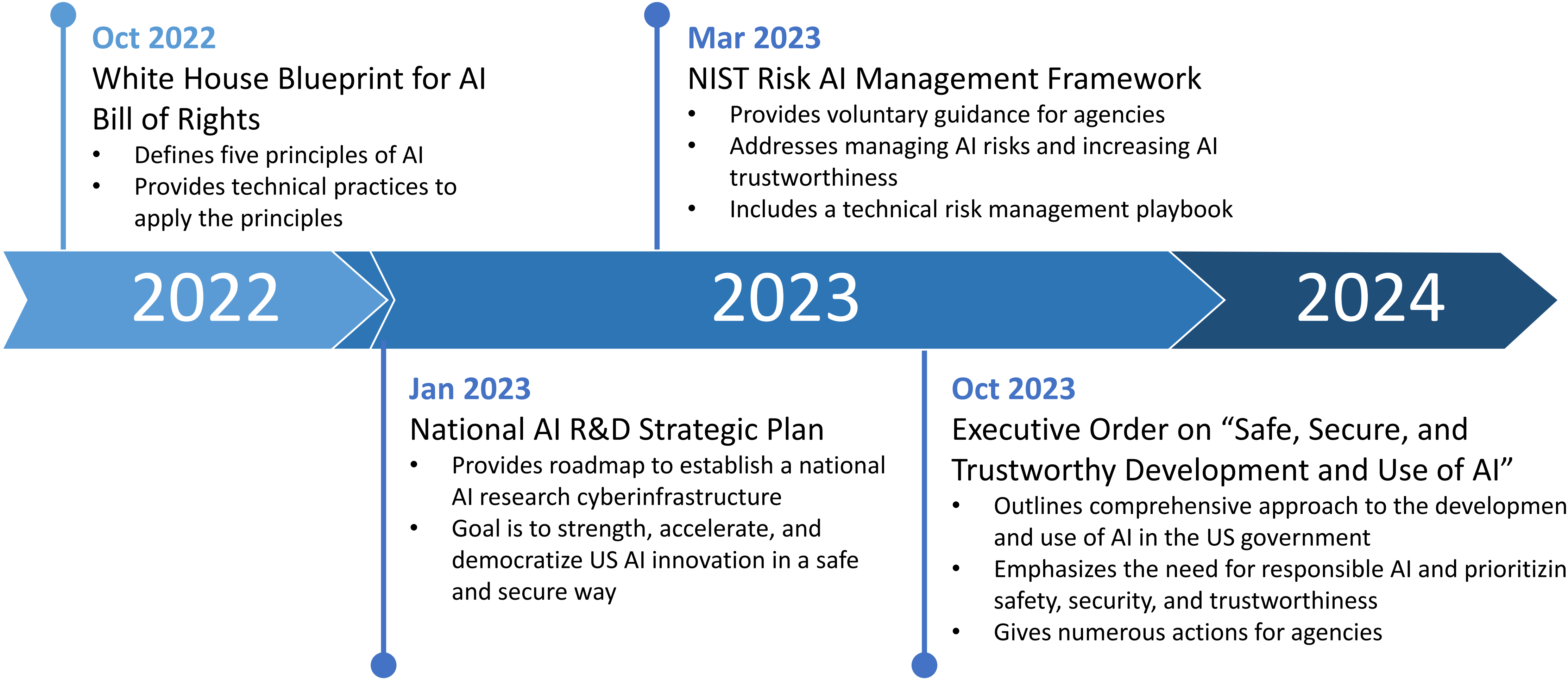
New ATM Requirement: Certification Framework for AI Technology

Spring REDAC: March 19, 2024

Outline

- Background
- Framework Overview

US Government Leadership on AI



Oct 2022

White House Blueprint for AI Bill of Rights

- Defines five principles of AI
- Provides technical practices to apply the principles

Jan 2023

National AI R&D Strategic Plan

- Provides roadmap to establish a national AI research cyberinfrastructure
- Goal is to strength, accelerate, and democratize US AI innovation in a safe and secure way

Mar 2023

NIST Risk AI Management Framework

- Provides voluntary guidance for agencies
- Addresses managing AI risks and increasing AI trustworthiness
- Includes a technical risk management playbook

Oct 2023

Executive Order on “Safe, Secure, and Trustworthy Development and Use of AI”

- Outlines comprehensive approach to the development and use of AI in the US government
- Emphasizes the need for responsible AI and prioritizing safety, security, and trustworthiness
- Gives numerous actions for agencies

