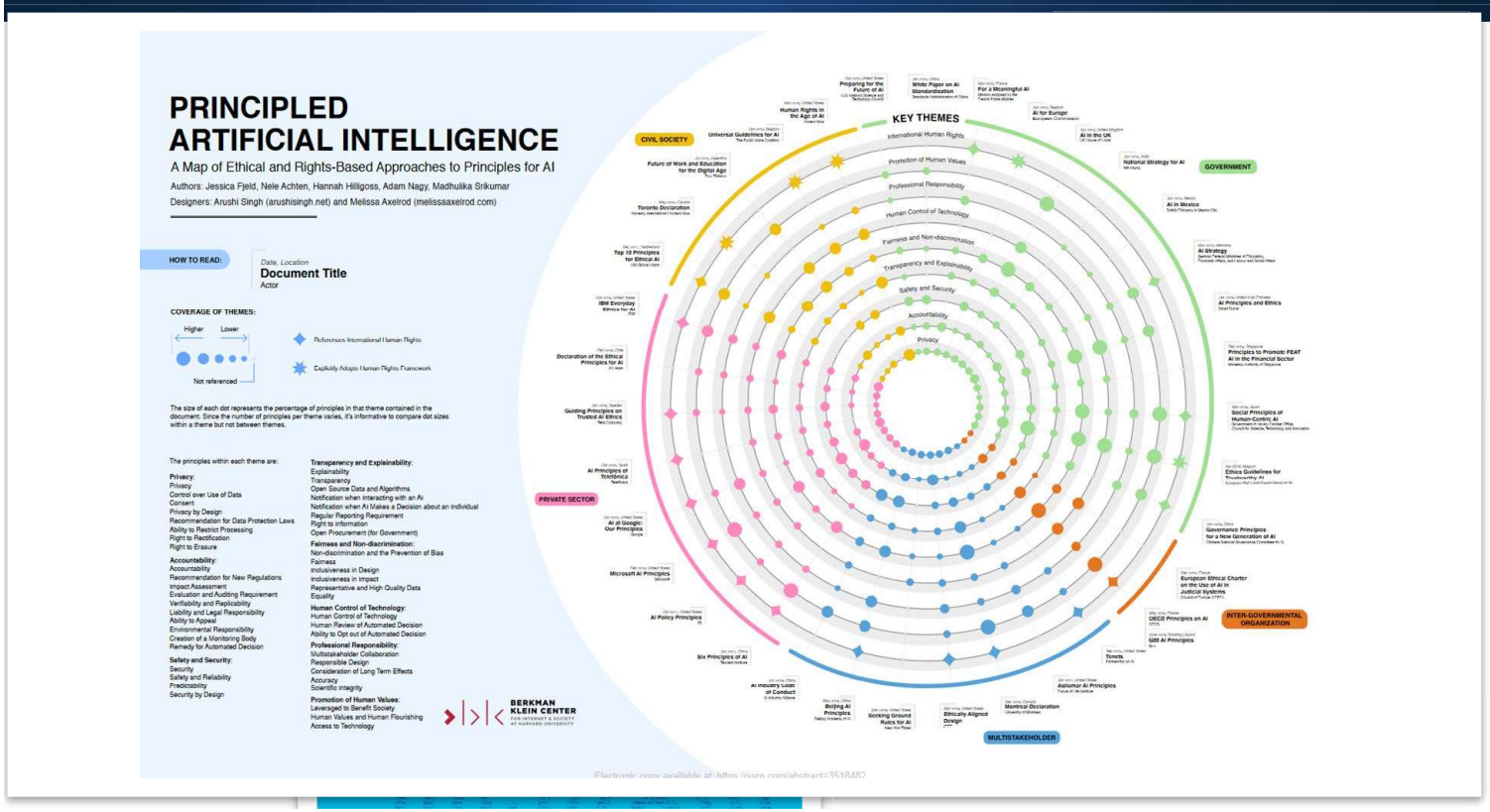
Jesse Dunietz
NIST ITL

The Artificial Intelligence Risk Management Framework (AI RMF 1.0)

NIST | NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

As risks from AI became more apparent, many frameworks of principles emerged— but they remained too high-level for implementers.



The AI RMF offers voluntary guidance to operationalize principles for AI governance into concrete targets and actions.

NIST

Table 1: Categories and subcategories for the GOVERN function.

