# S

**Security and Privacy**

One of the IBM's pillars of trust, transparency is an AI system's ability to include and share information on how it has been designed and developed.

# T

**Transparency**

Responsible AI aims to embed ethical principles into AI applications and workflows to minimize unwanted, negative outcomes associated with the use of AI.

# U

**Unwanted Outcomes**

Visibility into AI operations, data, and models is critical to address security, compliance, governance, accountability and bias.

# V

**Visibility**

AI and data platform to manage, monitor and govern AI applications and models to help operationalize responsible AI at scale.

# W

**watsonx.governance**

One of the foundational pillars of IBM Responsible AI, eXplainable AI (XAI) is a set of processes and methods that enable users to trust outcomes generated by AI/ML.

# X

**XAI: eXplainable Artificial Intelligence**

The practice of a system that continuously tracks and assess the performance and outputs of AI systems, and ensure they meet desired objectives and uphold ethical standards.

# Y

**Yield Monitoring**

An unwavering adherence to ethical principles and legal regulations to mitigate risks and unethical behaviours and ensure that AI systems are deployed responsibly.

# Z

**Zero Tolerance**

© Copyright IBM Corporation 2024.

IBM Corporation, United

Produced in  
United States of America,  
April 2024.

IBM, the IBM logo, ibm.com, watsonx, and watsonx.governance are trademarks or registered trademarks of International Business Machines Corporation, in the United States and/or other countries. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on <https://www.ibm.com/legal/copyright-trademark>.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

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

Statement of Good Security Practices: No IT system or product should be considered completely secure, and no single product, service or security measure can be completely effective in preventing improper use or access. IBM does not warrant that any systems, products or services are immune from, or will make your enterprise immune from, the malicious or illegal conduct of any party.

