NASA System-Wide Safety Project

R&D on Complex Autonomous Systems Assurance

April 24, 2025

Hosted by George Washington University and the National Aeronautics Research Institute

NASA System-Wide Safety Project

WELCOME

Airspace Operations and Safety Program (AOSP)

Akbar Sultan, Program Director

NASA Headquarters, Washington, DC

NASA System-Wide Safety Project

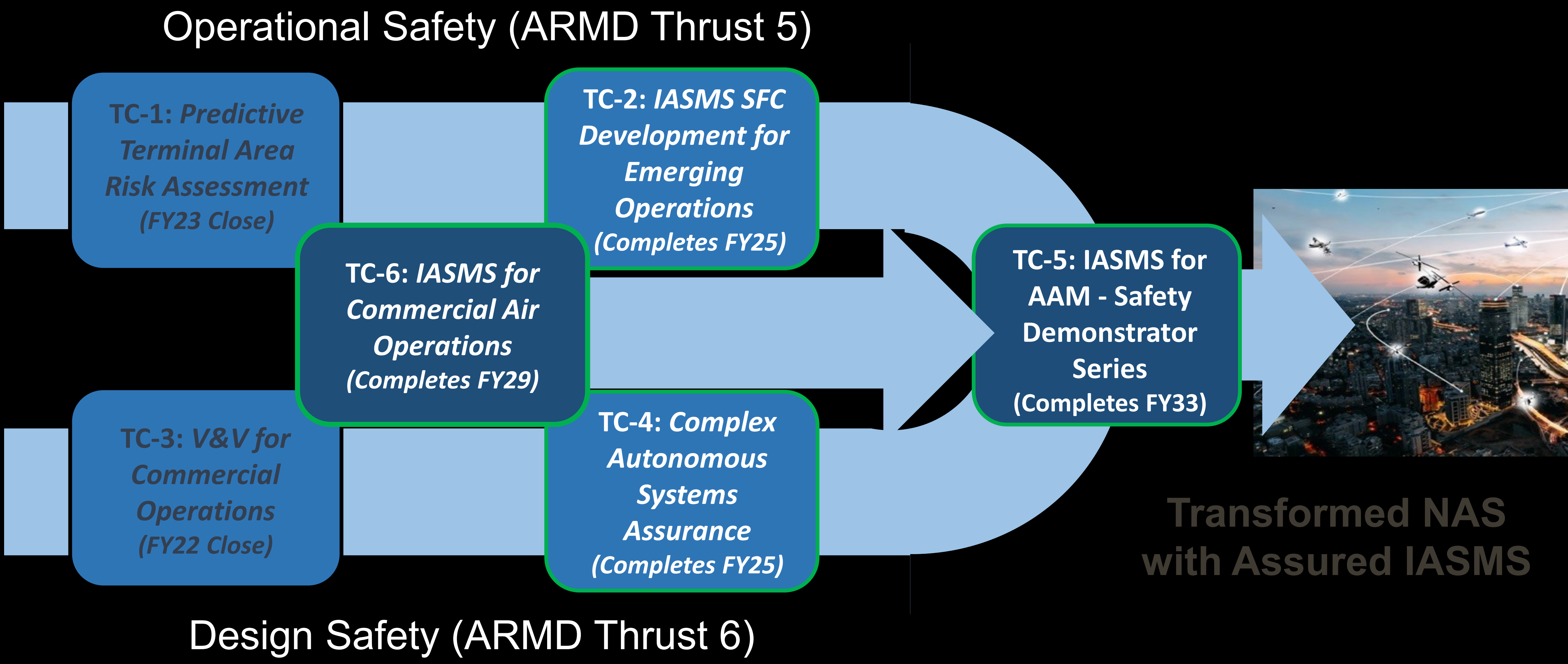
WELCOME

System-Wide Safety Project (SWS)

Summer Brandt, Project Manager

NASA Ames Research Center, Moffett Field, CA

SWS Technical Challenges



NASA System-Wide Safety Project

OVERVIEW

Mission, program, and project context

What is assurance of autonomous systems?

