AI Risk Management Framework

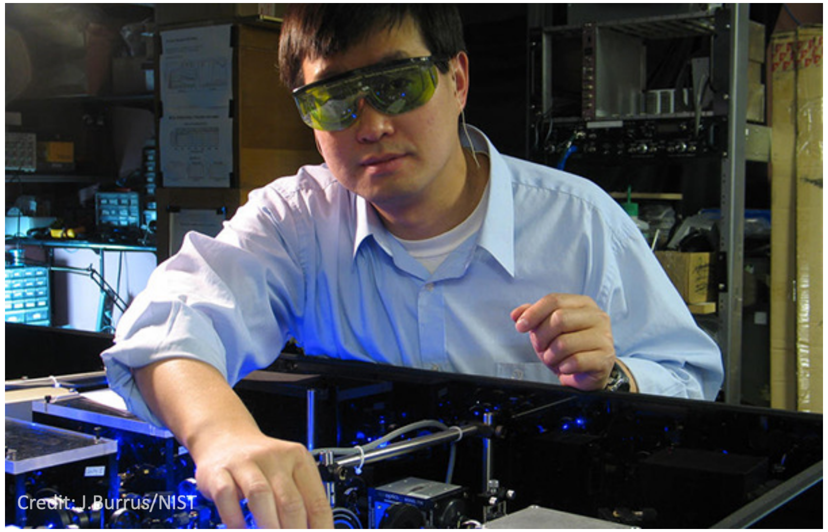
National Institute of Standards and Technology
U.S. Department of Commerce

26 OCT 2022

NIST's Mission

To promote U.S. innovation and industrial competitiveness by advancing **measurement science, standards,** and **technology** in ways that enhance economic security and improve our quality of life

©Robert Rathe



Credit: J.Burrus/NIST

©Nicholas McIntosh Photography

Artificial Intelligence (AI) is rapidly transforming our world. New AI-enabled systems are revolutionizing and benefitting nearly all aspects of our society and economy – everything from commerce and healthcare to transportation and agriculture. But its development and use are not without challenges and risks.

NIST AI Program

CONDUCT FOUNDATIONAL RESEARCH TO ADVANCE TRUSTWORTHY AI TECHNOLOGIES

ADVANCE AI RESEARCH AND INNOVATION ACROSS NIST'S LABORATORY PROGRAMS

ESTABLISH BENCHMARKS AND DEVELOP METRICS TO EVALUATE AI TECHNOLOGIES

PARTICIPATE AND LEAD IN DEVELOPING STANDARDS TO ADVANCE AI INNOVATION

CONTRIBUTE NIST'S TECHNICAL EXPERTISE TO DISCUSSIONS AND DEVELOPMENT OF POLICIES

ENSURE NIST HAS RESOURCES AND EXPERTISE TO CARRY OUT ITS AI PROGRAMS

Key NIST Roles for the Federal Government

NIST AI RISK MANAGEMENT FRAMEWORK

NATIONAL AI ADVISORY COMMITTEE

AI RESEARCH RESOURCE TASK FORCE

FEDERAL AI STANDARDS COORDINATOR

INTERAGENCY COORDINATION WH OSTP/NSTC, TTC, QUAD

STAKEHOLDER OUTREACH

Trustworthy and Responsible AI @ NIST

Cultivate trust in the design, development, use and governance of artificial intelligence technologies and systems.