Categories	Subcategories
GOVERN 1: Policies, processes, procedures, and practices across the organization related to the mapping, measuring, and managing of AI risks are in place, transparent, and implemented	**GOVERN 1.1:** Legal and regulatory requirements involving AI are understood, managed, and documented. **GOVERN 1.2:** The characteristics of trustworthy AI are integrated into organizational policies, processes, procedures, and practices. **GOVERN 1.3:** Processes, procedures, and practices are in place to determine the needed level of risk management activities based on the organization's risk tolerance. **GOVERN 1.4:** The risk management process and its outcomes are established through transparent policies, procedures, and other

Table 2: Categories and subcategories for the MAP function.

Categories	Subcategories
MAP 1: Context is established and understood.	**MAP 1.1:** Intended purposes, potentially beneficial uses, context-specific laws, norms and expectations, and prospective settings in which the AI system will be deployed are understood and documented. Considerations include: the specific set or types of users along with their expectations; potential positive and negative impacts of system uses to individuals, communities, organizations, society, and the planet; assumptions and related limitations about AI system purposes, uses, and risks across the development or product AI lifecycle; and related TEVV and system metrics. **MAP 1.2:** Interdisciplinary AI actors, competencies, skills, and capacities for establishing context reflect demographic diversity and broad domain and user experience expertise, and their par-

Table 3: Categories and subcategories for the MEASURE

Categories	Subcategories
MEASURE 1: Appropriate methods and metrics are identified and applied.	**MEASURE 1.1:** Approaches and metrics for mea risks enumerated during the MAP function are sele mentation starting with the most significant AI r or trustworthiness characteristics that will not measured are properly documented. **MEASURE 1.2:** Appropriateness of AI metrics a of existing controls are regularly assessed and upd reports of errors and potential impacts on affected **MEASURE 1.3:** Internal experts who did not ser developers for the system and/or independent

Table 4: Categories and subcategories for the MANAGE

Categories	Subcategories
MANAGE 1: AI risks based on assessments and other analytical output from the MAP and MEASURE functions are prioritized, responded to, and managed.	**MANAGE 1.1:** A determination is made as to system achieves its intended purposes and stated whether its development or deployment should p **MANAGE 1.2:** Treatment of documented AI risk based on impact, likelihood, and available resour **MANAGE 1.3:** Responses to the AI risks deemed identified by the MAP function, are developed, pla umented. Risk response options can include miti ring, avoiding, or accepting. **MANAGE 1.4:** Negative residual risks (defined unmitigated risks) to both downstream acquirers

- ✓ Detailed
- ✓ Flexible
- ✓ Systematic
- ✓ Sensitive to actors and context

Agenda

Motivation

AI RMF Overview

Tools for AI RMF Implementation

Managing risk entails several key challenges.

