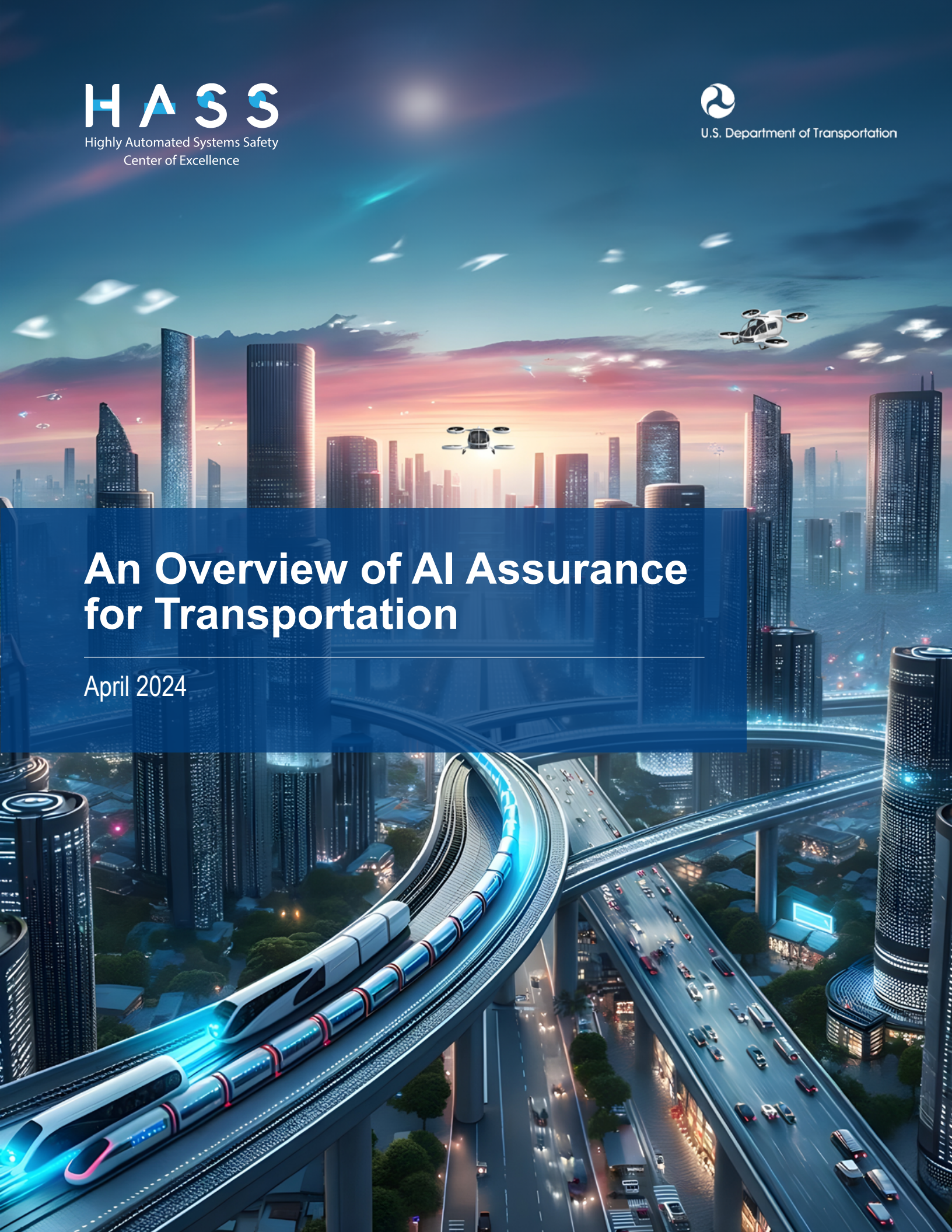
HASS
Highly Automated Systems Safety
Center of Excellence

U.S. Department of Transportation



An Overview of AI Assurance for Transportation

April 2024

Table of Contents

AI Assurance

AI assurance refers to the techniques, activities, and processes used to evaluate and assure expected properties of AI components or AI-enabled systems throughout the lifecycle of these components or systems. The properties may relate to performance, safety, or security. Typical examples of AI assurance techniques are verification, validation, and safety analysis of AI components and AI-enabled systems.

