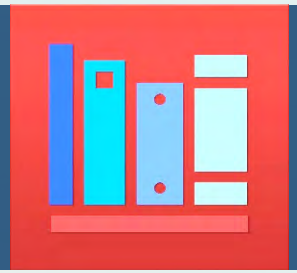
Development of Standards for Artificial Intelligence Systems



- Opening Remarks by Victor Hall, Deputy Division Director, RES/DSA

- **Session Chair:** Luis Betancourt, Branch Chief, RES/DSA/AAB

- **Panelists/Speakers:**
  - Elham Tabassi (NIST)
  - Jonnie Bradley (DOE)
  - Ahmad Al Rashdan (INL)
  - Trey Hathaway (NRC)
  - Thiago Seuaciuc-Osorio (EPRI)

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Artificial Intelligence Risk Management Framework (AI RMF 1.0)

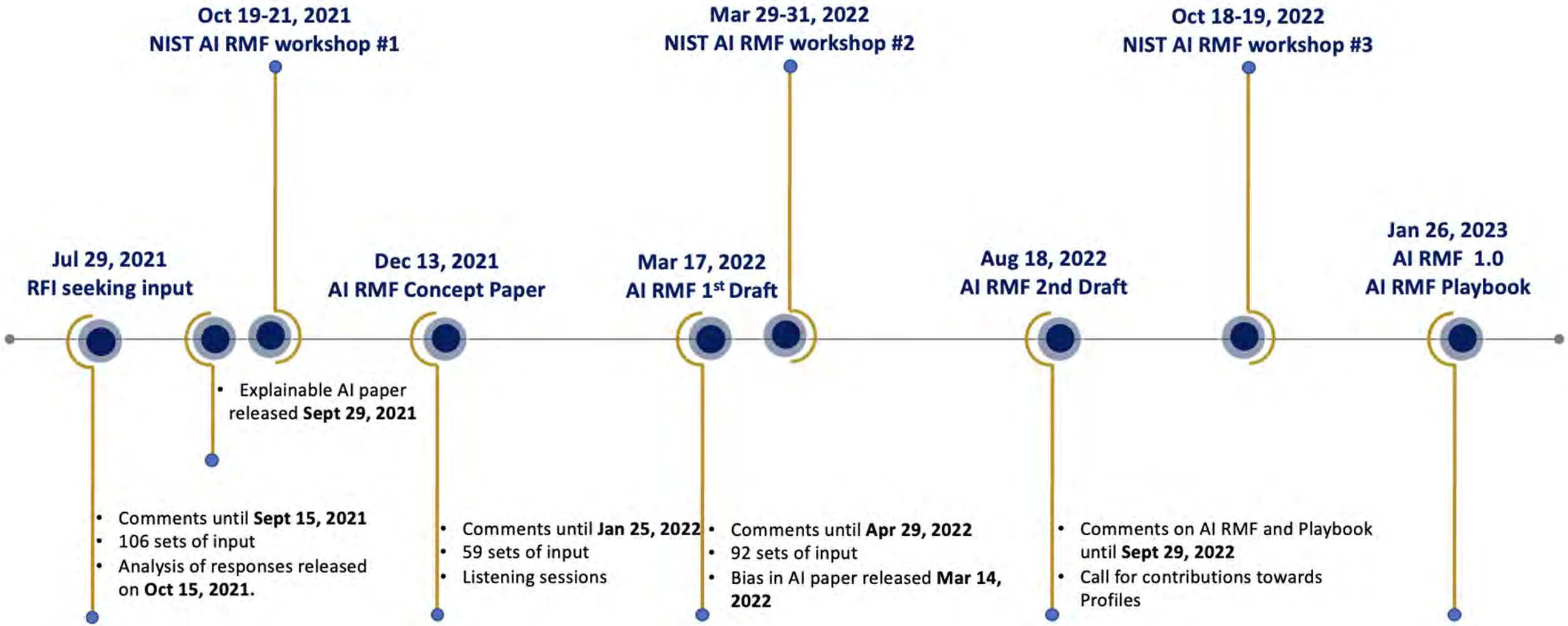
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

WHAT IS THE AI RMF?

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

- **Rights-preserving**
- **Flexibly applied**
- **Measurable**

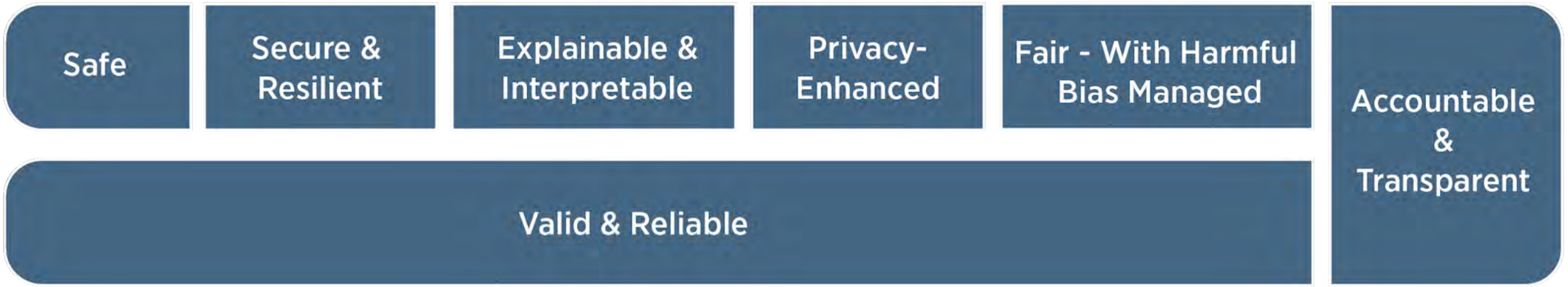
THE PATH TO AI RMF 1.0



KEY TERMINOLOGY

- **AI systems:** engineered or machine-based system that **generates outputs such as predictions, recommendations, or decisions** influencing real or virtual environments and operating with varying levels of autonomy.
- **Risk:** composite measure of an event's probability of occurring and the magnitude or degree of the consequences of the corresponding event. The impacts, or consequences, of AI systems can be **positive, negative, or both and can result in opportunities or threats**.