- Development of AI Risk Management
- AI Research, Standards and Evaluation
- Establishing National AI Advisory Committee

AI RMF:
Managing Risk Through Trustworthy and Responsible AI

What

- Address risks to individuals, communities, organizations, and society
- Congressionally mandated, living document for voluntary use
- Maximize positive impacts, minimize potential negative impacts
- Rights-preserving, aims to operationalize values
- Law and regulation agnostic

How

- Developed in an open, transparent, collaborative process (ongoing)
- Outcome based
- Across context and use cases
- Trustworthy characteristics
- Responsible practices and culture (consideration of impacts)
- Inclusive and equitable

AI RMF Timeline and Engagements

Jul 29, 2021
RFI seeking input

- Comments until **Sept 15, 2021**
- 106 sets of input
- Analysis of responses released on **Oct 15, 2021.**

- Explainable AI paper released **Sept 29, 2021**

Oct 19-21, 2021
NIST AI RMF workshop #1

Dec 13, 2021
AI RMF Concept Paper

- Comments until **Jan 25, 2022**
- 60 sets of input
- Listening sessions

Mar 17, 2022
AI RMF 1st Draft

- Comments until **Apr 29, 2022**
- 95 sets of input
- Bias in AI paper released **Mar 14, 2022**

Mar 29-31, 2022
NIST AI RMF workshop #2

Aug 18, 2022
AI RMF 2nd Draft

- Comments on AI RMF and Playbook until **Sept 29, 2022**
- Call for contributions towards Profiles
- 77 sets of input