AI and ML Research within NextGen

- The FAA's NAS 2040 envisions a more agile and dynamic environment capable of seamlessly adapting to the changing demands of the NAS
- To support this vision, the FAA is exploring the use of Artificial Intelligence (AI) and Machine Learning (ML) technologies to help enhance decision support function, processes, and other capabilities in support of Air Traffic Management (ATM) Operations
- AI/ML models can identify trends/patterns in existing data, predicting airspace operational behaviors and analyzing complex airspace situations

Techniques Investigated by AI/ML Areas

Natural Language Processing

- Embedding Techniques
  - TF-IDF, Word2Vec, GloVe, Doc2Vec
- Transformers
  - BERT, RoBERTa
- Regular Expressions

ML / Deep Learning

- Support Vector Machines
- Random Forest
- Reinforcement Learning
- Deep Q-Learning
- Conservative Q-Learning
- Long Short-Term Memory Networks

Speech Recognition

- Open AI Whisper Models
- Microsoft Azure speech studio
- Google Speech to Text (STT) models

Concerns Around the Use of AI/ML

- AI/ML isn’t always a suitable solution. **AI/ML suitability requires**
  - Learnable patterns
  - Data availability
  - Problems with a repetitive and predictive nature
  - Problems where wrong predictions are cheap (i.e., not catastrophic)
  - Patterns shared between training data and live data
- Even when AI/ML is suitable, **to develop and deploy these systems responsibility** within the **human-based safety-first environment of FAA**, teams need to consider
  - Reliability and robustness
  - Explainability and transparency
  - Data quality and bias
  - Trust, overreliance, and automation bias
  - Continuous monitoring and maintenance

Safe and Secure Use of AI

- To ensure the responsible implementation of AI/ML into the NAS, ongoing efforts in this area include:

AI Certification Framework

- Developing systematic processes to certify AI technologies by considering the entire AI development life cycle

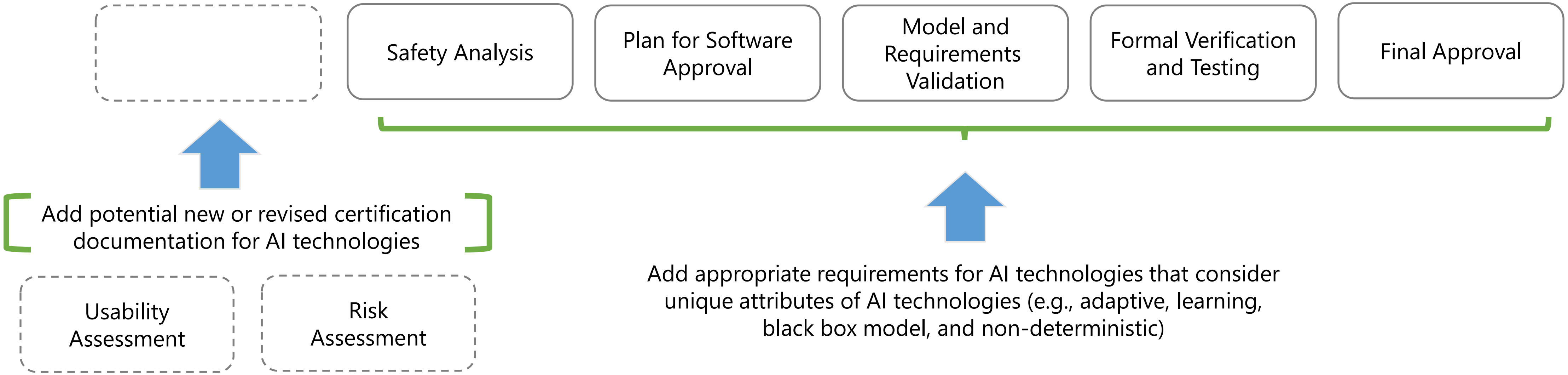
NASA Aeronautical Research Institute (NARI) - FAA Collaboration

- Exploring AI problems, use cases, and prototypes for the NAS

AI Impacts on Traditional Software Certification

- The traditional software assurance processes used by the FAA will require modification to appropriately validate and verify AI software:
  - New checkpoints or documentation
  - Amendments to existing requirements

Notional review process to assess software assurance for traditional software:



Safety Analysis

Plan for Software Approval

Model and Requirements Validation

Formal Verification and Testing

Final Approval

Add potential new or revised certification documentation for AI technologies

Usability Assessment

Risk Assessment

Add appropriate requirements for AI technologies that consider unique attributes of AI technologies (e.g., adaptive, learning, black box model, and non-deterministic)

Regulatory Gaps for AI Certification

- Current industry standards are more suited for traditional software development and do not provide sufficient consideration of AI
  - These standards must be amended to address unique technical aspects of AI
- FAA specifies processes for new, COTS, and previously certified software development
  - The certification process for new development of AI technologies should be considered within the current FAA (e.g., AMS) software lifecycle
  - Due to the learning aspects of some ML that inherently alter the established algorithm based on implementation context, the criteria to certify COTS and reused AI software may be different from traditional software