AI RISKS AND TRUSTWORTHINESS



AI RISK MANAGEMENT CHALLENGES

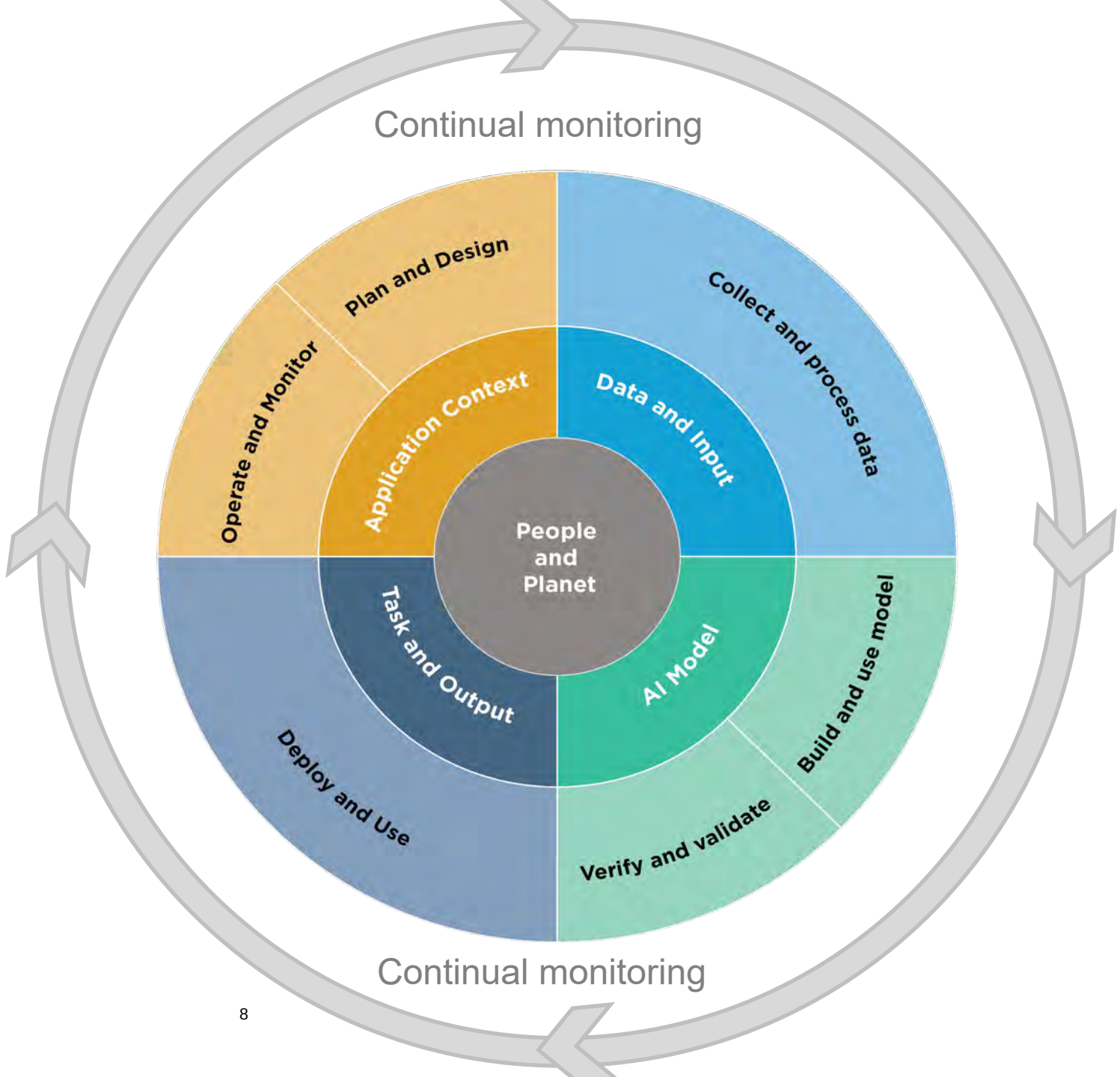
Risk measurement

Risk tolerance

Risk prioritization

Risk integration & management

AUDIENCE: AI LIFECYCLE AND AI ACTORS



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

AI RMF PLAYBOOK

Type	
Govern	19
Manage	13
Map	18
Measure	22

AI Actors	
Affected Individuals and Commun...	12
AI Deployment	37
AI Design	12
AI Development	13
AI Impact Assessment	23
Domain Experts	18

Topics	
Accountability and Transparency	6
Adversarial	3
AI Deployment	3
AI Incidents	3
Contestability	1
Context of Use	7

Download the NIST AI RMF Playbook

Playbook PDF **Playbook CSV** **Playbook Excel** **Playbook JSON**

AI RMF PROFILES

Implementations of the AI RMF functions, categories, and subcategories for a specific setting or application based on the requirements, risk tolerance, and resources of the Framework user.

Use-case profiles; e.g., hiring or fair housing

Temporal profiles; e.g., current state vs. the target state

Cross-sectoral profiles; e.g., large language models, cloud-based services or acquisition

AI RMF CROSSWALKS

alignment with international standards as is top priority for NIST and the AI community.

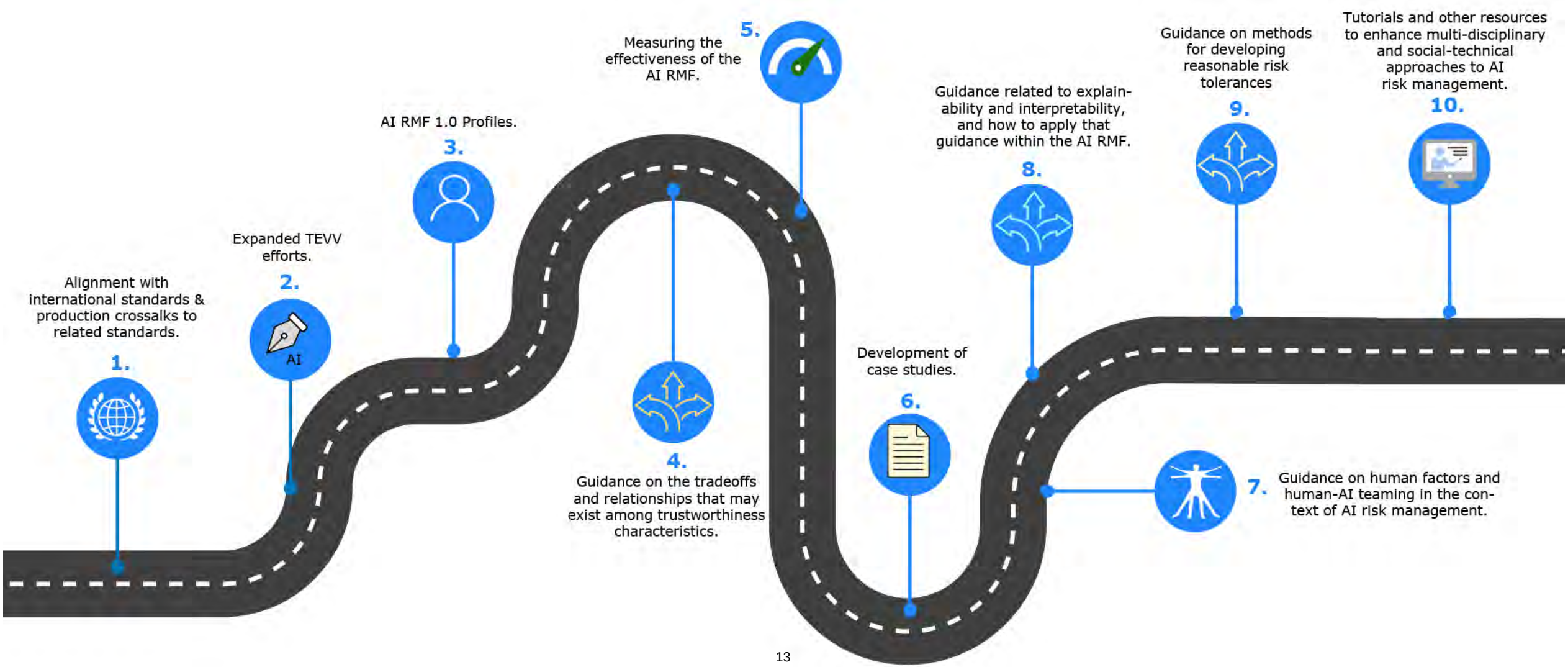
Crosswalk
AI RMF (1.0) and ISO/IEC FDIS 23894 Information technology - Artificial intelligence Guidance on risk management

**Note ISO/IEC FDIS 23894 is a final draft international standard and expected to be published in 2023.*

AI RMF 1.0 Function	ISO/IEC FDIS 23894
GOVERN: Culture of risk management is cultivated and present.	5.2 Leadership and commitment 5.3 Integration (references ISO 31000:2018) 5.4 Design 5.4.1 Understanding the organization and its context 5.4.2 Articulating risk management commitment (references ISO 31000:2018) 5.4.3 Assigning organizational roles, authorities, responsibilities, and accountabilities (references ISO 31000:2018) 5.4.4 Allocating resources (ref. ISO 31000:2018) 5.4.5 Establishing communication and consultation (references ISO 31000:2018)
MAP: Context is recognized and risks related to context are identified.	5.4.1 Understanding the organization and its context 6.3.2 Defining the scope 6.3.3 External and internal context 6.3.4 Defining risk criteria 6.4.2 Risk identification 6.4.2.3 Identification of risk sources 6.4.2.4 Identification of potential events and outcomes 6.4.2.6 Identification of consequences 6.4.3 Risk Analysis 6.7 Recording and reporting
MEASURE: Identified risks are assessed, analyzed, or tracked.	5.7 Improvement (references ISO 31000:2018) 6.3.4 Defining risk criteria 6.4.2.5 Identification of controls 6.4.3 Risk Analysis 6.4.3.2 Assessment of consequences 6.4.3.3 Assessment of likelihood 6.6 Monitoring and review (references ISO 31000:2018)