R&D scope, assumptions, objectives, and approach

Summary list of activities and capabilities

SWS TC-4: R&D on Complex Autonomous Systems Assurance

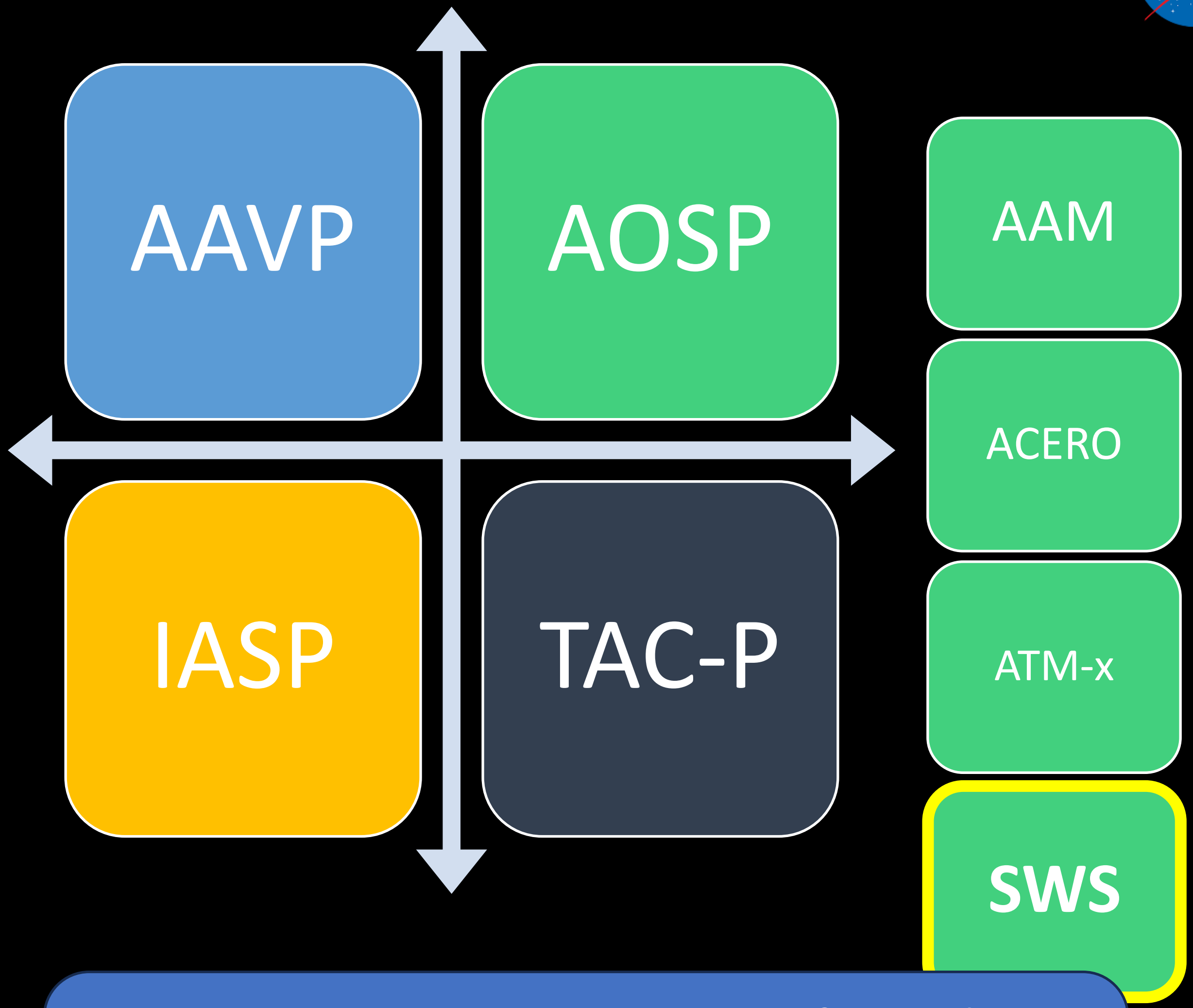
Terry Morris, PhD, Sub-project Manager, NASA Langley Research Center, Hampton, VA

Guillaume Brat, PhD, Sub-project Manager, NASA Ames Research Center, Moffett Field, CA

NASA's Mission context

"NASA explores the unknown in air and space, innovates for the benefit of humanity, and inspires the world through discovery."