Industry Certification Standards

- RTCA DO-178C: *Software Considerations in Airborne Systems and Equipment Certification*
- RTCA DO-278A: *Software Integrity Assurance Considerations for Communication, Navigation, Surveillance and Air Traffic Management (CNS/ATM) Systems*
- A EUROCAE and SAE working group is currently assessing considerations for AI use, development, and certification in aviation

FAA Certification Standards

- Allows RTCA DO-178C for software assurance for aircraft systems and equipment
- Software assurance specified in ATO's *Safety Risk Management Guidance for System Acquisitions* is based on RTCA DO-278A

Technical Gaps for AI Certification

- Unique attributes of AI technologies inhibit the use of traditional software assurance processes

Use of Learning Techniques

- Training and testing datasets are unique to ML techniques
- Requirements and review for these datasets and the learning process are not covered by traditional software assurance

Validating Requirements

- Traditional traceability requirements may be difficult to fulfill for AI (e.g., lack of clear traceability between requirements and data or between algorithm components and outputs)
- In particular, it is difficult to establish traceability to low-level requirements for ML due to the use of learning techniques

Verifying Results

- Comprehensive testing of all possible ranges of inputs and outputs for AI may be difficult
- Repeatable testing requirements may only be possible for locked A techniques
- Traditional code and model testing requirements may not be suitable for AI
- Current practices do not support the probabilistic nature of AI

AI Certification Framework Assumptions

01 The framework focuses on the certification of new and emerging technologies under consideration by the FAA, and, in particular, on AI-based technologies.

02 The framework will consider the certification of any type of AI-related software and service within the FAA supporting aircraft, ATM, and uncrewed components.

03 The framework will consider the certification of software for all systems and service classifications and the “system of systems” aspect of AI implementation, including but not limited to human factors, system design/architecture, system integration, safety, and life-cycle management.

04 The framework will assume the certification of hardware components is addressed by existing certification requirements.

05 The framework will consider the certification of software under the following pathways:

a) New FAA-led development of AI software from start to finish,
b) COTS AI software received in near-final form,
c) Modified COTS AI software received and altered for use, and
d) Previously certified AI software now used in another system or being updated in the same system.

06 The framework will consider the integration of any new or modified certification processes within the context of current FAA software approval processes (e.g., Verification and Validation (V&V) and Test and Evaluation (T&E)) and safety assurance expectations.

07 The framework will maintain that all AI technologies must meet the certification requirements deemed applicable by the civil authority; the certification pathway and level of rigor will differ depending on intended use, criticality, risk, and other factors.

08 The framework will evaluate the requirements of data management, particularly for separation of training data from testing data and from data used for certification compliance test case demonstrations.

Framework Overview

Need for a Flexible Framework

- The certification process to support a range of projects:

Maturity Levels

Spectrum between new, ongoing, and close to completion

COTS vs New

Spectrum between COTs, modified COTs, or new development

Impact Criticalities

Spectrum between low, medium, and high criticalities

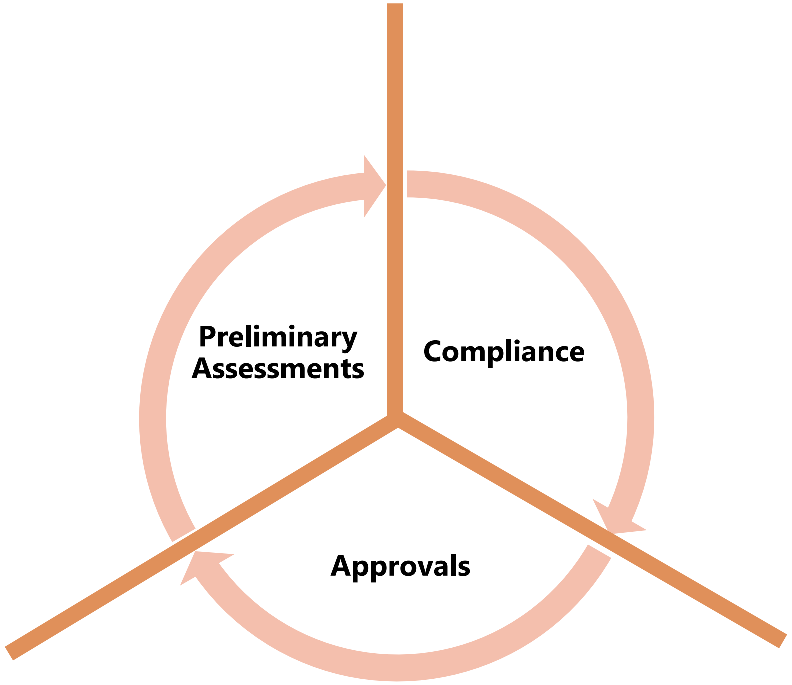
AI Technologies

Any number of technologies such as natural language processing, classification, imaging, etc.