AI assurance use cases include performance and safety assessment of highly automated systems and their ecosystems, from perception, localization, planning, and control functions, to AI-based data analysis.

AI Assurance Concepts

Six key concepts make up AI assurance for applications in automated transportation systems (see Figure 1):

Figure 1. AI assurance concepts.

Concept	Definition
AI use cases	Use cases of AI in highly automated systems and analytical tools.
AI risk assessment and mitigation	Includes hazard/risk identification and assessment, safety assessment, and risk mitigation techniques for the AI components and AI-enabled systems.
AI data assurance	Covers all data collected, generated, processed, and maintained in the AI lifecycle.
AI design assurance	Assurance of AI models in the development phase, including models in the pre-processing, training, optimization, and post-processing activities.
AI implementation assurance	Assurance of AI models in the compiling, interpretation, or inference phase. Model transformation/conversion, optimization, or synthesis are usually performed against these models.
Support for assessment of AI or regulatory efforts	Includes quantitative and qualitative assessment of AI regarding predefined properties of performance, safety, and security, as well as how to use assessment results to support AI-related regulatory efforts.

AI Assurance Challenges

The increasingly complex and scalable nature of advanced AI systems poses assurance challenges related to technologies and government.

Technical Challenges

- Maturity, usability, and scalability of assurance technologies, tools, and processes
- Appropriate and sufficient evaluation metrics and criteria for the assessment of safety goals
- Standardized requirements, benchmarks, testing platforms, and simulation environments
- Adaption to the fast evolution of AI and its data
- Standards update to keep up with the evolution of AI

Challenges to Government

- Lack of expertise, experience, tools, and workforce for the assessment of AI systems
- Limited access to AI techniques and data owned by industry or other stakeholders
- Absence of powerful and credible testing/evaluation environment and benchmarks
- Missing guidance or regulatory requirements

AI Assurance Technologies

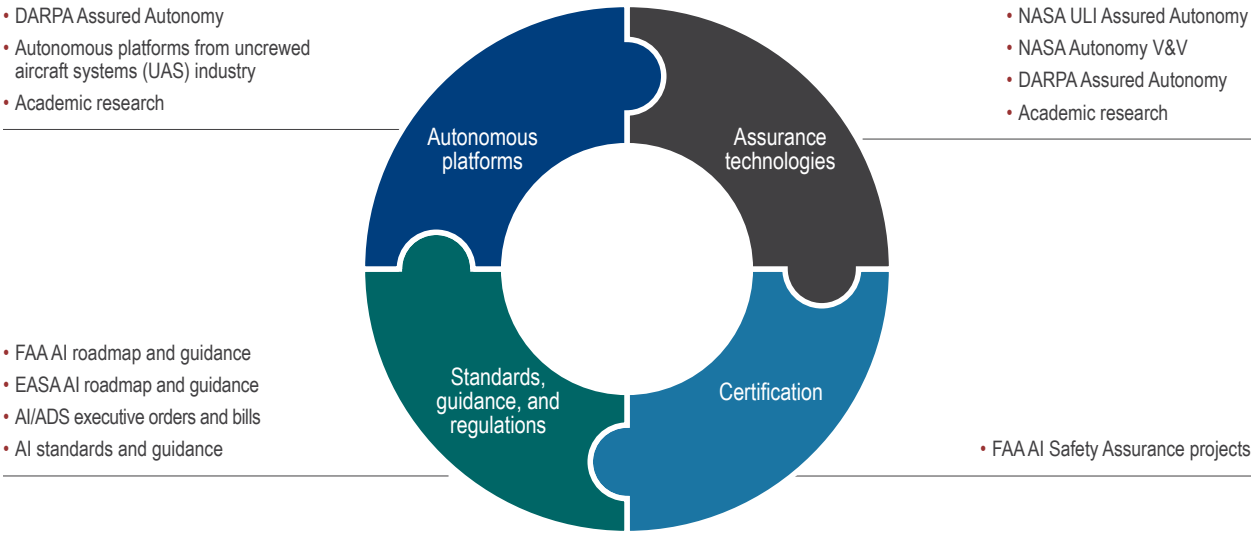
AI assurance technologies include AI-related data analysis, AI model verification and validation, model performance analysis, AI hardware analysis and verification, model integration verification, run time monitors and verification, and contingency management of AI components, their related data, and the AI-enabled systems or vehicles. Testing and verification objectives include safety or performance properties such as correctness, accuracy, sensitivity, reliability, explainability, robustness, fairness, security, and privacy.