Continued on next page

AI RMF ROADMAP



NIST TRUSTWORTHY AI RESOURCE CENTER

AI RMF PLAYBOOK

AI RMF PROFILES

AI RISK GLOSSARY

AI METRICS HUB

…AND MORE

AI INSTITUTE
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 FOUNDED A.D. 1865

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

Generative AI Public Working Group

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

THANK YOU

www.nist.gov/artificial-intelligence

airc.nist.gov

AI RISK MANAGEMENT PLAYBOOK

DOE/NRC

AI/ML Meeting

September 13, 2023

Ms. Jonnie Bradley, Sr. Program Manager/
Responsible AI Official, Artificial Intelligence
and Technology Office

U.S. DEPARTMENT OF ENERGY

AGENDA

- Overview
- AI RMP Features
- Pathways
- Future Developments
- Walkthrough

U.S. DEPARTMENT OF ENERGY

DELIVERING ON AITO'S MISSION AND VISION

1. **From** AI inventory data tracking **to** AI Portfolio and Program **Optimization** & **Impacts**
   - Strategic portfolio analysis and alignment of AI/ML investments, ensuring alignment with national security priorities, facilitates department-wide responsible and trustworthy AI across program offices (ex**. Federated Learning)**
2. **From** management of AI projects **to** orchestration of AI **Strategy** and **Partnership** Development
   - Builds robust AI partnerships and customer excellence across internal, external, and international boundaries and addresses strategic communications for the Department.

U.S. DEPARTMENT OF ENERGY

DELIVERING ON AITO'S MISSION AND VISION

Visionaries and practitioners of Modern-Day Governance for AI and Autonomous Innovations

- Responsible AI Official
- Convene
- Coordinate
- Facilitate
- Orchestrate
- Advocate
- Integrate
- Risk Management

U.S. DEPARTMENT OF ENERGY

OVERVIEW: AI Integrated Ecosystem

AIX System

Enables better coordination across DOE programs and alignment with strategic priorities in AI

AI RMP

Guidance for leaders of new AI/ML projects and a reference of risks and mitigation techniques for technical users

AI Pathways

A space for the DOE AI community to post questions, share knowledge, and connect with SMEs

AI Strategy

Certificate upskilling program with two tracks: Management and Technical

AI Advancement Council

A space to test and validate algorithms and models using the latest techniques

DOE AI Community Hub

U.S. DEPARTMENT OF ENERGY

Active or underway

In planning stages


AI RISK
MANAGEMENT
PLAYBOOK
U.S. DEPARTMENT OF
ENERGY

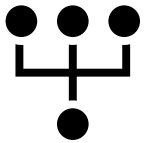
AI RMP VALUE

The **AI Risk Management Playbook** (AI RMP) is a comprehensive reference guide for AI risk identification with recommended mitigations to support responsible and trustworthy (R&T) AI use and development.

U.S. DEPARTMENT OF ENERGY

TIMELINE: AI RISK MANAGEMENT PLAYBOOK

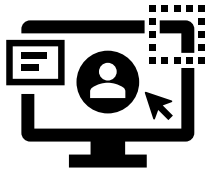
April 2020 – AI Asset Risk Management Framework



2021 – DOE review, edit, comment, & concurrence, and intergovernmental and industry review/comment

September 2022 – AI RMP Public Q&A

August 2020 – Internal AI Risk Management Playbook (AI Ethics, Data Compliance, Asset Management, Secure Model Usage)



August 2022 – Public Release – energy.gov/ai/airmp

U.S. DEPARTMENT OF ENERGY

AI RMP: FEATURES

Essential Guidance

Dynamic system featuring 141 unique risks and mitigation techniques with ability to easily and continuously expand

Intelligent Search

Ability to filter according to lifecycle stage, assets, as well as mapping to project roles and direct keyword searching

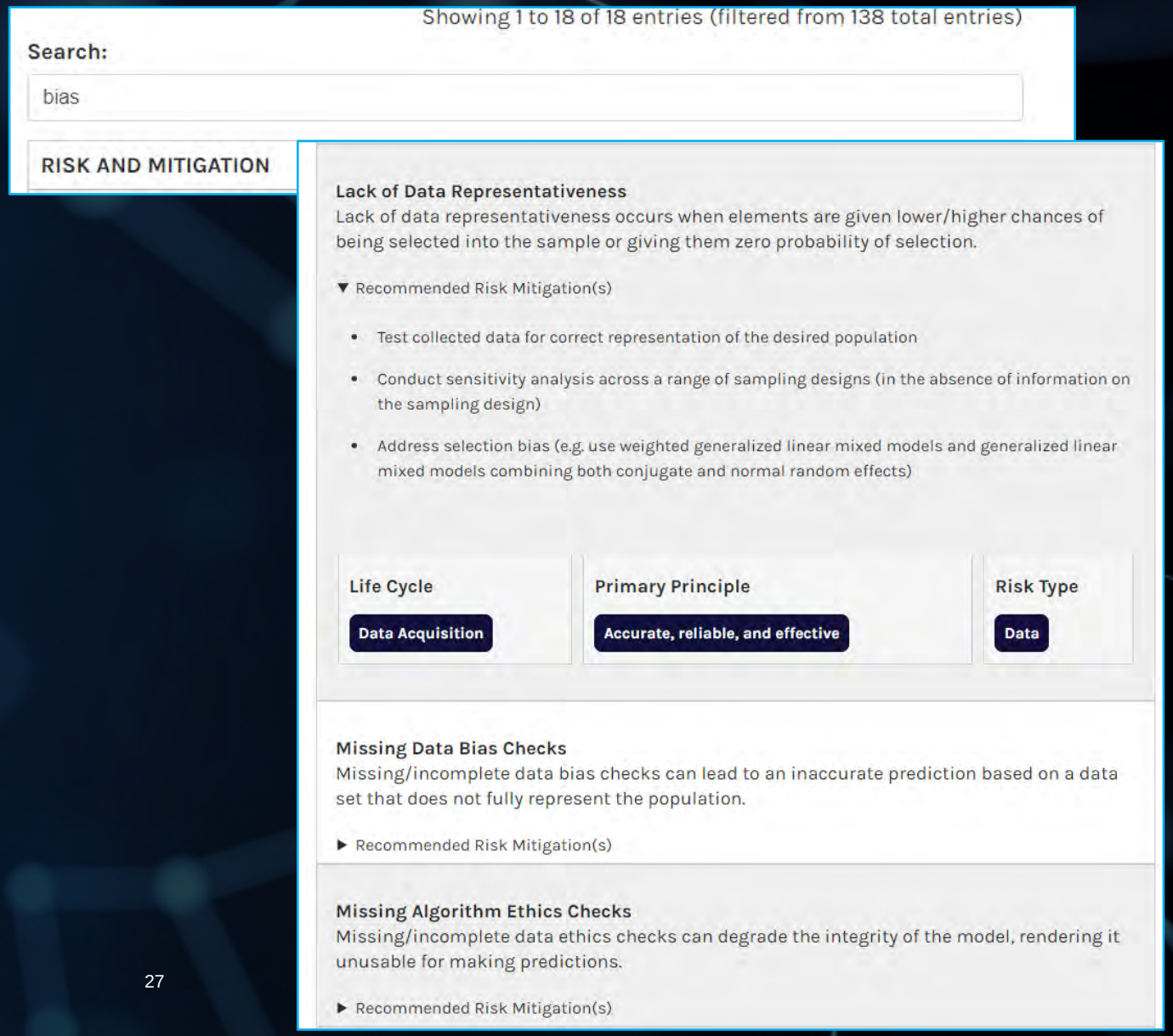
Trustworthy AI

Integration with EO 13960: Promoting the Use of Trustworthy AI, including ability to filter by principle