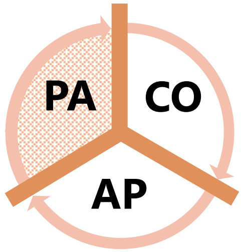
Certification Methodology

- There are three main components in the certification methodology.

	Phase	Purpose	Outcome
01	Preliminary Assessments	Evaluate the initial usability, risk, response, and safety aspects of the AI technology	• Level of certification rigor • Certification pathway
02	Compliance	Define and compile all documentation necessary for approval	• Compliance documentation
03	Approvals	Coordinate with stakeholders to review and approve documentation	• Certification decision



ICN



Preliminary Assessments Phase

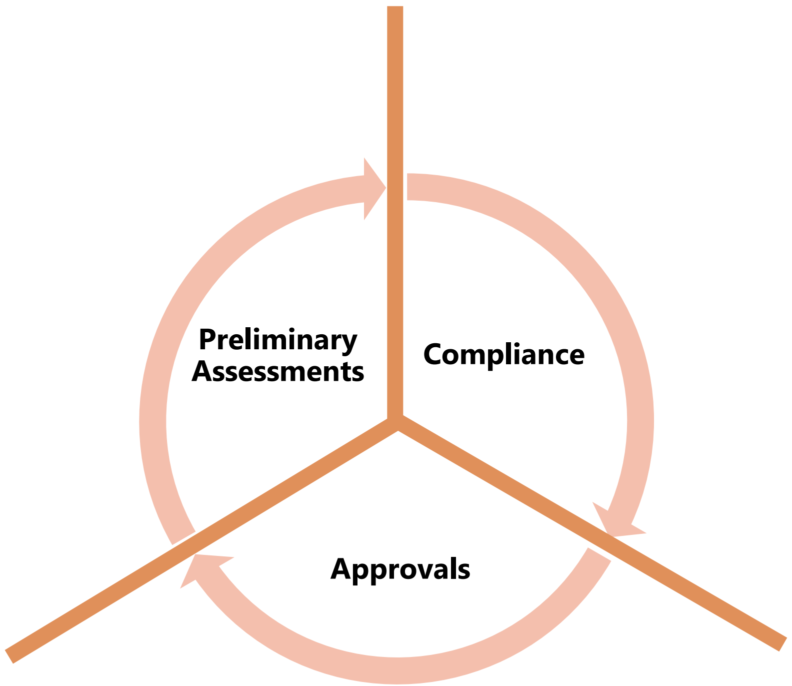
	Phase	Purpose	Outcome
01	Preliminary Assessments	Evaluate the initial usability, risk, response, and safety aspects of the AI technology	• Level of certification rigor • Certification pathway

AI Usability Assessment: Evaluate the use of AI technology for this problem based on FAA policy, cost-benefit, and conceptual and technical suitability

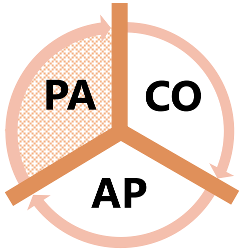
Risk/Response Assessment: Evaluate the AI technology based on the AI risk factors, roles, and function that will affect level of certification rigor

Safety Assessment: Evaluate the safety of the system and corresponding assurance level

Pathway Determination: Certifying Authority approves the pathway and level of rigor required and tailored for certification of the AI technology