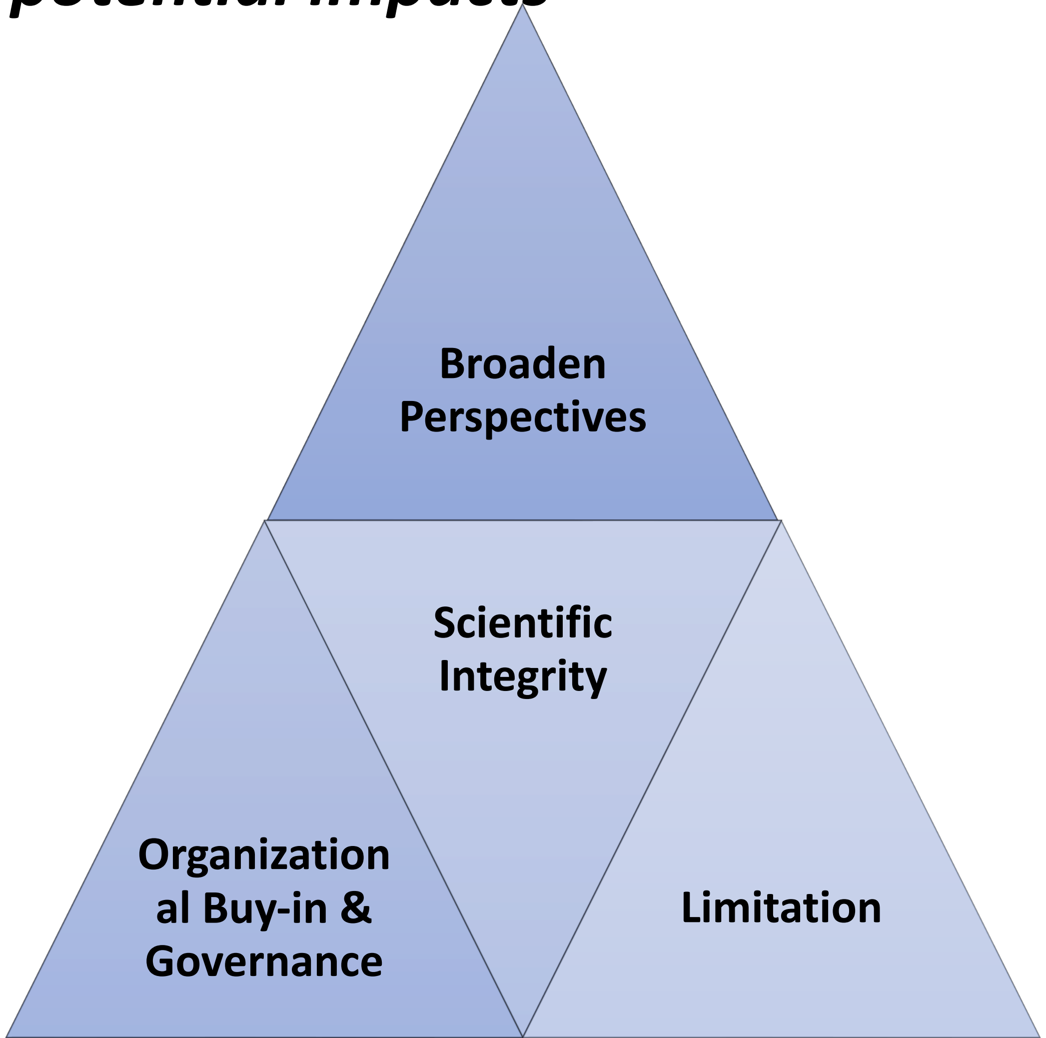
Oct 18-19, 2022
NIST AI RMF workshop #3

Jan 2023
AI RMF 1.0

Transforming Culture – Socio-technical Systems Approach

Takes into consideration the larger social context in which AI operates, its purpose and potential impacts

- Manage risk within/connected to specific operational **context**
  - utilize broader set of perspectives and expertise
  - apply **human-centered** design to AI systems
- Apply the **scientific method** to AI systems
- Set up **governance** structures for the people who build and maintain AI systems
- Consideration of **limitations** from an impact and values-based perspective