U.S. DEPARTMENT OF ENERGY

USING AI RMP

AI Use Case	Example Risk
Using data for AI systems to determine peak charging hours for EV owners	Unintended biases in datasets used to build AI systems; disadvantaged communities are not included in the datasets
Using data for AI systems to determine which neighborhoods to install EV supercharging stations	
Using data for AI systems to predict power outages and which communities will be serviced after an outage	



AI RMP: PATHWAYS

MANAGEMENT

- Additional filters designed for project managers and AI novices
- Provides understanding of risks according to development stage
- Connects risks with Trustworthy AI principles

TECHNICAL

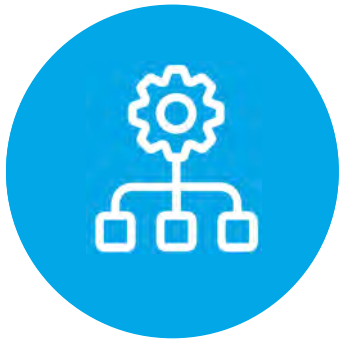
- Ability to directly search via keyword according to issues encountered and mitigation techniques
- Quickest pathway for technical experts to find relevant risk or mitigation recommendation

U.S. DEPARTMENT OF ENERGY

AI RMP: FUTURE DEVELOPMENT

Additional Content

Continuously add new risks and mitigation techniques; Eventual integration with NIST's AI Risk Framework



Enhanced Searching

New search capabilities co-developed with playbook users including the ability to search by project role

Increased Engagement

Additional features to drive community engagement including contributor badges and enhanced editing capabilities

U.S. DEPARTMENT OF ENERGY

Q&A/ Discussion

U.S. DEPARTMENT OF ENERGY

ADDITIONAL RESOURCES

- Executive Order 13960 - Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government
- NIST AI Technical Standards
- Blueprint for an AI Bill of Rights
- National Artificial Intelligence Initiative
- Algorithmic Discrimination Protections
- AI Now Institute's Annual Reports
- Partnership on AI
- Alan Turing Institute's Fairness, Transparency, Privacy group
- Harvard's "Embedded Ethics" Modules
- DAIR Institute
- Fairness Tutorial

U.S. DEPARTMENT OF ENERGY

THANK YOU

Artificial Intelligence and Technology Office
U.S. Department of Energy
September 13, 2023

Website: energy.gov/ai
AIRMP: energy.gov/ai/airmp

Inquiries, comments may be sent to:
doeaimailbox@hq.doe.gov

U.S. DEPARTMENT OF ENERGY

DOE DRAFT RESPONSIBLE AND TRUSTWORTHY AI PRINCIPLES

DOE AI Principles Draft
Equitable
Traceable
Reliable
Governable
Accountable

U.S. DEPARTMENT OF ENERGY

NRC Standards Forum

LWRS LIGHT WATER REACTOR SUSTAINABILITY

Ahmad Al Rashdan, INL

Considerations for the Development of a Standards-compatible AI

Acknowledgment: Ted Quinn and Roman Shaffer

September 13, 2023

Ahmad Al Rashdan, Ph.D.

Senior Research and Development Scientist

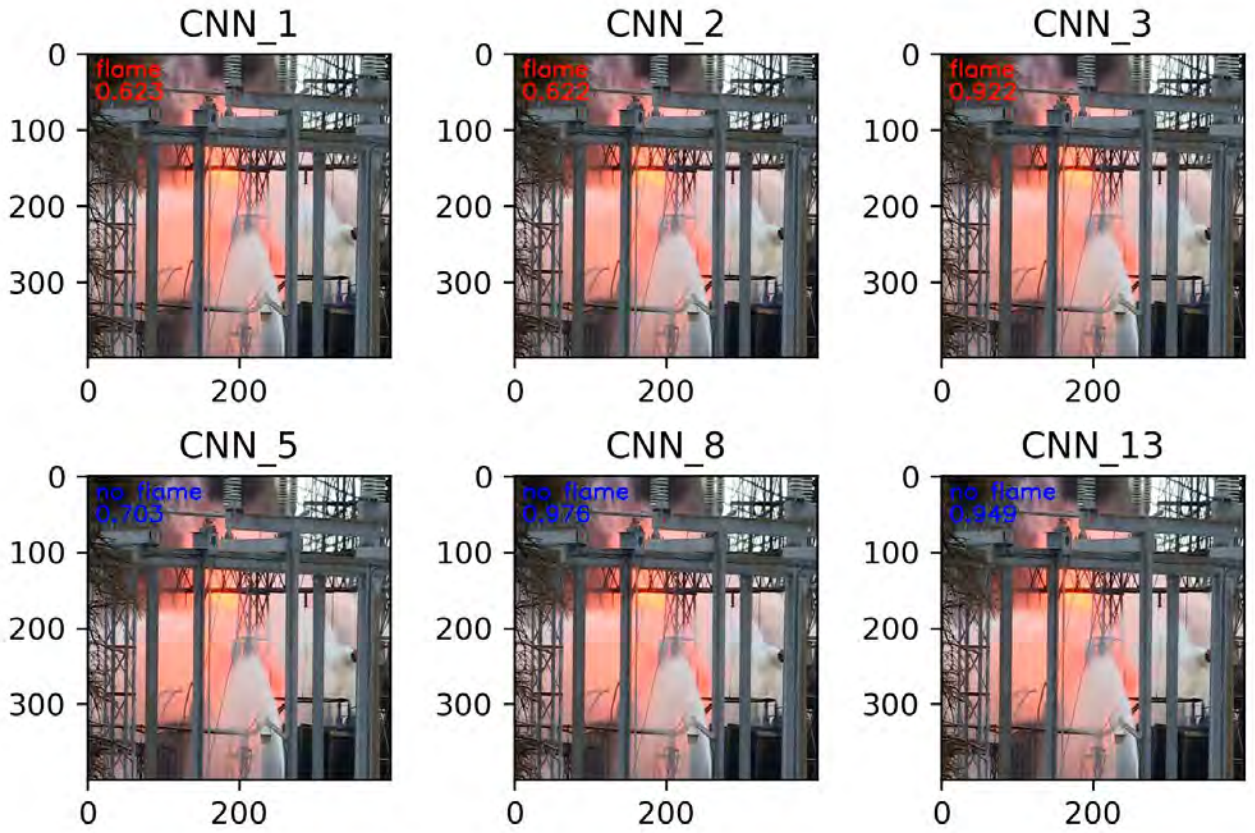
Idaho National Laboratory, USA

- E-mail = ahmad.alrashdan@inl.gov
- Phone = +1 979 422 4264

- A senior scientist with 17 years of industrial and research experience in automation, artificial intelligence/machine learning (AI/ML), and instrumentation and controls (I&C).
- Lead of several efforts under the U.S. Department of Energy (DOE) focusing on plant modernization using AI methods and advanced analytics.
- Author of more than 80 reports and journal papers, 9 patent applications and awards, and developer/co-developer of 8 Copyrighted software packages.
- Recipient of several funding and recognition awards including a recent 2022 R&D 100 and the prestigious INL's Director Excellent Engineering Achievement Award.
- Holder of several professional leadership positions
- Standards involvement:
  - Member of the IEC SC45A WG12 to create a standard for AI Application for Nuclear Installations.
  - Member of the ANS Large Light Water Reactor Consensus Standards Committee.
  - Vice Chair to the ANS Simulators, Instrumentation, Control Systems, Software & Testing standards Subcommittee.