Current/Past AIA Efforts in Aviation

The U.S. Government and aviation industry are making considerable efforts in the development of safety assurance technologies for AI-enabled systems. Major investments from DARPA, NASA, and industry focus on the development of autonomous platforms, assurance technologies, guidance and standards with consideration of future certification requirements.

Figure 2 shows recent efforts in assuring the safety of AI-enabled systems.

Figure 2. Recent efforts in assuring the safety of AI-enabled systems.



EASA: European Union Aviation Safety Agency **ULI:** University Leadership Initiative **V&V:** Verification and Validation

Guidance and Standards Considered in AIA

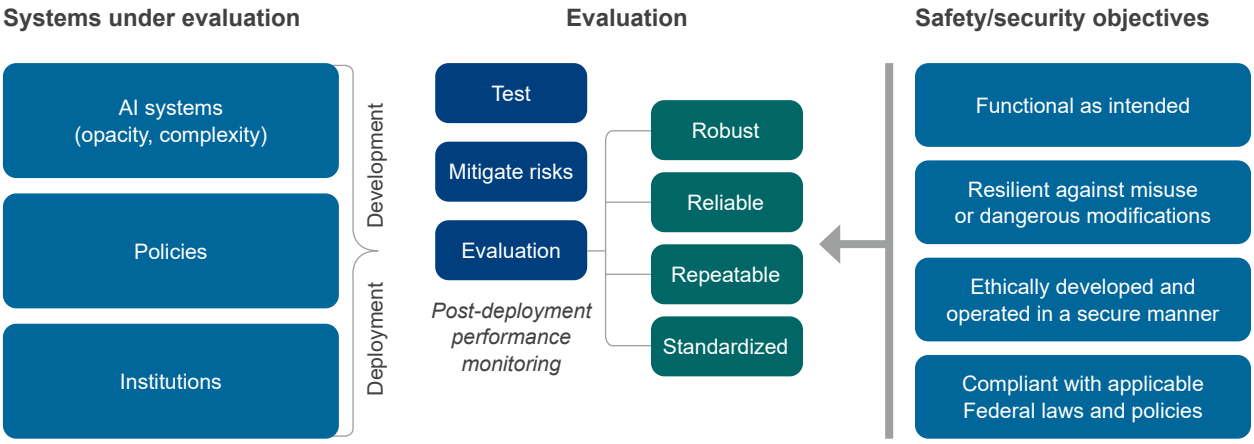
No AI assurance standards exist yet, but consideration of current safety standards is important to identify, understand, and fill the gaps for AI assurance:

- FAA AI roadmap
- RTCA DO-178 and DO-254
- SAE ARP4761
- NHTSA Federal Motor Vehicle Safety Standards (FMVSS) and Automated Driving Systems: A Vision for Safety
- ISO 26262 and ISO/PAS 21448

Safety Approach Defined in the AI Executive Order

The Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence was published on October 30, 2023. The basic evaluation and testing approach for safety and security (see Figure 3) provides high-level guidance for the development of AI assurance technologies and frameworks.

Figure 3. Evaluation and testing approach for safety and security.