Risk is hard to measure

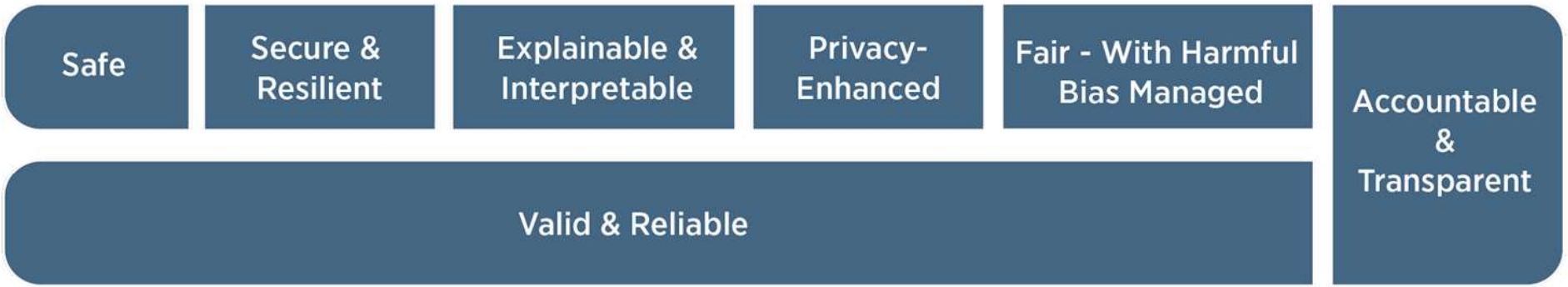
Risk tolerances vary

Risks must be prioritized

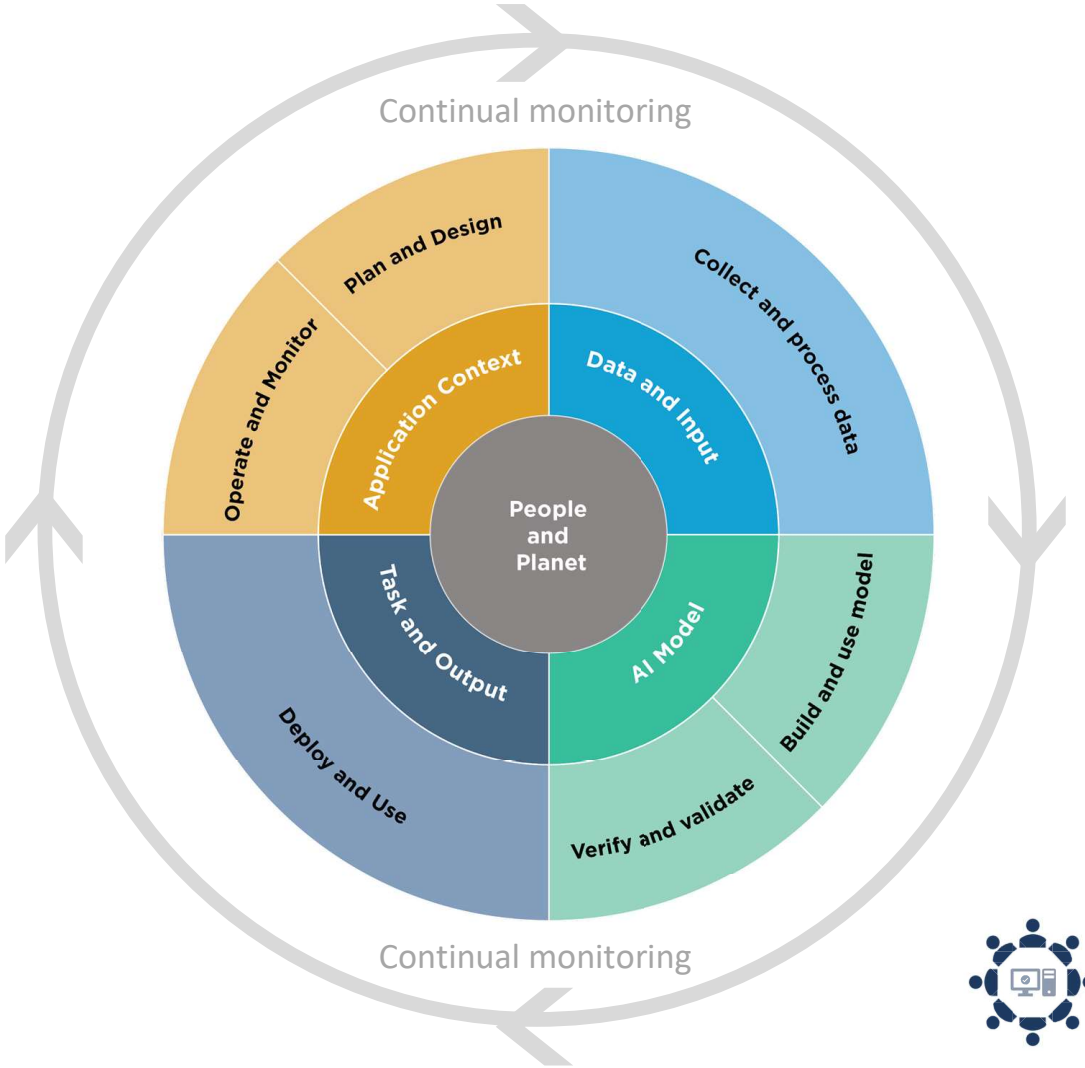
Risk management must be integrated

The core precept of the AI RMF is
AI system trustworthiness within a
culture of responsible AI practice and use.