AI Usability Assessment



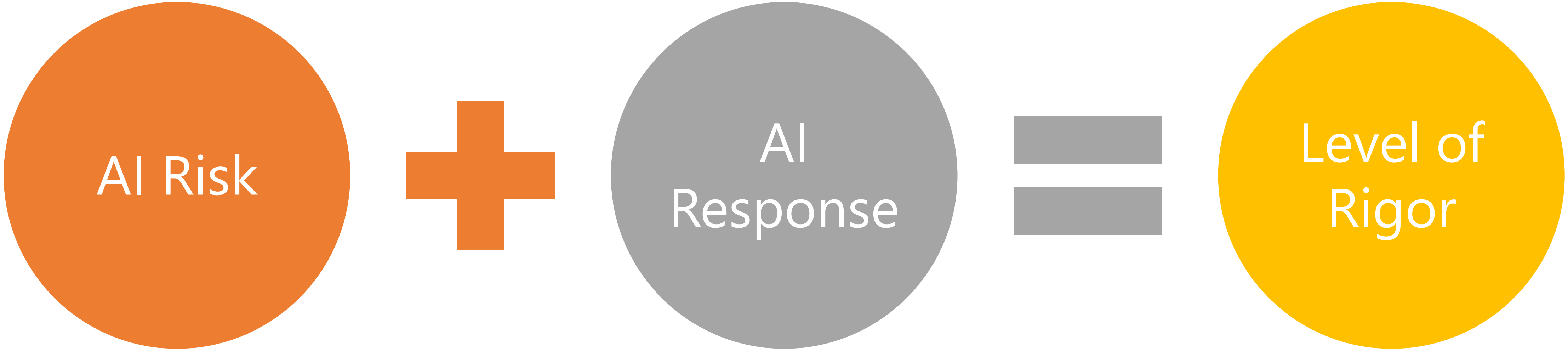
- Decision tree of Yes/No questions that cover the following areas:
  - FAA Policy
    - *E.g., Does the proposed technology comply with FAA 1370.121A guidance?*
  - Cost-Benefit
    - *E.g., Does the proposed technology provide added value over the current approach?*
  - Conceptual Suitability
    - *E.g., Is the problem clearly defined?*
  - Technical Suitability
    - *E.g., Are proposed data sources consistent, sufficient, and accurate?*

Possible Outcomes

- AI solution is applicable
- AI-enabled solution not advisable under these conditions
- Potential applicability issues, subjective assessment needed

Risk/Response Assessment

- Two components are used to estimate the level of rigor needed for the certification of the AI technology

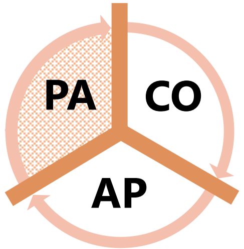


Risk Assessment

- Considers risk commonly associated with AI
- Each risk component contains one or more negotiable risk elements
  - A lower and higher risk description are defined for each element to establish a scoring range from 1 to 10
  - Average of the element scores is the component score

Risk Section	Risk Components
Scope	Intent
	Use Cases
	Roles
	Performance
	Benefits
Inputs	Data
	Bias
	Protected data
	Quality
	Assumptions
Model	Approach
	Prior instances
	Assumptions
	Feature selection
	Method of learning
Architecture	System accessibility
	External connectivity
	SWaP-C requirements
Oversight	Human involvement in decision-making
	Reporting requirements
Implementation	Documentation
	Intended deployment

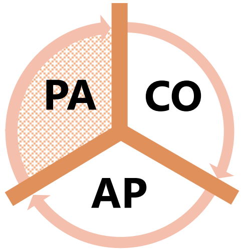
Response Assessment



- Identifies the AI function: what type of actions or activities is the technology aiding

Function	Use Case	Definition*
Planning	Advisory	AI application provides <u>informational output</u> result in near real-time (service) and provides a user (analyst) outcomes to be evaluated for final decisions.
	Situational Awareness	AI application provides <u>informational output</u> result that is made available for a user (analyst) to be evaluated for final decisions.
Decision Making	Strategic	AI application provides <u>notification output</u> of expected values for the detection of an event or condition, the information is supplied to a human for action.
	Tactical	AI application provides <u>notification output</u> of expected values for imminent or probable future event or condition evaluated; the information is supplied to a human for action.
Perceptive	Detection	AI application provides <u>alerts</u> in response to detecting a condition or event in real time.
Management/ Control	Human to System Interaction	AI application provides capability for the <u>human to interact with a system</u> and optimize a given function. The AI/ML model provide a recommended action to reach a "best" outcome.
	Multi-Systems Interaction	AI application uses data from multiple systems to make enhanced predictions with <u>output from system-to-system interaction</u> to determine a "best" course of action to avoid or mitigate problems or adjust to reach a better state in terms of safety or efficiency.

* Definitions are loosely based on intervention levels described in FAA Order 1370.121A Appendix 31



❗ Safety Assessment

- The safety component of the assessment will consider the functionality and use of the AI technology
- For framework v1.0, only low safety use cases are considered as defined by FAA's SRM process

Comply with guidelines from **FAA's Safety Risk Management (SRM)** process for system acquisition