AI Assurance Framework

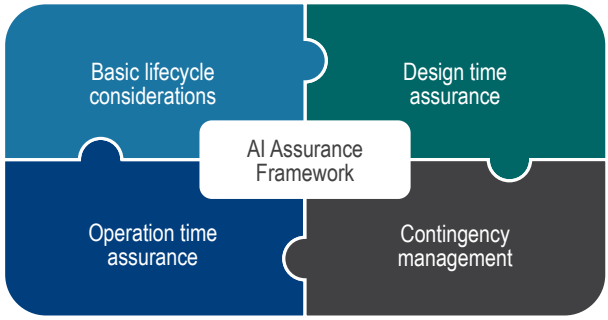
Basic Lifecycle Considerations

Basic lifecycle considerations (see Figure 4) include, but are not limited to, **AI requirements**, **data**, and **safety assessment**, addressed throughout the AI development, deployment, and operation phases.

Considering the large number of AI assurance techniques, an assurance framework is needed for the identification, positioning, and coordination of assurance techniques in the AI development lifecycle to achieve expected safety and performance objectives (see Figure 4).

The framework should be open and flexible, which is not subject to mandated order or regulatory requirement. The AI assurance framework described in the following section is a specific implementation of the previous safety approach (see Figure 3) for safety-critical transportation systems.

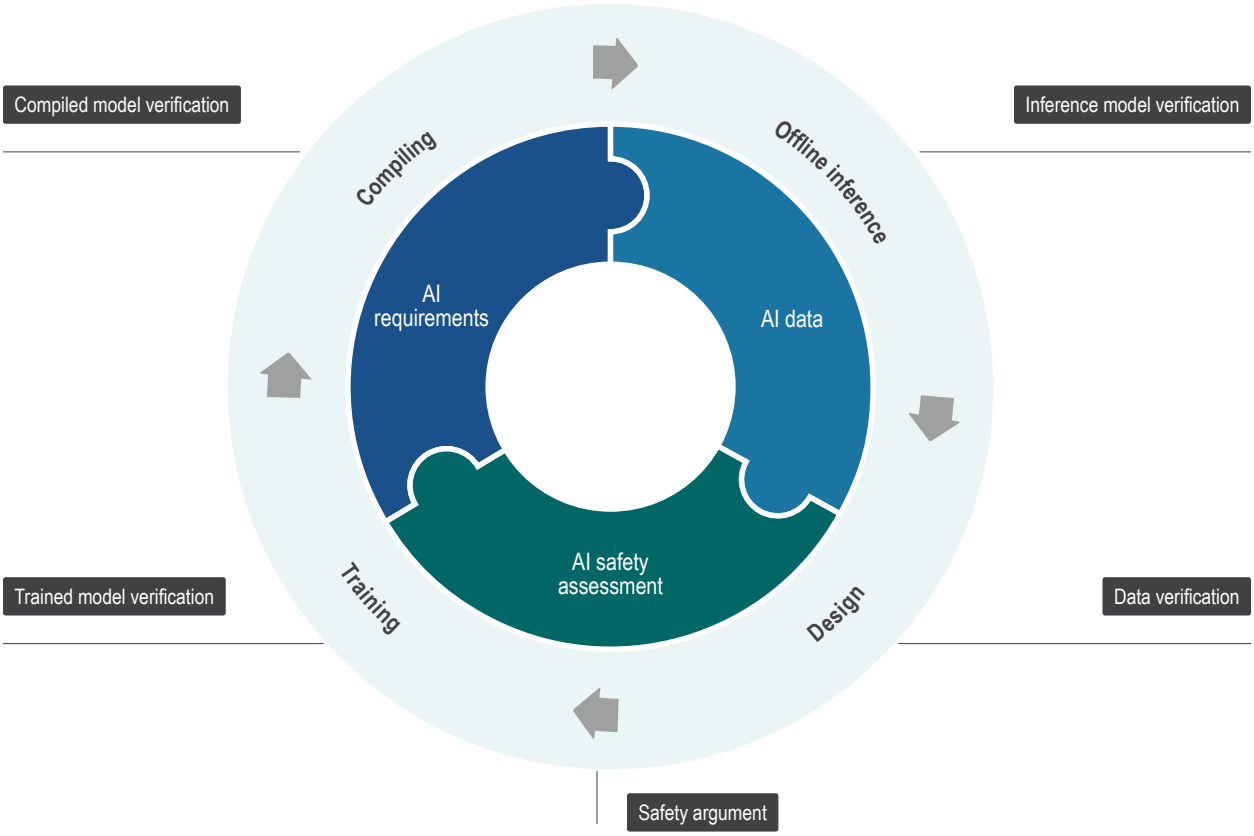
Figure 4. AI assurance framework components.