Trustworthy AI Characteristics

VALID & RELIABLE

SAFE

FAIR & BIAS IS MANAGED

SECURE & RESILIENT

EXPLAINABLE & INTERPRETABLE

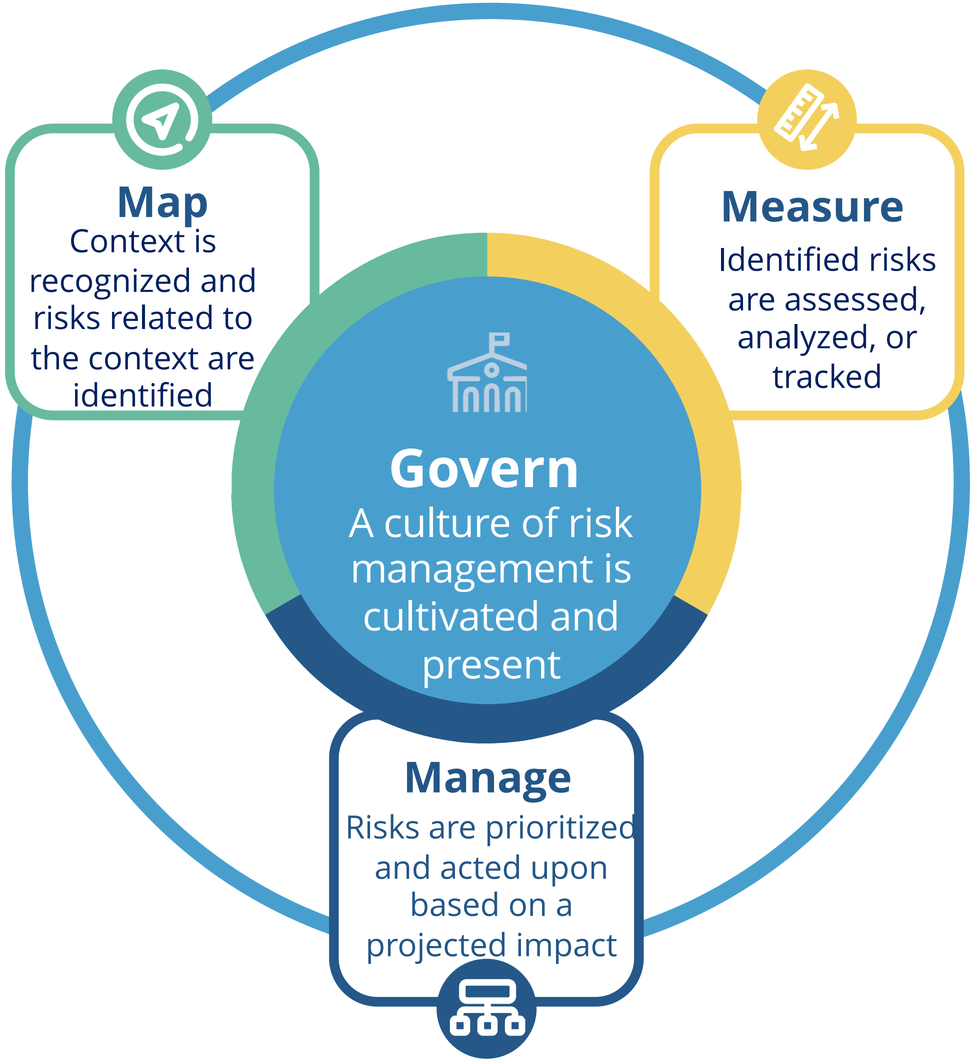
PRIVACY-ENHANCED

ACCOUNTABLE & TRANSPARENT

Trustworthy AI systems should achieve a high degree of control over risk while retaining a high level of performance quality. Achieving this difficult goal requires a comprehensive approach to risk management, with tradeoffs among the trustworthiness characteristics.

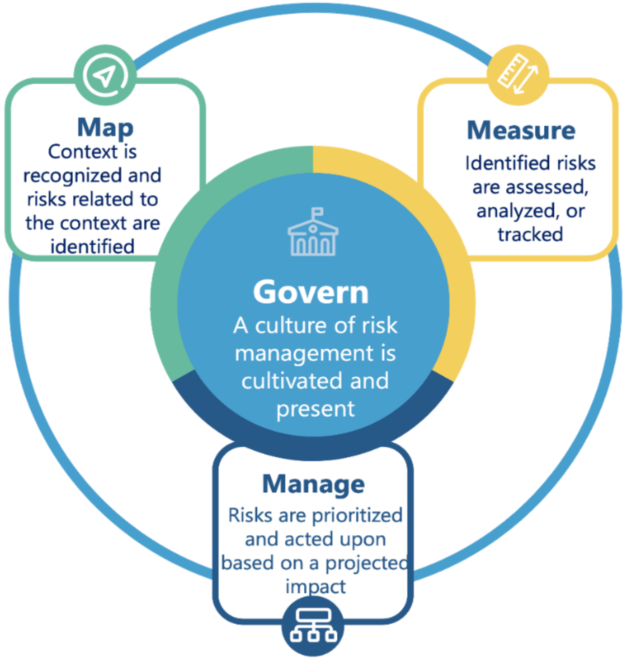
AI Risk Management Framework Core

NIST



Transforming Culture - Socio-technical approach takes into consideration the larger social system in which AI operates, its purpose and potential impacts

NIST AI Risk Management Framework Playbook



Review FAQs

Welcome to the draft NIST AI Risk Management Framework (AI RMF) Playbook – a companion resource for the AI RMF.

The Playbook includes suggested actions, references, and documentation guidance for stakeholders to achieve the outcomes for "**Map**" and "**Govern**" – two of the four proposed functions in the AI RMF. Draft material for the other two functions, **Measure** and **Manage**, will be released at a later date.

This draft Playbook is being released to allow interested parties the opportunity to comment and contribute to the first complete version, to be released in January 2023 with the AI RMF 1.0. The Playbook is an online resource and will be hosted temporarily on GitHub Pages.

NIST welcomes feedback on this draft Playbook.

AI RMF Profiles

Use-case profiles

- Instantiations of the AI RMF functions, categories, and subcategories for a certain application or use case based on the requirements, risk tolerance, and resources of the Framework user.