Computer Vision Machine Learning (CVML) in Fire Watch

- Automatically identifying a fire in a video stream to eliminate/reduce the need for fire watches.
- Sixteen different models were evaluated.
- An ensemble of models is developed to improve accuracy (>99% achieved)
- Smoke is being integrated. Temporal effects are being considered to eliminate false positives (mist, fog, steam, etc.)



CVML in Gauge Reading

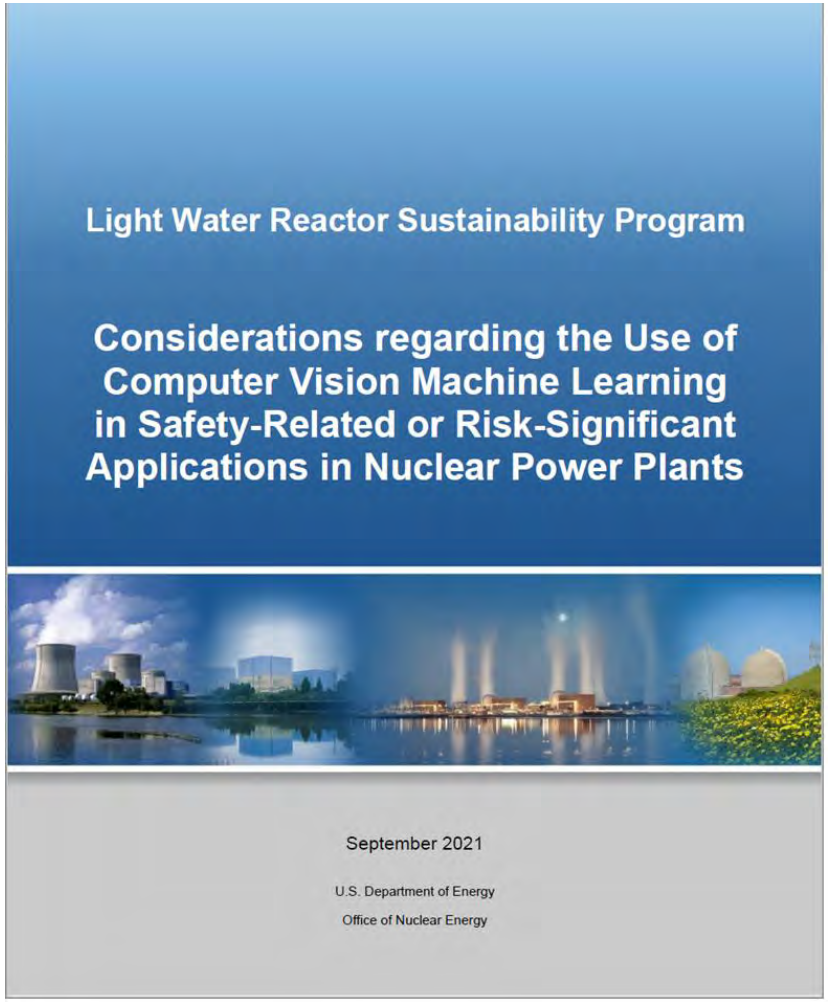
- Automating manual logging of analog gauges (i.e., a method to recognize gauges in oblique angles and read their values)

Automated gauge reading impacts a wide spectrum of activities in a plant including operator rounds, gauges calibration, and peer verification, and improves data fidelity for online monitoring.

The Need for Compatible AI and What Does that Mean?

- Digital Instrumentation And Controls (DI&C) regulatory requirements would need to be satisfied for any and every AI application that impacts safety related and risk-significant applications
- Regulations frequently cite standards.
  - Can AI be customized to meet those standards requirements?
  - Do we need new AI-focused standards?

"To evaluate how example AI technologies align with the safety framework, and discusses how they could be analyzed, modeled, tested, and validated in a manner similar to typical DI&C technologies."



Example of Considerations

CVML models often utilize open-source datasets and feature extraction engines or models.

It is not always possible to determine the level of overlap among open-source datasets. Open-source models could use similar fundamental concepts. This impacts the independence of the developed CVML models:

- Causes the CVML system to be susceptible to common cause failure
- Overestimates the software verification results
- Introduces a cybersecurity concern

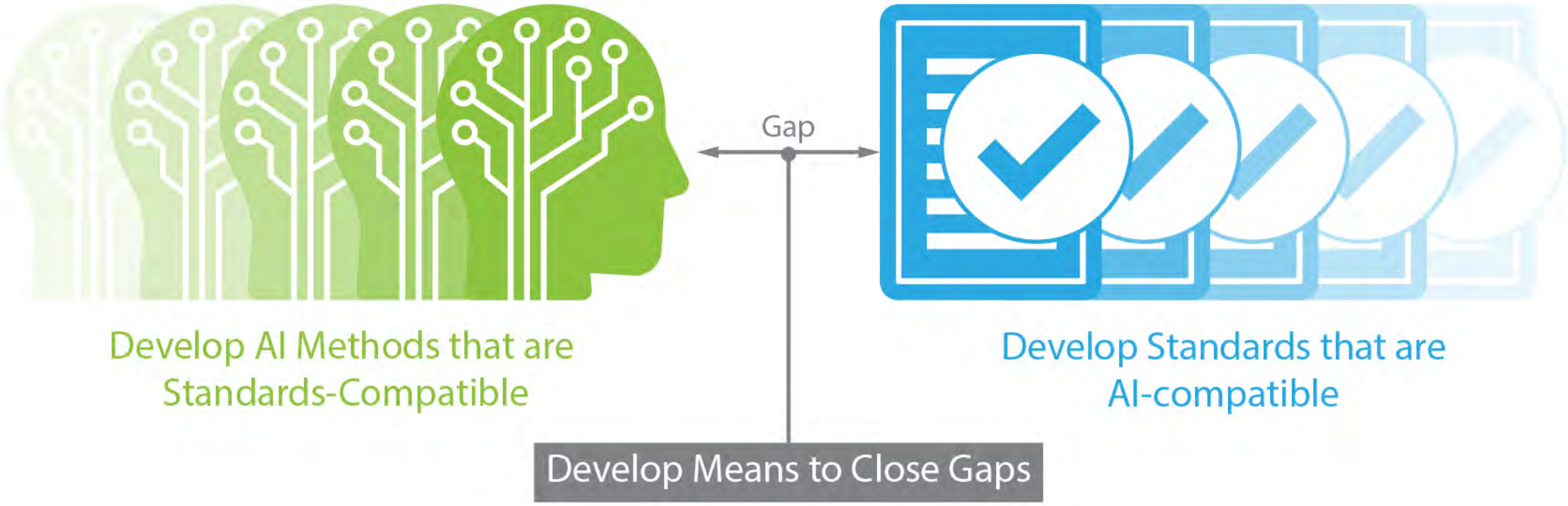
Methods to create independent datasets may be needed (e.g., GANs)

Methods to quantify independence may be needed

CVML Compatibility with the Current Requirements

Characteristic/Consideration	Independence	Defense in Depth	CCF	V&V	QA	Configuration Control	Cyber Security	CGD	Maintainability	Traceability	Design Control	Repeatability	Deterministic Nature	Explinability	Reliability	FMEA	Simplicity	Justification	Trustworthiness
Open-source data and model	■	■	■	■	■														
Frequent updates to source						■	■	■	■										
Massive amounts of data					■	■		■		■	■					■			
Periodic training						■	■		■		■								■
Probabilistic and stochastic				■								■	■			■			■
Various performance metrics				■											■	■			
Incomprehensible to reviewers				■										■			■		■
Inherited bias	■		■	■															
Non-systematic approach				■								■						■	
Robustness to new conditions			■									■	■		■	■			■
Special skillset						■			■	■				■		■			

Conclusions



Sustaining National Nuclear Assets

lwrs.inl.gov

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

IEC/SC45A/WGA12

Artificial Intelligence for Nuclear Facilities

Trey Hathaway, Ph.D.
Reactor Systems Engineer
U.S. Nuclear Regulatory Commission
RES/DSA/AAB

2023 NRC Standards Forum

September 13, 2023

AI Standardization Efforts

- ISO/IEC JTC 1/SC 42 – Artificial Intelligence (AI)
  - Created in 2017
  - Secretariat: ANSI
  - Scope: Standardization encompassing Artificial Intelligence
    - Provide guidance to JTC 1, IEC, and ISO committees developing artificial intelligence applications
- IEC – International Electrotechnical Commission
  - IEC/SC 45A – Instrumentation, control and electric power systems of nuclear facilities
    - Scope:

"Prepare standards applicable to electronic and electrical functions and associated systems and equipment used in nuclear energy generation facilities to improve efficiency, safety and security of nuclear energy generation."