Design Time Assurance

Design time assurance (DTA) activities include analysis, verification, validation, testing, and simulation of AI models and data in development, as well as use of the assurance results in structured argument to demonstrate that predefined safety and performance goals are achieved (see Figure 5). DTA covers AI development activities, from data collection/generation, model design, training, optimization, and compiling, to inference. To achieve design time assurance, appropriate and sufficient metrics and criteria for quantitative or qualitative evaluation are required in the whole development. Well-established or standardized requirements, benchmarks, testing platforms, and simulation environments are also necessary.

Figure 5. Design time assurance activities



Operation Time Assurance

AI operation time assurance (OTA) deals with operational constraints, uncertainties, faults/failures, and any unexpected situations that cannot be addressed by DTA. OTA monitors AI components and their enabled systems to detect any violations of operation time requirements; and with safety guards, OTA assures only safe functions are performed.

OTA includes runtime verification for deployment, integration, online data, inference models, and safety arguments regarding expected safety and performance goals based on OTA results and operational conditions (see Figure 6). Safety goals are usually directly derived from analysis of safety hazards and risks identified from AI safety assessment activities.

Figure 6. Operation time assurance activities.

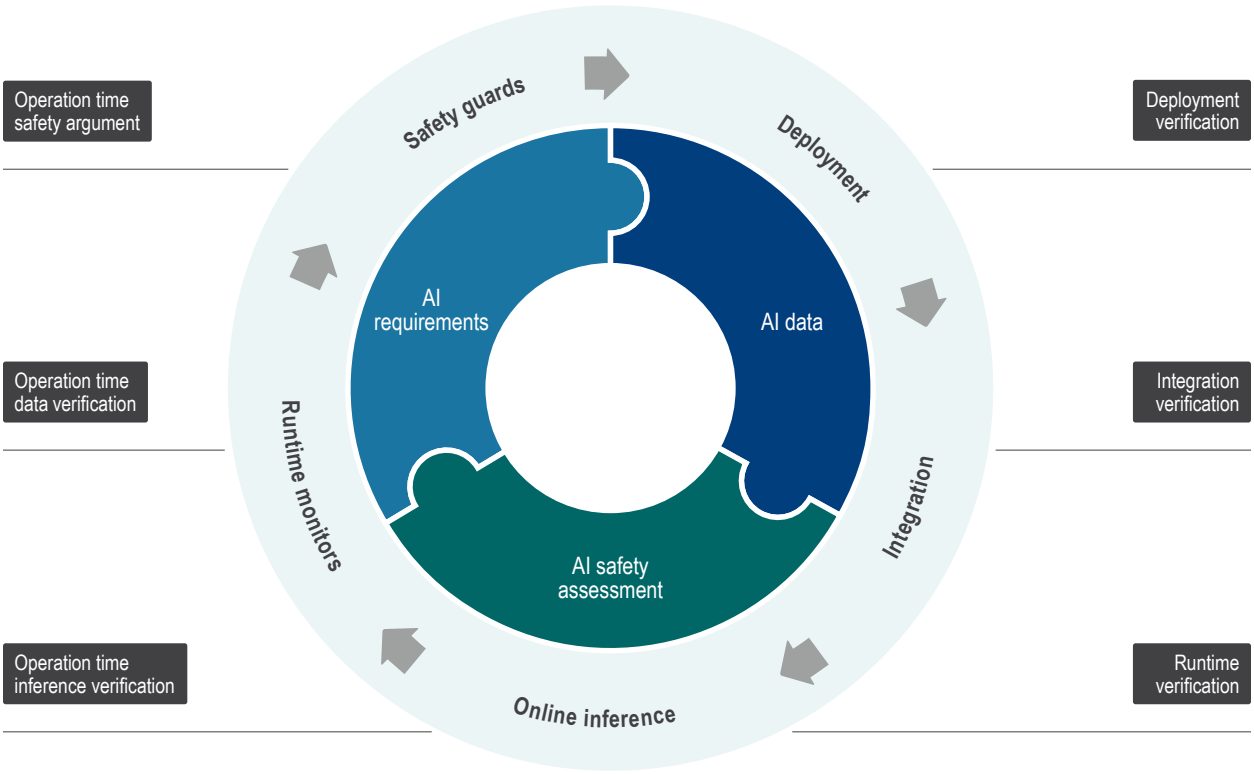