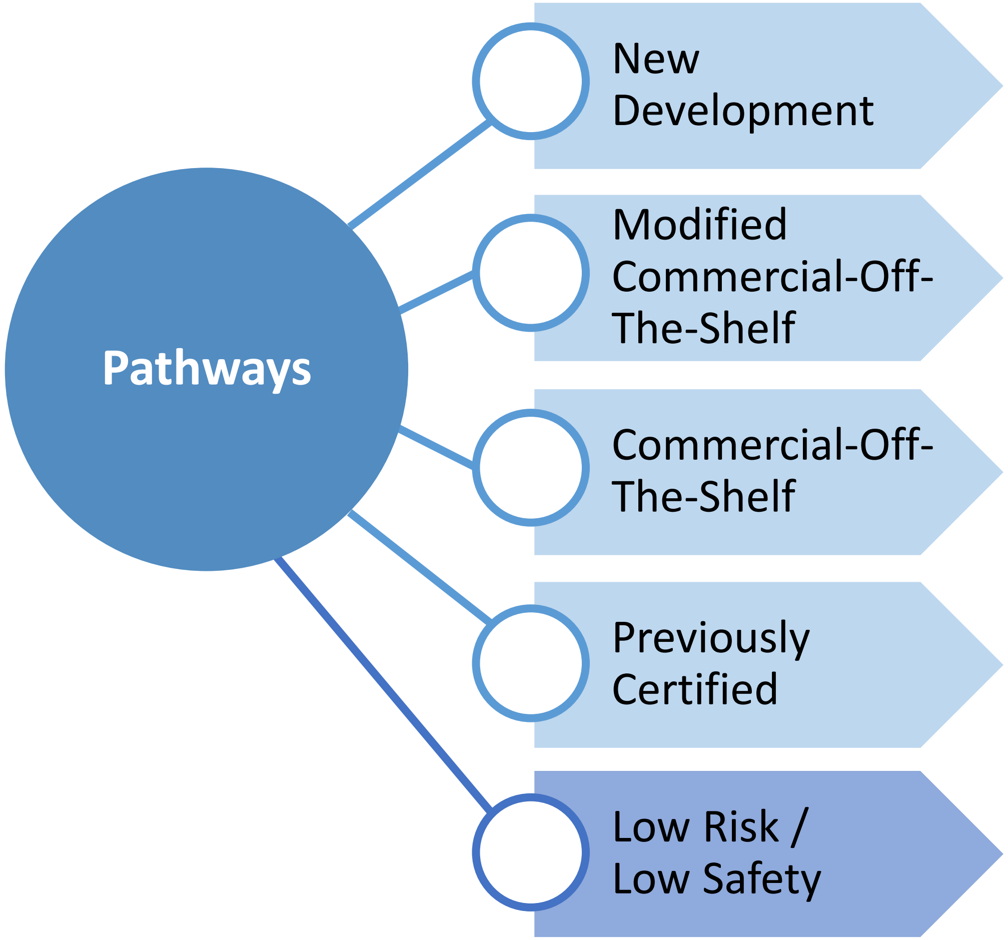
Pathway Determination

- The pathways may differ by **timing**, **criteria**, and **roles** but each pathway must comply with all objectives and criteria for approval at the required software assurance level

01	**Timing**	When does the certification process take place? E.g., continuous review during software development lifecycle or one aggregate review at the end of development
02	**Criteria**	What are the requirements for certification? E.g., previously certified software may be exempt from certain steps while COTS products may already have proof of compliance
03	**Roles**	Who is the certifying authority and what is their level of involvement? E.g., the certifying organization may be the FAA (different lines of businesses) or an approved 3rd party

Pathway Determination

PA CO AP



Pathways

New Development

AI technology completed, approved, and placed under **continuous configuration control throughout the software lifecycle**

Modified Commercial-Off-The-Shelf

AI technology independently developed, used **with modification** requiring some new development, and not previously certified

Commercial-Off-The-Shelf

AI technology independently developed, self-contained and used **without modification**, and not previously certified

Previously Certified

AI technology certified in a prior project and seeking approval in a **follow-on certification** project (e.g., re-training/upgrading in current system or new use in another system)

