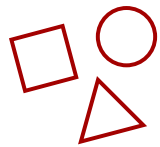# VCAT Briefing:
Artificial Intelligence

Elham Tabassi

Associate Director for Emerging Technologies

Information Technology Laboratory

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

13 JUN 2023

NIST AI Program

CONDUCT FOUNDATIONAL RESEARCH TO ADVANCE TRUSTWORTHY AI TECHNOLOGIES
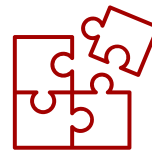

ADVANCE AI RESEARCH AND INNOVATION ACROSS THE NIST LABORATORY PROGRAMS

ESTABLISH BENCHMARKS AND DEVELOP METRICS TO EVALUATE AI TECHNOLOGIES

PARTICIPATE AND LEAD IN THE DEVELOPMENT OF STANDARDS TO ADVANCE AI INNOVATION

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

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

Trustworthy and Responsible AI @ NIST

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

Guidance and Tools
AI Risk Management Framework, AI Standards

Research and development
National AI Institute on Trustworthy AI (TRAILS)

Lead and Convene
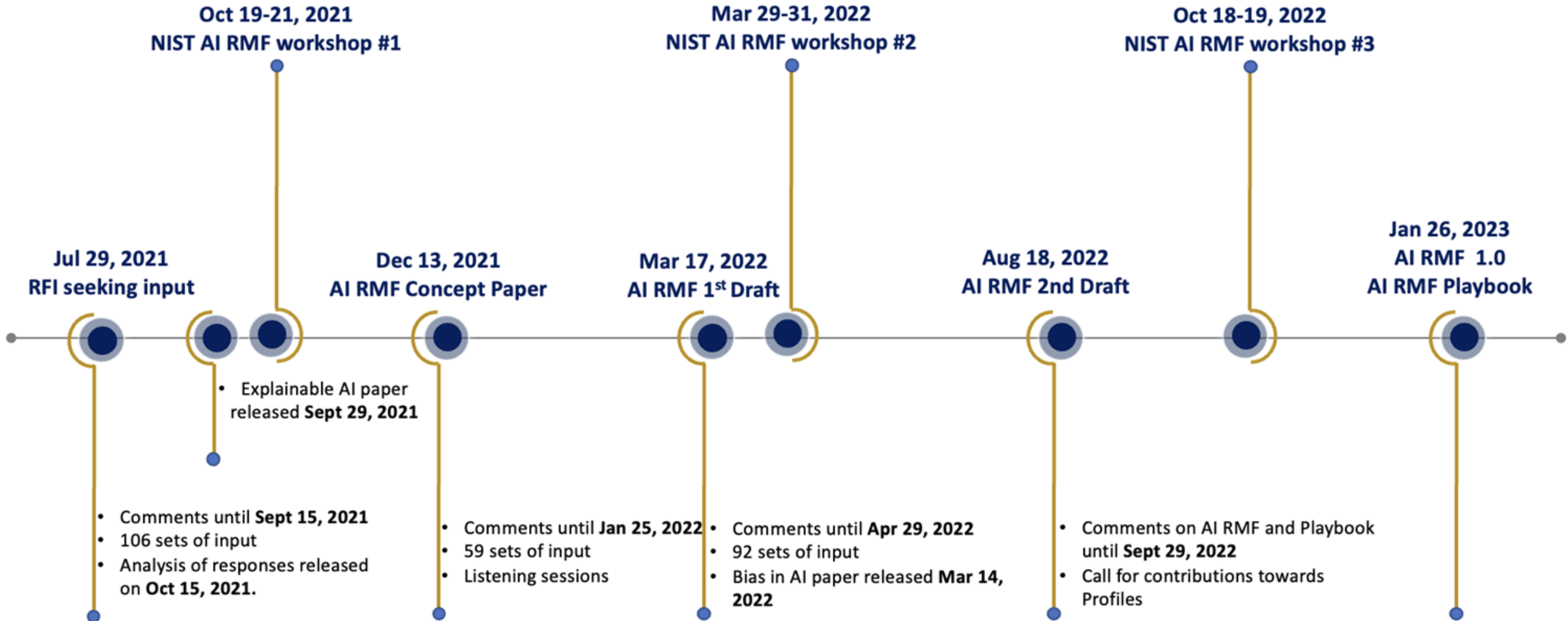National AI Advisory Committee, international engagements, and more.