WGA 12

- WGA 12 – AI for nuclear facilities
  - First meeting held on August 22, 2023
- Composed of a multi-disciplinary international team
  - Currently 35 experts
    - NRC has four staff involved
      - Ismael Garcia, Kim Lawson-Jenkins, Tanvir Siddiky – NSIR
      - Trey Hathaway - RES

WGA 12

- Tasks
  - Develop and maintain standards and reports for AI applications in nuclear facilities
  - Provide guidance to stakeholders developing, deploying, and overseeing AI applications for nuclear facilities
  - Cover fundamental characteristics of AI of nuclear facility applications
  - Applicable to the entire nuclear facility life cycle

WGA 12

- Provide overview of AI from a nuclear perspective
  - Discuss concepts, applications, and challenges
- Explore definition of AI for nuclear facility applications
- Discuss AI applications with focus on IEC/SC45A cross-cutting areas
  - Instrumentation and control
  - Annex to discuss applications of AI to other aspects of nuclear facility uses beyond IEC/SC45A scope

WGA 12

Cross-cutting Topics with other Working Groups in SC45A

Working Group	Working Group Titles
WGA2	Sensors and measurement techniques
WGA3	Instrumentation and control systems: architecture and system specific aspects
WGA5	Special process measurements and radiation monitoring
WGA7	Functional and safety fundamentals of instrumentations, control and electrical power systems
WGA8	Control rooms
WGA9	System performance and robustness toward external stress stems
WGA10	Ageing management of instrumentation, control and electric power systems in NPP
WGA11	Electrical power systems: architecture and system specific aspects

WGA 12

- Three levels of documents to be produced
  - Level 2 document on cross-cutting areas
    - General Requirements
    - “Horizontal Standard”
      - e.g., Trustworthiness & Risk Management, Data Processing & Management, Testing and V&V, Performance Assessment
  - Level 3 document on AI specific applications in nuclear facilities
    - “Vertical Standard”
    - Standards for specific applications
      - e.g., Virtual sensors, anomaly detection
  - Technical reports that support Level 2 and Level 3 standards
- Documents created through IAEA Consultancy meetings will be leveraged in the IEC work

Status and Next Steps

- Participants will notify working group convener of areas where they would like to offer input – by September 9, 2023
- Meetings with other working groups at IEC General Meeting in October 2023 to discuss cross-cutting areas
- Develop the Level 2 Standards first, then begin to explore application specific standards

AI-Assisted Ultrasonic Inspections in the Nuclear Power Industry

Thiago Seuaciuc-Osorio
Principal Technical Leader

2023 NRC Standards Forum
September 13, 2023

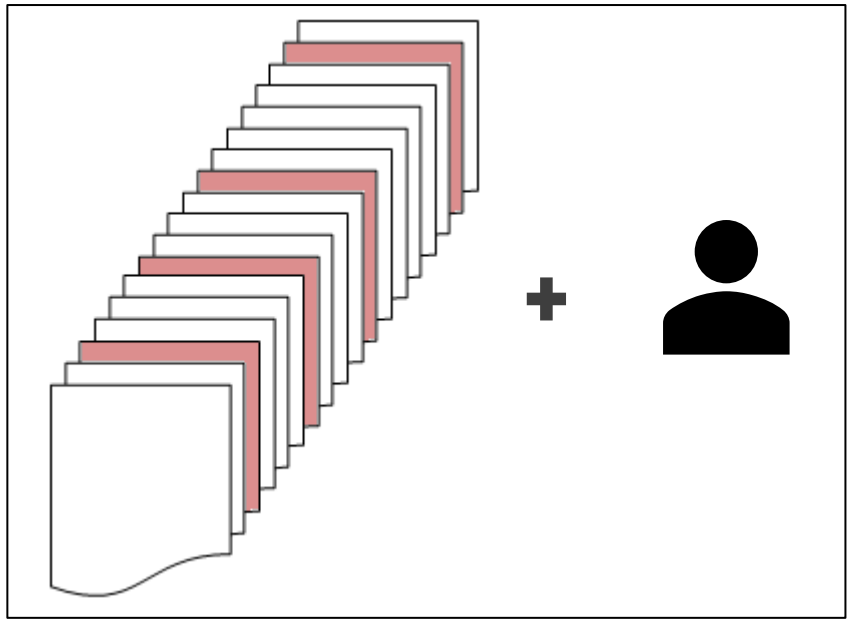
AI-Assisted Analysis of UT Inspections

- UT inspections are an important part of the scope of an NDE program
- Some inspections are challenging or have large volumes of data
- Machine learning tools can potentially **assist** in the analysis of the data
  - Increase reliability
  - Decrease analysis time
- *Assist* means AI flags regions for review: final decisions still rest with the qualified inspector.

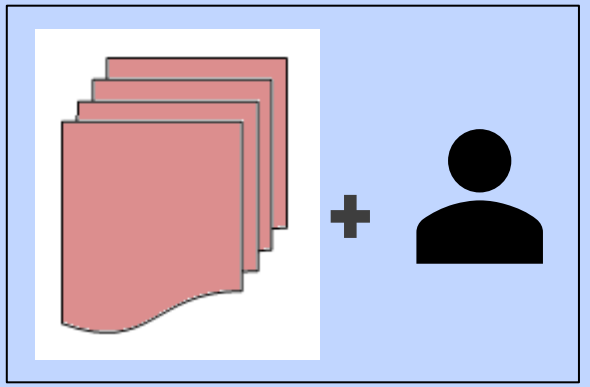
***Goal:* Develop auto-analysis tools to assist in UT inspections**

How Would AI Assist in UT Inspections?