Low Risk / Low Safety

AI technology determined to be **low risk** and pose **no safety hazards**

ICN

Notional Certification Flowchart

Legend

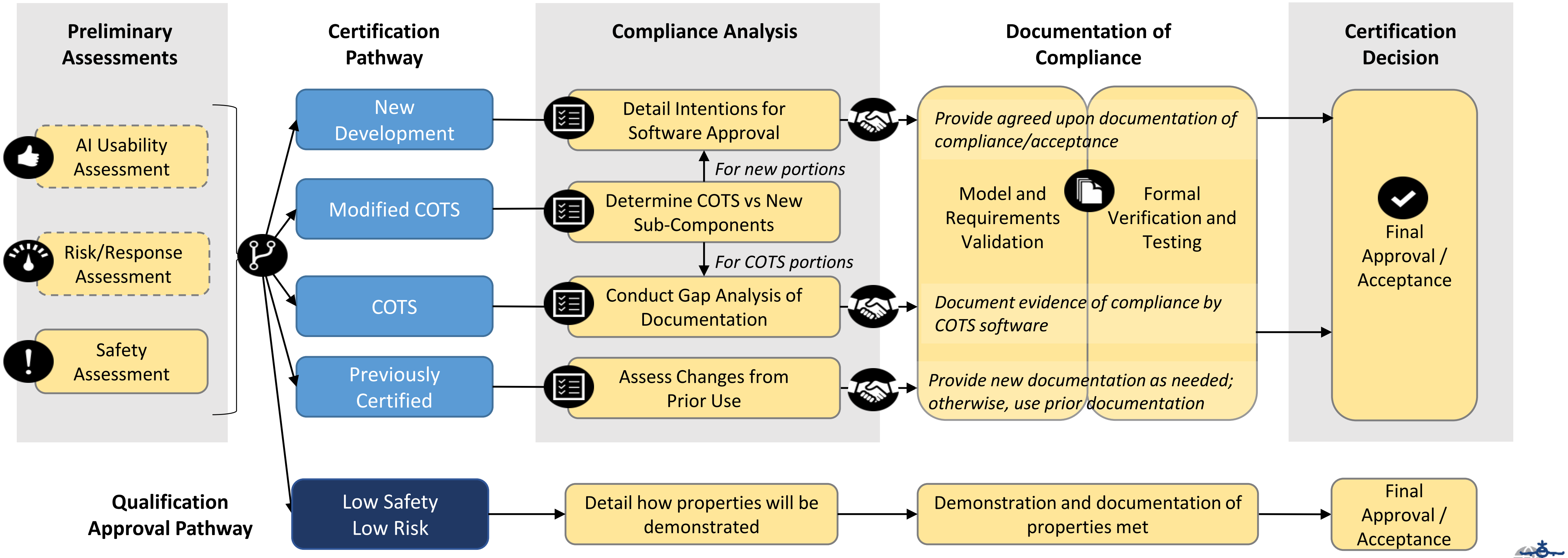
Contains additional level of flowchart, checklists or other recommendations for AI technologies

Certification Pathways

Qualification Pathways

New documents for AI certification

- The notional certification process shows where new AI-specific checkpoints or analyses may be needed for different types of technologies



Preliminary Assessments

AI Usability Assessment

Risk/Response Assessment

Safety Assessment

Certification Pathway

New Development

Modified COTS

COTS

Previously Certified

Compliance Analysis

Detail Intentions for Software Approval

For new portions

Determine COTS vs New Sub-Components

For COTS portions

Conduct Gap Analysis of Documentation

Assess Changes from Prior Use

Documentation of Compliance

Provide agreed upon documentation of compliance/acceptance

Model and Requirements Validation

Formal Verification and Testing

Document evidence of compliance by COTS software

Provide new documentation as needed; otherwise, use prior documentation

Certification Decision

Final Approval / Acceptance

Qualification Approval Pathway

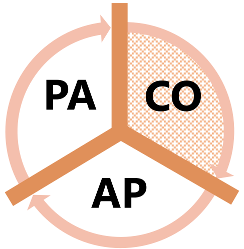
Low Safety Low Risk

Detail how properties will be demonstrated

Demonstration and documentation of properties met

Final Approval / Acceptance

*Based on the current software assurance processes used by the FAA from RTCA DO-178C and 278A.