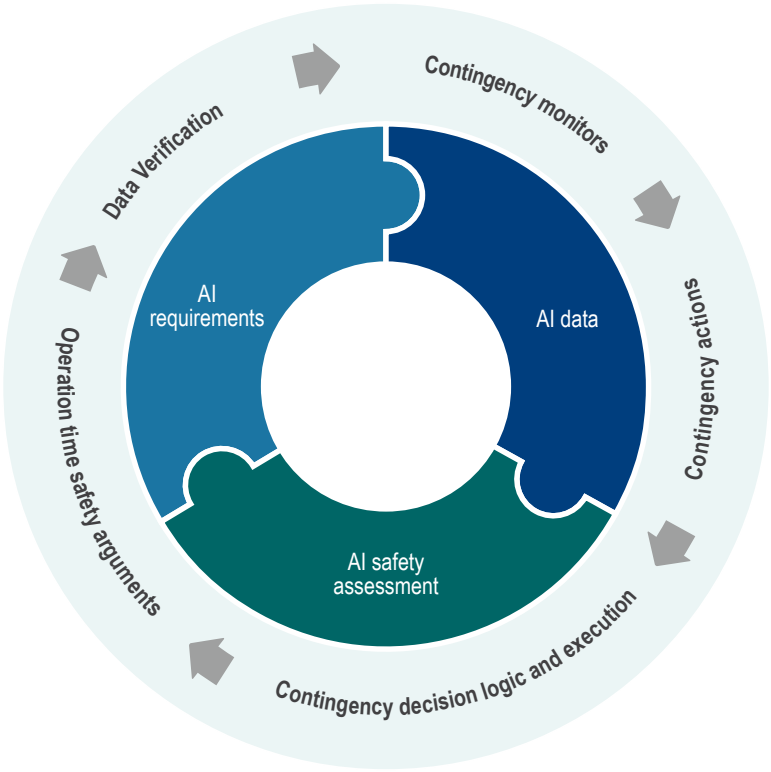


Figure7. Contingency management activities.



Contingency Management

AI Contingency Management addresses system failures and situations that cannot be addressed by AI and its assurance (see Figure 7), as a very last measure to guarantee system and operation safety with potentially degraded service or end of service (e.g., emergency stop or contingency landing).

AI Assurance Workshops and Working Group

To promote coordination and collaboration on AI assurance across U.S. DOT Operating Administrations, HASS COE organized the first AI Assurance Workshop on November 30, 2023, at U.S. DOT Headquarters. Additional workshops in 2024 invite industry and academic partners to exchange knowledge and collaborate on AI assessment and assurance for applications in automated transportation systems. The workshops welcome all interested U.S. DOT employees to share their needs, challenges, current state, and potential solutions for AI assurance across all modes of transportation.

The AI Assurance Working Group aims to facilitate coordination, collaboration, and knowledge sharing across U.S. DOT. The working group is open to U.S. DOT employees and will involve regular meetings, reports, and technical talks.

Contact:

Huafeng Yu, Ph.D.
Senior Scientist, HASS COE
huafeng.yu@dot.gov
transportation.gov/hasscoe

Co-authors:
Taylor Lochrane, Ph.D., P.E. (HASS COE);
Trung Pham, Ph.D. (FAA); Srini Mandalapu, Ph.D. (FAA);
George Romanski (FAA); Denise Bakar (HASS COE)

U.S. Department of Transportation

HASS
Highly Automated Systems Safety
Center of Excellence