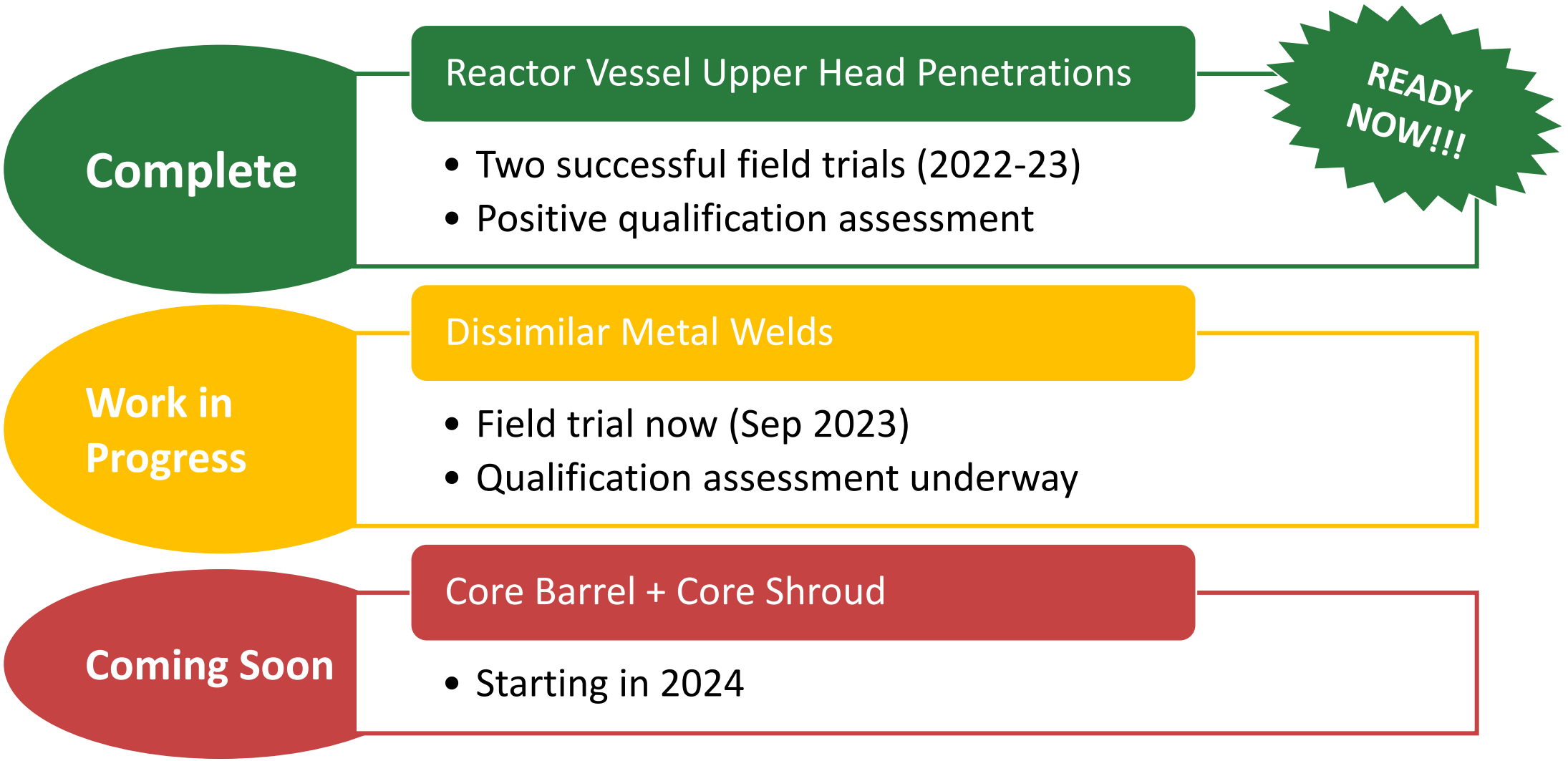
- Current Inspection
  - Examiners distribute their energy across a high volume of (mostly benign) data



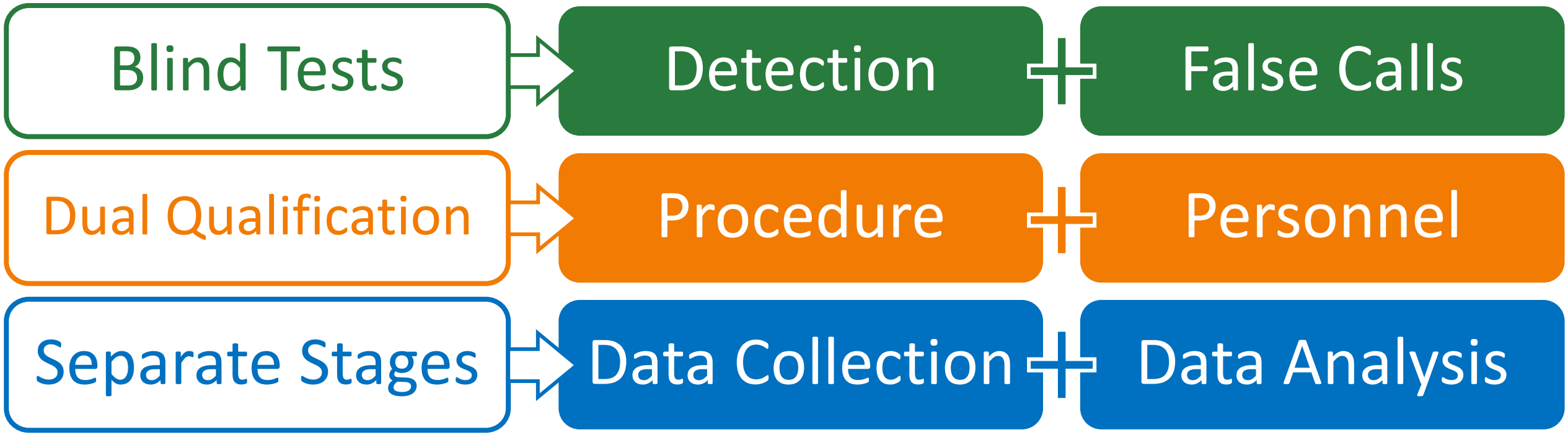
- AI Assisted Inspection
  - Examiners focus their energy on the regions that require more careful review, while AI takes care of the more monotonous portion