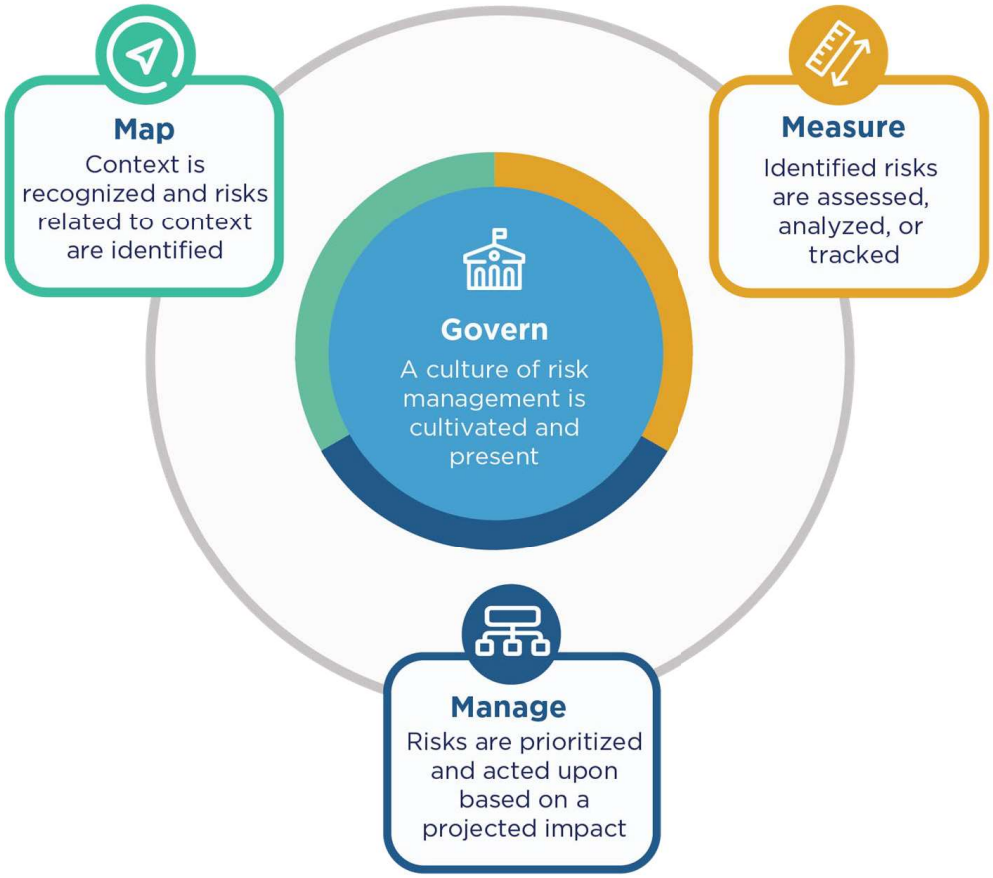
AI system trustworthiness can be defined in terms of well-understood characteristics.



Beyond the system, a culture of responsible practice and use must pervade activities across the entire AI lifecycle.



The AI RMF Core lays out four organizational functions to facilitate trustworthy systems and responsible practice and use.



The **GOVERN** function is about fostering a risk-aware culture.

GOVERN 2: Accountability structures are in place so that the appropriate teams and individuals are empowered, responsible, and trained for mapping, measuring, and managing AI risks.

GOVERN 4: Organizational teams are committed to a culture that considers and communicates AI risk.

GOVERN 5: Processes are in place for robust engagement with relevant AI actors.

The **MAP** function establishes the context in which risks could materialize.

MAP 1: Context is established and understood.

MAP 3: AI capabilities, targeted usage, goals, and expected benefits and costs compared with appropriate benchmarks are understood.

MAP 5: Impacts to individuals, groups, communities, organizations, and society are characterized.

The **MEASURE** function sets up objective, repeatable, and scalable processes for test, evaluation, verification, & validation (TEVV).

MEASURE 1: Appropriate methods and metrics are identified and applied.

MEASURE 2: AI systems are evaluated for trustworthy characteristics.

MEASURE 3: Mechanisms for tracking identified AI risks over time are in place.

MEASURE 4: Feedback about efficacy of measurement is gathered and assessed.

The **MANAGE** function is how organizations forestall **MAP**ped and **MEASURE**d risks, and respond to them when they materialize.