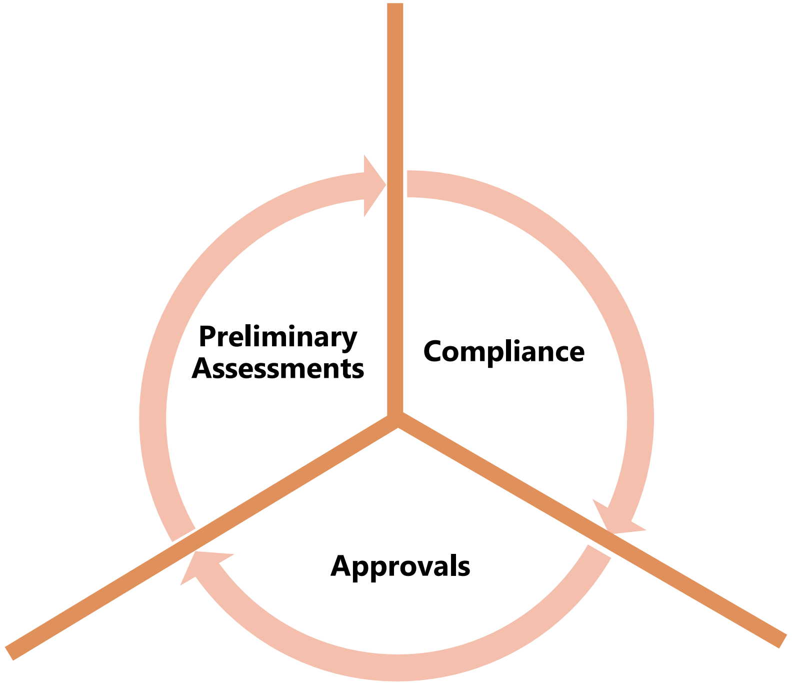
Compliance Phase

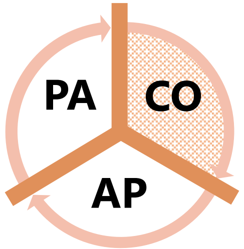
	Phase	Purpose	Outcome
02	Compliance	Define and compile all documentation necessary for approval	• Compliance documentation

Documentation Needed: Based on the types of components in the technology, propose any missing, new, or prior documentation to be submitted as evidence of compliance

Documentation Determination: Certifying Authority approves the plan for documentation required and tailored for certification of the AI technology

Document Compliance: Provide evidence of compliance regarding model and validation requirements as well as testing and verification



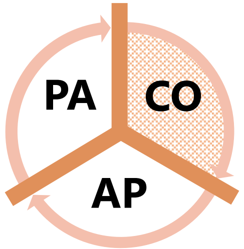


Compliance Objectives

- Each compliance category has a checklist of objectives
- Documentation is required to show evidence that objectives are satisfied
- Any objectives that are not applicable should be justified as such

Example Framework Guidance for a Compliance Objective

- Category: Configuration Management
- Objective: All supporting and resultant items to be configured are defined.
- Suggested Actions:
  - Define the training, testing, validation, and other datasets as configuration items.
  - Define the algorithms and trained model parameters as configuration items.
  - Etc.
- Documentation: A compilation of definitions of each identified configuration item.



Compliance Phase Activities

Compliance Analysis

Documentation Needed

- Identify based on Certification Pathway
- Assess gaps in COTS documentation
- Assess changes from prior use
- Propose which compliance objectives are applicable and what documentation will be provided

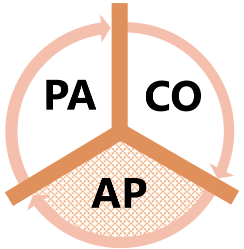
Documentation Determination

- Negotiate tailored documentation needs for the AI technology with the appropriate FAA line of business

Documentation of Compliance

Document Compliance

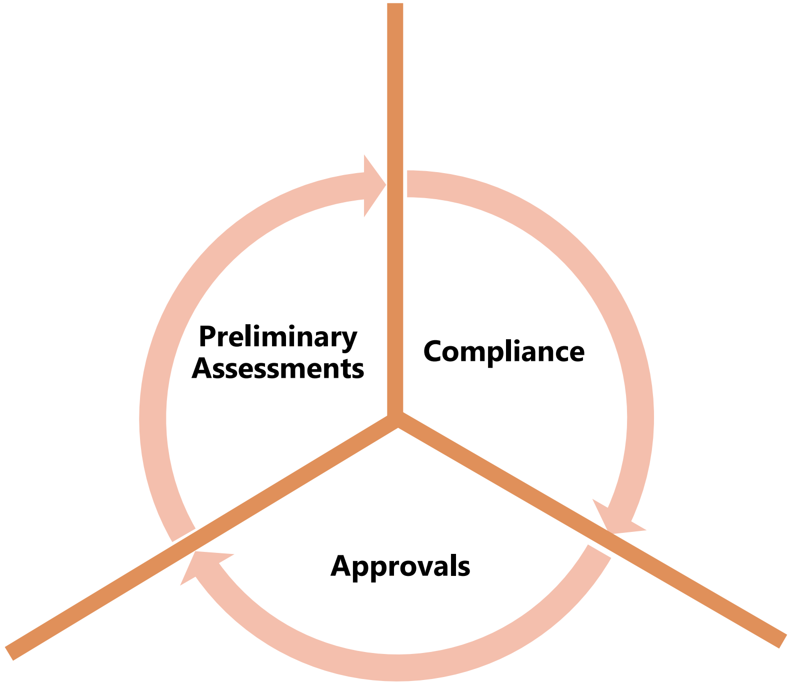
- Provide documentation to show evidence of meeting the agreed upon objectives
  - ❑ New documentation
  - ❑ Old documentation from prior use
  - ❑ Existing documentation from COTS vendor



Approvals Phase

	Phase	Purpose	Outcome
03	Approvals	Coordinate with stakeholders to review and approve documentation	• Certification decision

Final Approval / Acceptance: Certifying Authority reviews documentation and approves technology



QUESTION