Temporal profiles

- descriptions of either the current state or the desired, target state of specific AI risk management activities within a given sector, industry, organization, or application context

NIST welcomes contributions towards development of AI RMF use case profiles as well as current and target profiles.

Crosswalks

Table 1: Mapping of AI RMF taxonomy to AI policy documents.

AI RMF	OECD AI Recommendation	EU AI Act (Proposed)	EO 13960
Valid and reliable	Robustness	Technical robustness	Purposeful and performance driven Accurate, reliable, and effective Regularly monitored
Safe	Safety	Safety	Safe
Fair and bias is managed	Human-centered values and fairness	Non-discrimination Diversity and fairness Data governance	Lawful and respectful of our Nation's values
Secure and resilient	Security	Security & resilience	Secure and resilient
Transparent and accountable	Transparency and responsible disclosure Accountability	Transparency Accountability Human agency and oversight	Transparent Accountable Lawful and respectful of our Nation's values Responsible and traceable Regularly monitored
Explainable and interpretable	Explainability		Understandable by subject matter experts, users, and others, as appropriate
Privacy-enhanced	Human values; Respect for human rights	Privacy Data governance	Lawful and respectful of our Nation's values

NIST AI RMF Related Resources

AI RMF PLAYBOOK

AI RMF PROFILES

AI RMF GLOSSARY

AI STANDARDS HUB

AI METRICS HUB

...AND MORE

What's Next?

NIST

AI RMF Profile(s)

Work with SDOs on AI standards

Evaluations of AI RMF effectiveness

AI evaluations and Test beds

Trustworthy AI Resource Center

Crosswalks to other standards, frameworks, etc.

And more …

THANK YOU

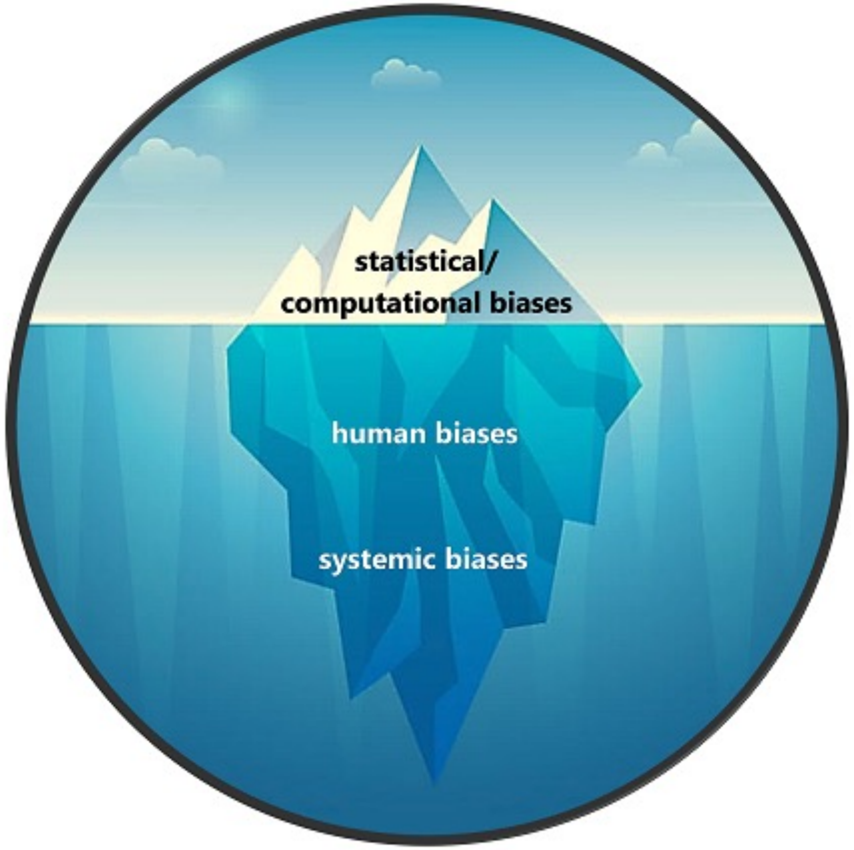
Contact us via email at
aiframework@nist.gov