3

NIST AI Risk Management Framework

A voluntary resource for organizations designing, developing, deploying, or using AI systems to manage AI risks and promote trustworthy and responsible AI

THE PATH TO AI RMF 1.0

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

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

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

Jul 29, 2021
RFI seeking input

Dec 13, 2021
AI RMF Concept Paper

Mar 17, 2022
AI RMF 1st Draft

Aug 18, 2022
AI RMF 2nd Draft

Jan 26, 2023
AI RMF 1.0
AI RMF Playbook

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

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

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

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

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

5

AI RMF CORE

Map
Context is recognized and risks related to context are identified

Measure
Identified risks are assessed, analyzed, or tracked

Govern
A culture of risk management is cultivated and present

Manage
Risks are prioritized and acted upon based on a projected impact

Perspectives about NIST AI RMF

"Google congratulates NIST on the release of its AI Risk

Partnership on AI

"The automotive industry

"The Federation of American Scientists is pleased to see the
National Institute of Standards and Technology's release of the
AI Risk Management Framework (AI RMF 1.0). We believe that

"IBM applauds NIST on the release of the AI Risk Management

"Credo AI welcomes the release "The Bipartisan Policy Center is pleased to see

"On behalf of Amazon Web Services, I would like to commend the NIST

We support the AI RMF and look forward to its
continued development to benefit our research
community, as well as for society as a whole."

National Science Foundation

David Danks, UC San Diego Halıcıoğlu Data Science Institute
(HDSI) and Department of Philosophy professor

"The National Fair Housing Alliance (NFHA)

Applauds NIST on the launch of its AI Risk

Management Framework (AI RMF 1.0) which

"The Center for Security and Emerging Technology (CSET),

We appreciate the opportunity to engage with
forward to working with BSA members to deter
practices to advance consistent approaches ar