NIST

Prevention measures

- Data management
- Risk transfer mechanisms (e.g., insurance, warranties)
- System modification (e.g., model editing)
- Software quality assurance

Response measures

- Decommissioning mechanisms ("kill switches")
- Incident response plans
- Recourse and feedback mechanisms
- Monitoring (bias, performance, security)
- Information sharing

Agenda

Motivation

AI RMF Overview

Tools for AI RMF Implementation

The RMF is accompanied by a suite of tools in the Trustworthy and Responsible AI Resource Center (AIRC).

Crosswalk Documents

NIST AI RMF Crosswalks are produced by by NIST or other organizations and are intended to provide a mapping of concepts and terms between the AI RMF and other guidelines, frameworks, standards and regulation documents. Organizations are encouraged to submit crosswalks to NIST at aiframework@nist.gov for potential posting on this page. The below list includes crosswalks that have been submitted, reviewed and accepted to date.

Glossary

NIST is releasing "The Language of Trustworthy AI: An In-Depth Glossary of Terms". This effort seeks to promote a shared understanding and improve communication among individuals and organizations seeking to operationalize trustworthy and responsible AI through approaches such as the NIST AI Risk Management Framework (AI RMF). The Glossary is being released in beta format as a spreadsheet, as approaches to visualize the relationships between and among these terms continues. A final glossary release will be launched at a later date.

Technical and Policy Documents

The section provides direct links to NIST documents related to the AI RMF (NIST AI-100) and NIST AI Publication Series, as well as NIST-funded external resources in the area of Trustworthy and Responsible AI. New documents will be added as they are completed.

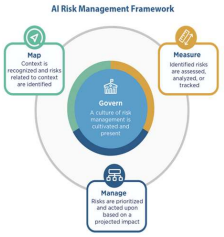
NIST AI RMF Playbook

The Playbook provides suggested actions for achieving the outcomes laid out in the AI Risk Management Framework (AI RMF) Core (Tables 1-4 in AI RMF 1.0). Suggestions are aligned to each sub-category within the four AI RMF functions (Govern, Map, Measure, Manage).

The Playbook is neither a checklist nor set of steps to be followed in its entirety.

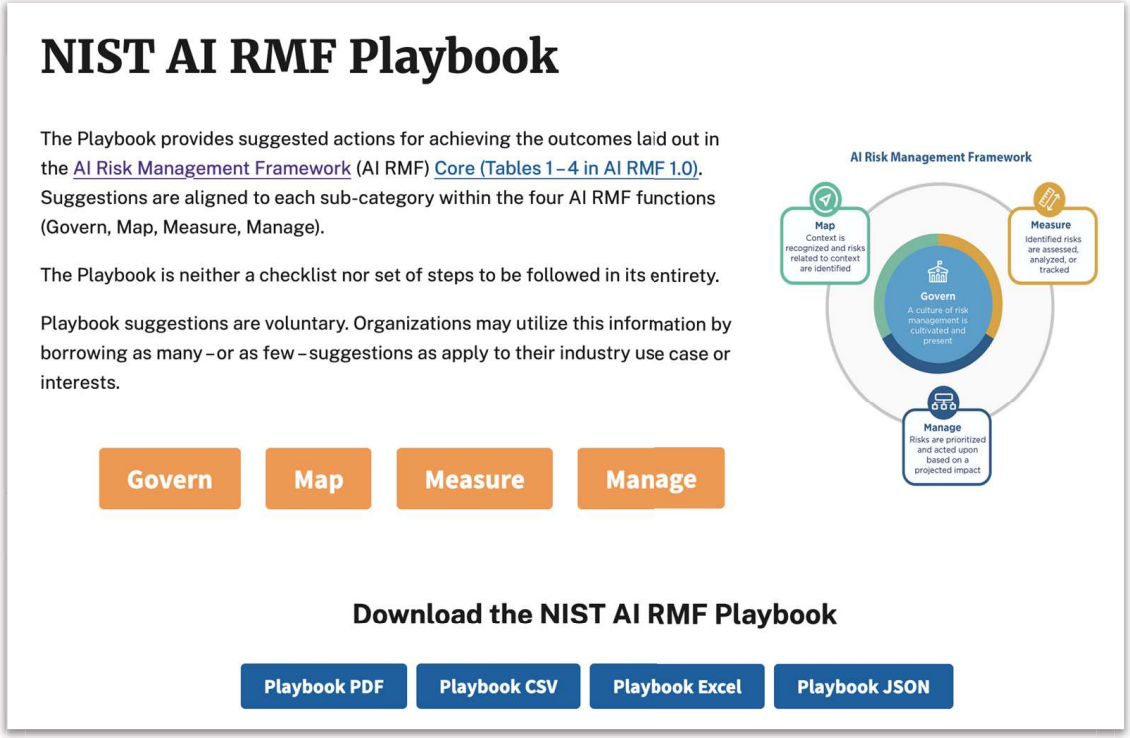
Playbook suggestions are voluntary. Organizations may utilize this information by borrowing as many - or as few - suggestions as apply to their industry use case or interests.