For more info on the NIST AI RMF, visit
https://www.nist.gov/itl/ai-risk-management-framework

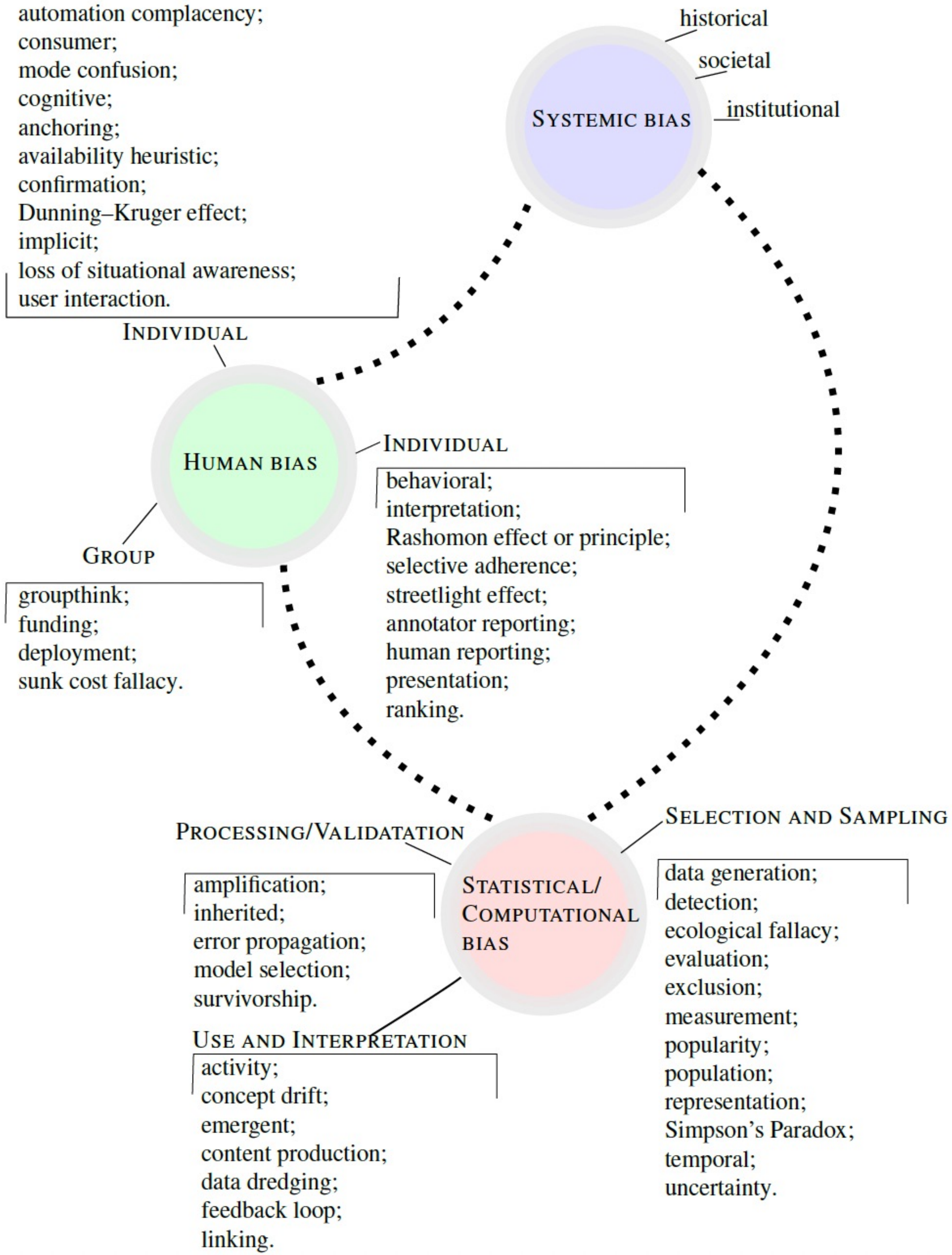
17

Extra

Taxonomy of AI Bias

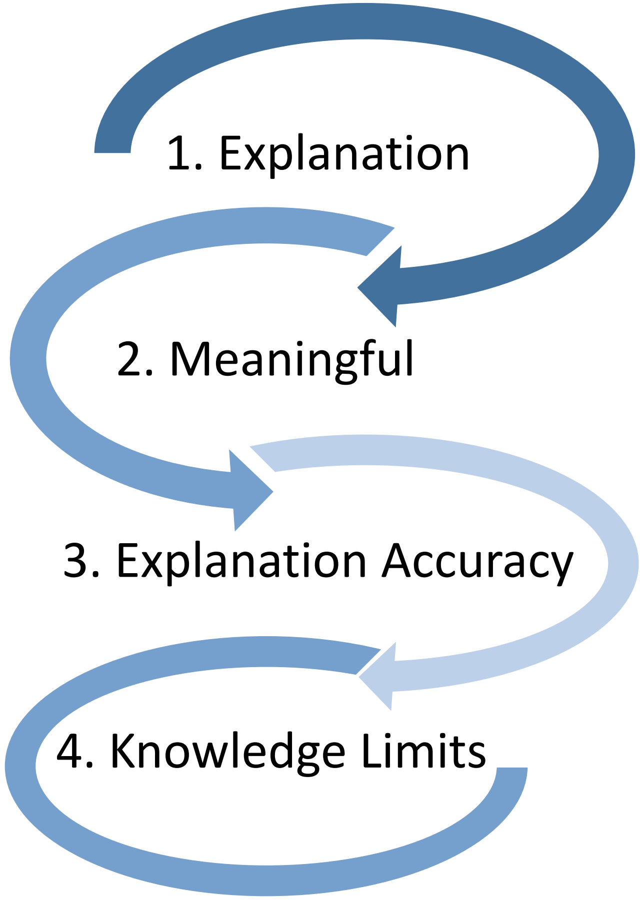


Current focus on computational/statistical bias obfuscates the other two categories