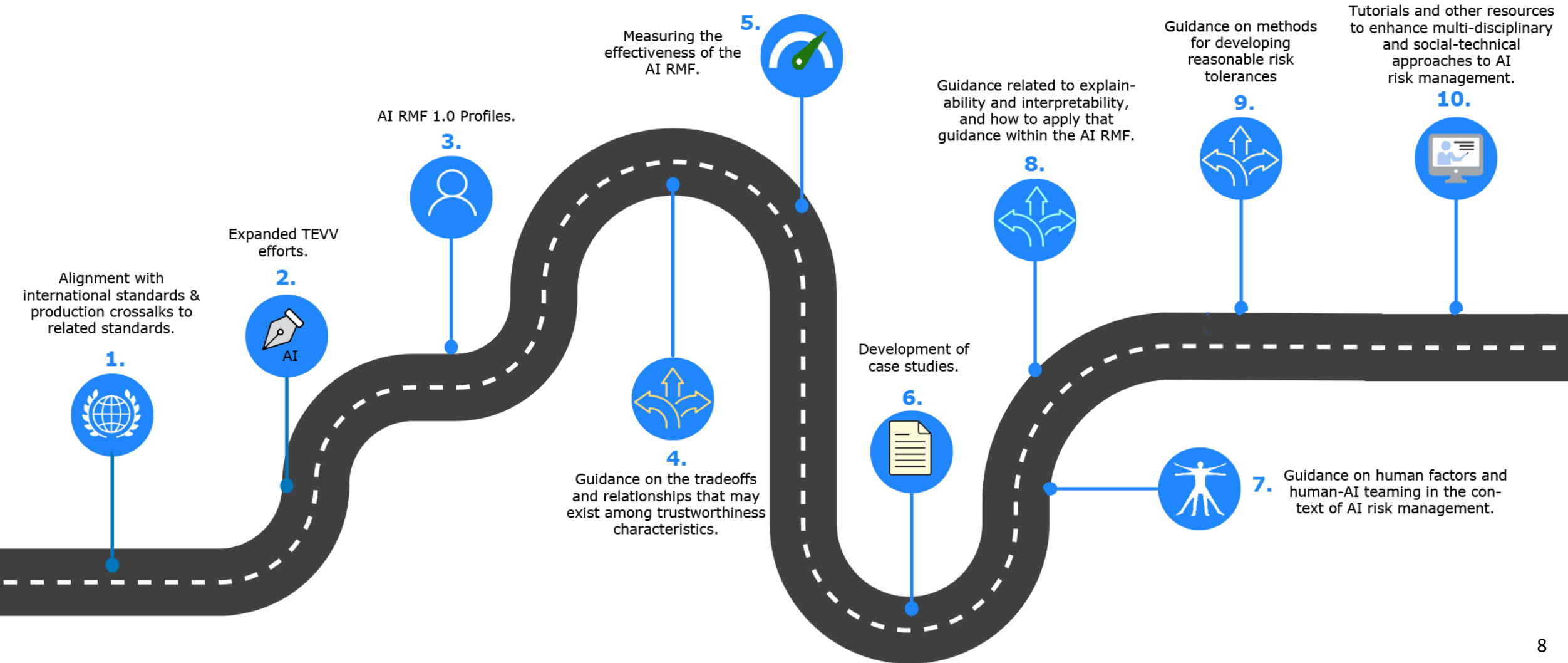*BSA | The Software Alliance*

Special Competitive Studies Project

"The Chamber appreciates

"The NIST AI Risk Management Framework
represents a collaborative, thoughtful approach to
addressing the socio-technical nature of AI systems.
Natasha Crampton, Chief Responsible AI Officer –
Office for Responsible AI, Microsoft

CTIA applauds NIST

AI RMF Roadmap

Alignment with international standards & production crosswalks to related standards.
1.

Expanded TEVV efforts.
2.

AI RMF 1.0 Profiles.
3.

Guidance on the tradeoffs and relationships that may exist among trustworthiness characteristics.
4.

Measuring the effectiveness of the AI RMF.
5.

Development of case studies.
6.

Guidance on human factors and human-AI teaming in the context of AI risk management.
7.

Guidance related to explainability and interpretability, and how to apply that guidance within the AI RMF.
8.

Guidance on methods for developing reasonable risk tolerances
9.

Tutorials and other resources to enhance multi-disciplinary and social-technical approaches to AI risk management.
10.

NIST TRUSTWORTHY AI RESOURCE CENTER

AI RMF PLAYBOOK

AI RMF PROFILES

AI RISK GLOSSARY

AI METRICS HUB

...AND MORE
airc.nist.gov

USG AI Standards Coordinator

U.S. LEADERSHIP IN AI:
A Plan for Federal Engagement in Developing
Technical Standards and Related Tools

Prepared in response to Executive Order 13859
Submitted on August 9, 2019

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

Outreach to connect with all known federal efforts relating to AI standards development and use.

Support and contribute to development of AI standards. Develop crosswalk to international standards.

Trustworthy AI in Law and Society (TRAILS)

*"Today, the ability to measure AI system trustworthiness and its impacts on individuals, communities and society is limited. **TRAILS can help advance our understanding of the foundations of trustworthy AI, ethical and societal considerations of AI,** and how to build systems that are trusted by the people who use and are affected by them."* Under Secretary of Commerce for Standards and Technology and NIST Director Laurie E. Locascio

Led by UNIVERSITY OF MARYLAND

In partnership with THE GEORGE WASHINGTON UNIVERSITY WASHINGTON, DC — MORGAN STATE UNIVERSITY — CORNELL UNIVERSITY

Co-Funded by NSF — NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

Example of International Engagements

US-UK COOPERATION
ON AI R&D

OECD NETWORK
OF EXPERTS ON AI

TRADE AND TECHNOLOGY
COUNCIL TECHNOLOGY
(STANDARDS, AI)

QUADRILATERAL
SECURITY DIALOGUE
STANDARDS SUB-GROUP

National AI Advisory Committee (NAIAC)

THANK YOU

www.nist.gov/artificial-intelligence
airc.nist.gov