Development Status



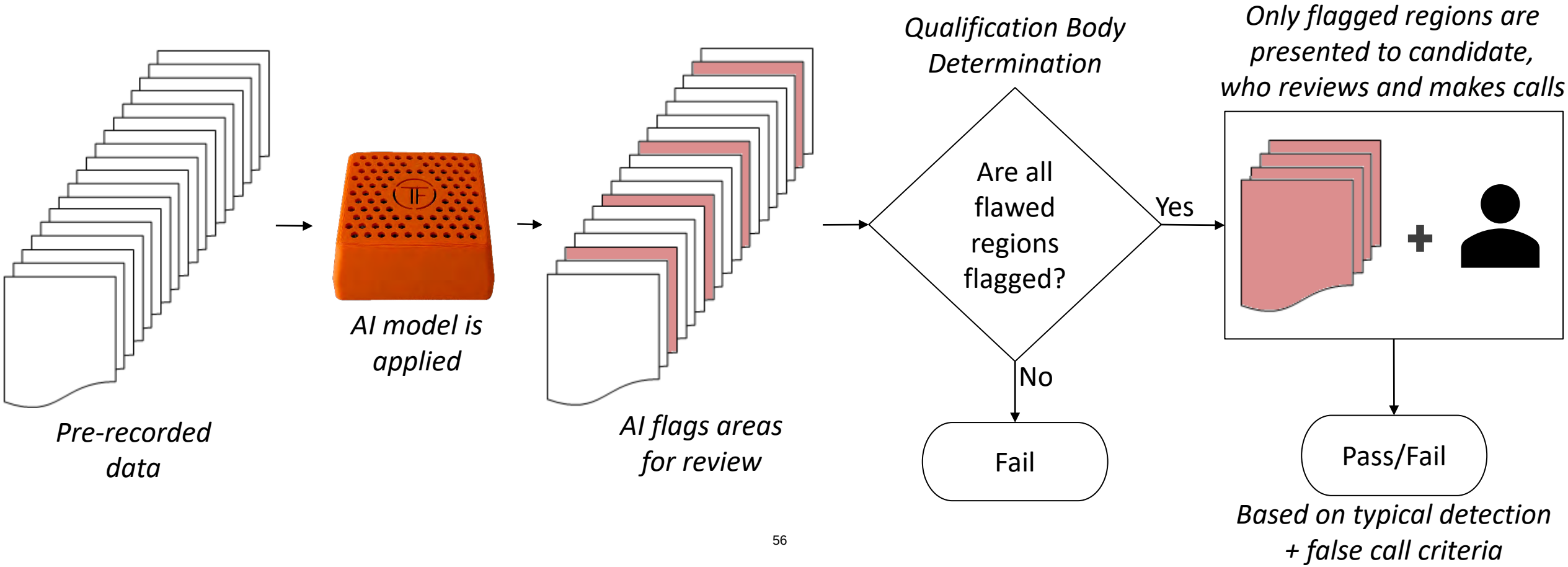
Current Qualification Framework



Develop AI-Assisted Analysis that fits this overall framework

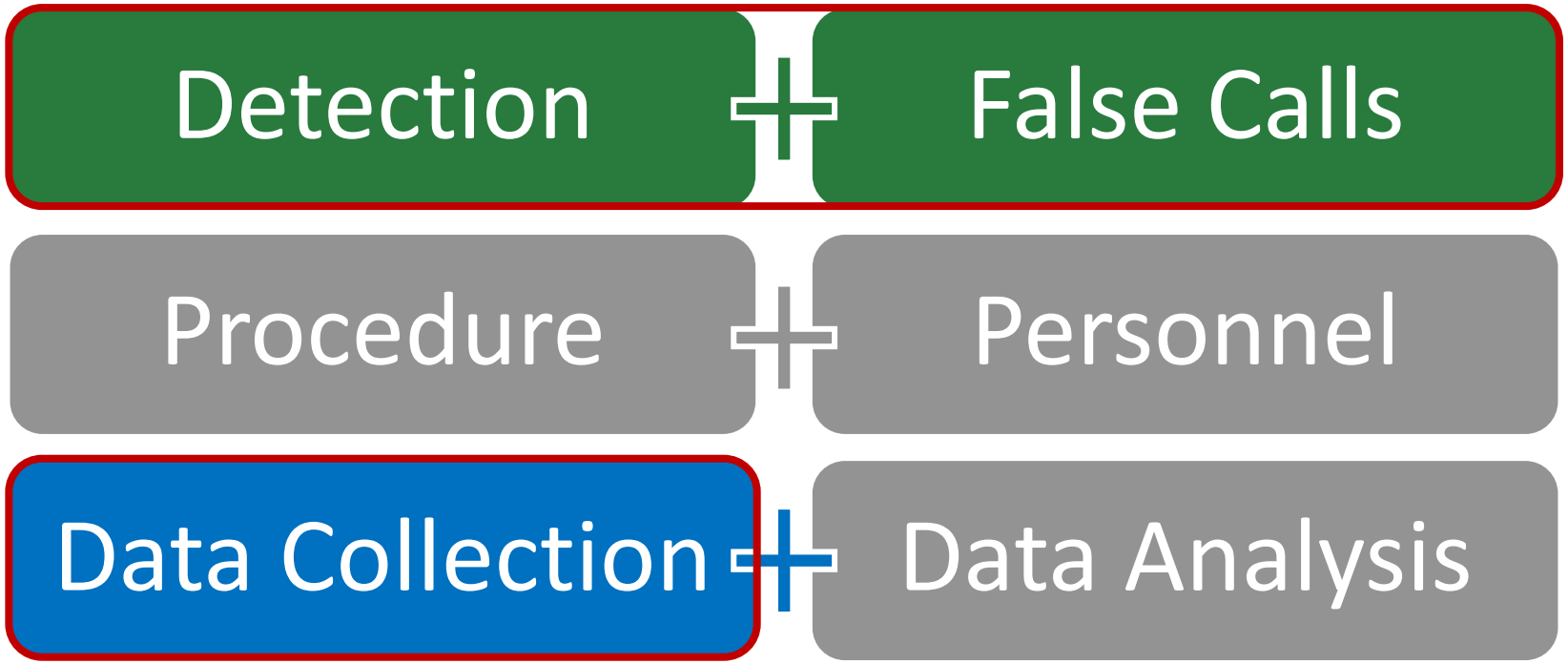
What initial qualification *may* look like

- Procedure would be updated to include a defined process for the AI evaluation
- AI algorithms for qualification would be developed and provided to the qualification body: model is **frozen** at this stage



Elements that do *NOT* Change

Same blind tests with same criteria



Same data enables:

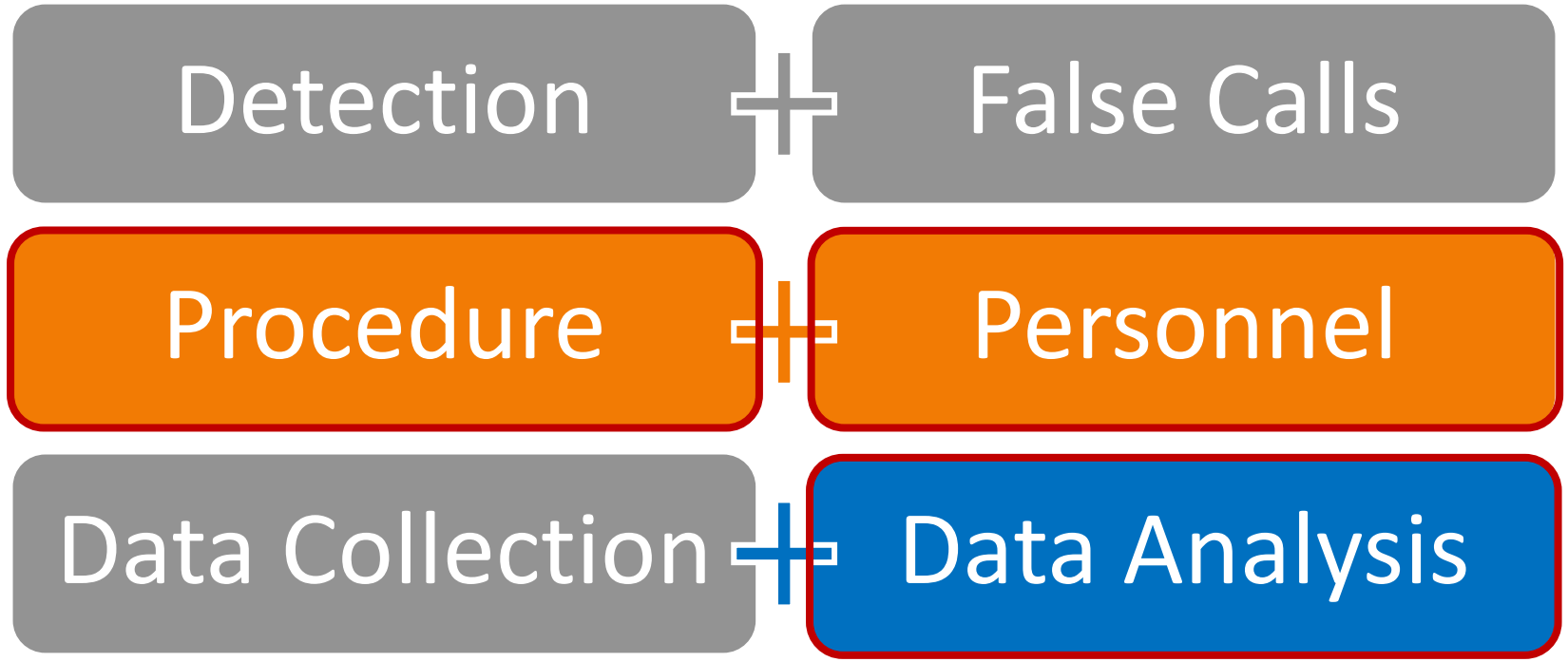
- Ease of implementation
- Parallel deployment

Qualification framework is maintained

Elements that *DO* Change

Procedure adds AI analysis stage:

- **Freezes** model
- 100% detection required



Personnel only reviews flagged regions:

- Same test
- More restrictive

- Initial screening potentially on tool
- Further analysis remain the same

Modifications are typical of any new procedure

Going Forward

Ready for Use

- Utilities can use for oversight & planning
- Vendors can seek qualification

Parallel & gradual deployment:
AI-Assisted as one of the two required independent reviews
- *Gain operational experience*
- *Different failure modes*

Supporting Research (2024 onwards)

- Assessment of reliability of AI-assisted analysis
  - UT, VT
- Qualification protocol
  - Re-qualifications

60