Interpretable and Explainable AI

Four Principles of Explainable AI

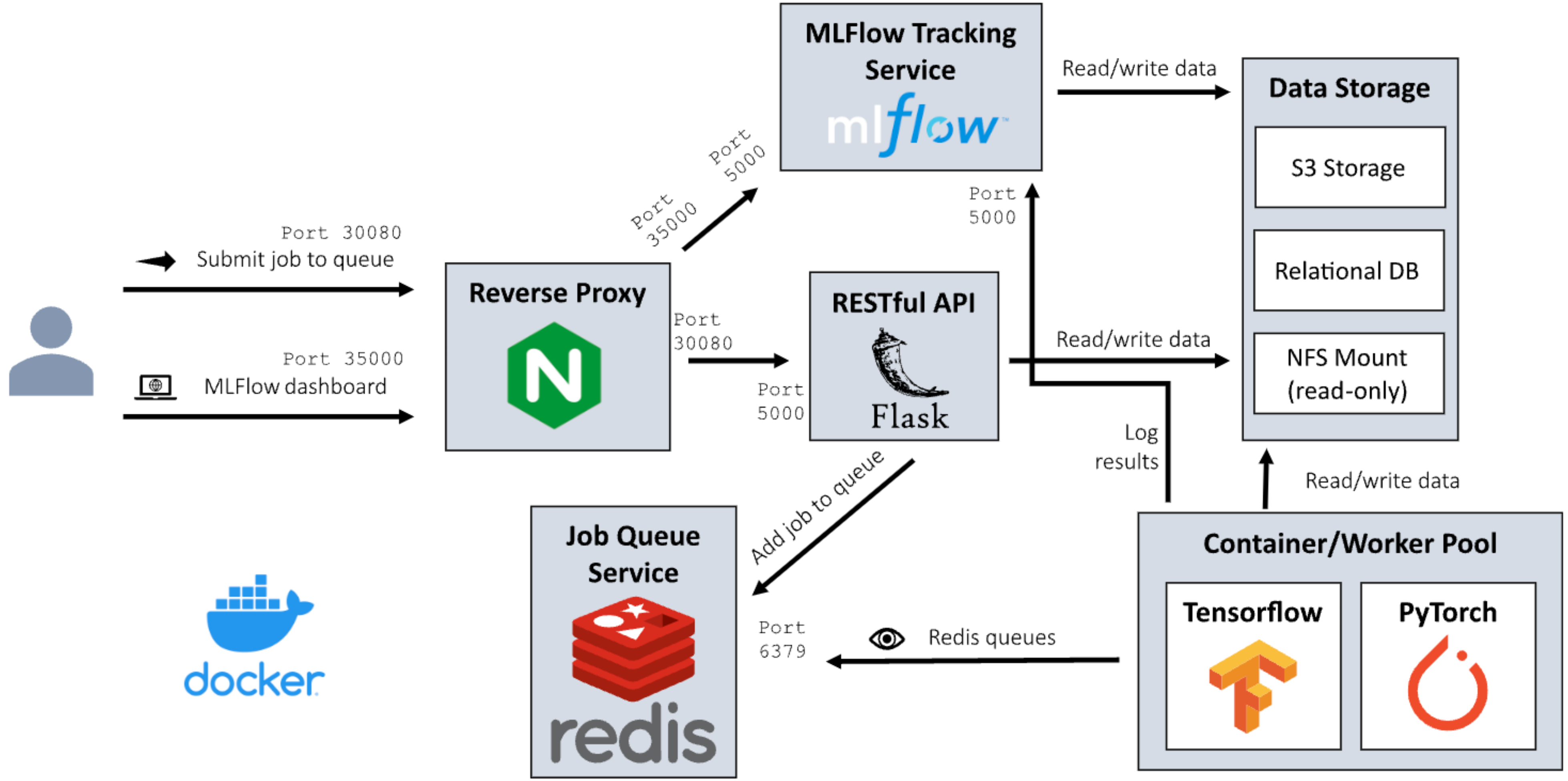


NISTIR 8312: Four Principles of Explainable Artificial Intelligence

NISTIR 8367: Psychological Foundations of Explainability and Interpretability in AI

Secure AI

NIST



Dioptra – Architecture overview

Privacy-preserved AI

NIST

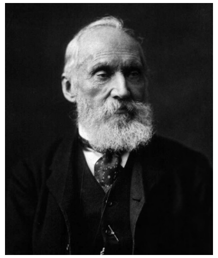
BRIEFING ROOM

U.S. and U.K. Launch Innovation Prize Challenges in Privacy-Enhancing Technologies to Tackle Financial Crime and Public Health Emergencies

JULY 20, 2022 • PRESS RELEASES

Planning for the challenges is being led by the U.S. White House Office of Science and Technology Policy, the U.S. National Institute of Standards and Technology, and the U.S. National Science Foundation in the United States, and the U.K. Centre for Data Ethics and Innovation and Innovate U.K. in the United Kingdom. The U.S. challenge is funded and administered by the U.S. National Institute of Standards and Technology and the U.S. National Science Foundation.

AI Evaluations



"If you cannot measure it, you cannot improve it."

Devise and revise metrics

Testbeds and test data

Assessing real world performance

Interoperable tests and test results

USG AI Standards Coordinator

Facilitate ongoing discussions between the U.S. private sector and federal agencies to strengthen private-public sector coordinator

Participate at and contribute to AI standards development activities