Govern | Map | Measure | Manage

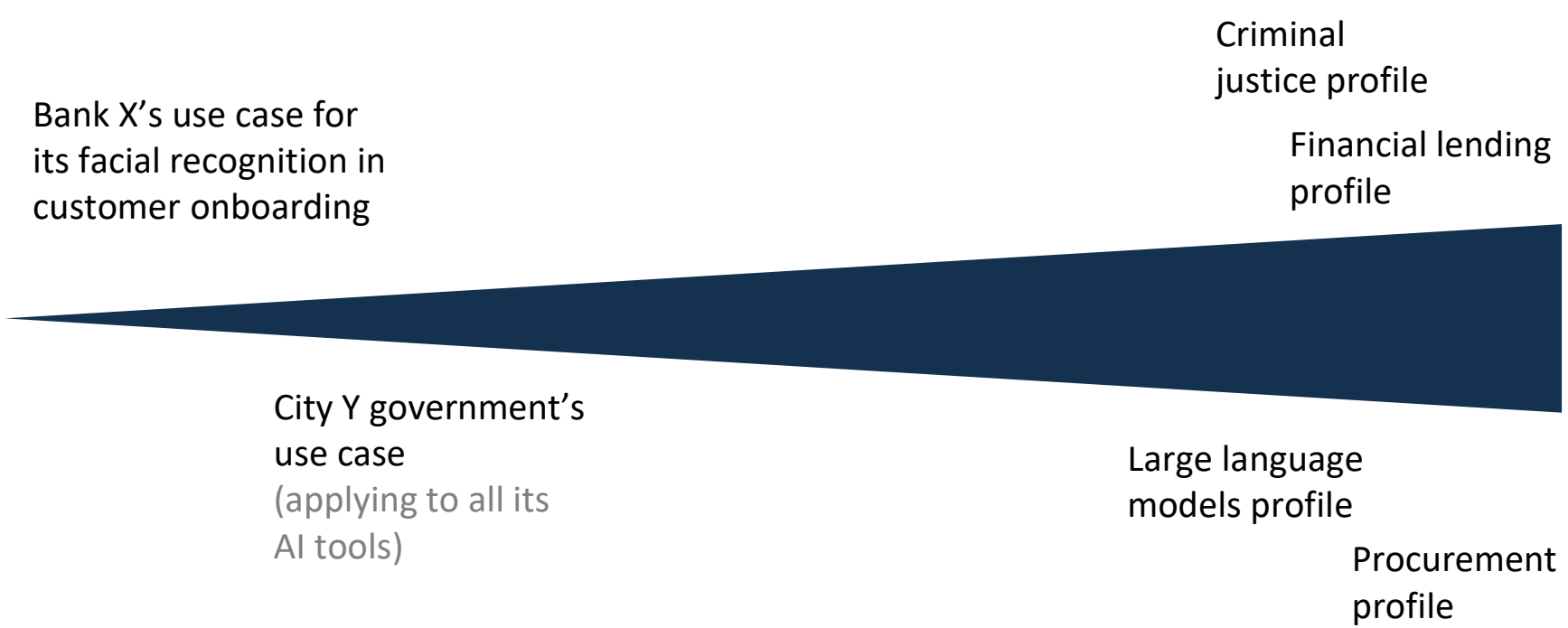


...

The Playbook was developed to give organizations a more detailed how-to for achieving the outcomes described in the Framework Core.



The AI RMF is being implemented at many scales, from individual systems'/organizations' "use cases" to "profiles" for entire sectors or technologies.



For more information, we encourage you to access NIST resources, or reach out directly!

https://ww.nist.gov/itl/ai-risk-management-framework

https://airc.nist.gov/

AIFramework@nist.gov