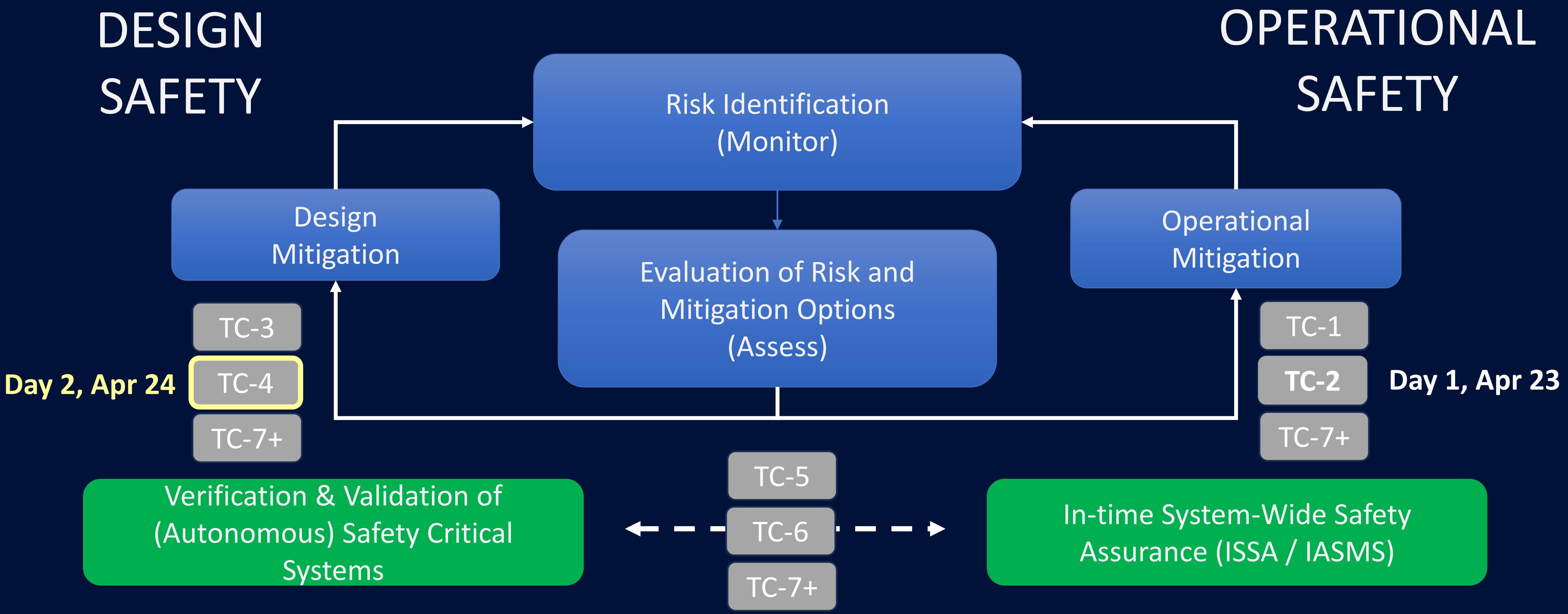
- Aeronautics Research is organized into four programs; each having multiple projects and sub-projects
- The System-Wide Safety (SWS) project defines sub-projects based on Technical Challenges associated with ensuring the safety of flight



SWS TC-4: R&D on Assurance of Complex Autonomous Systems

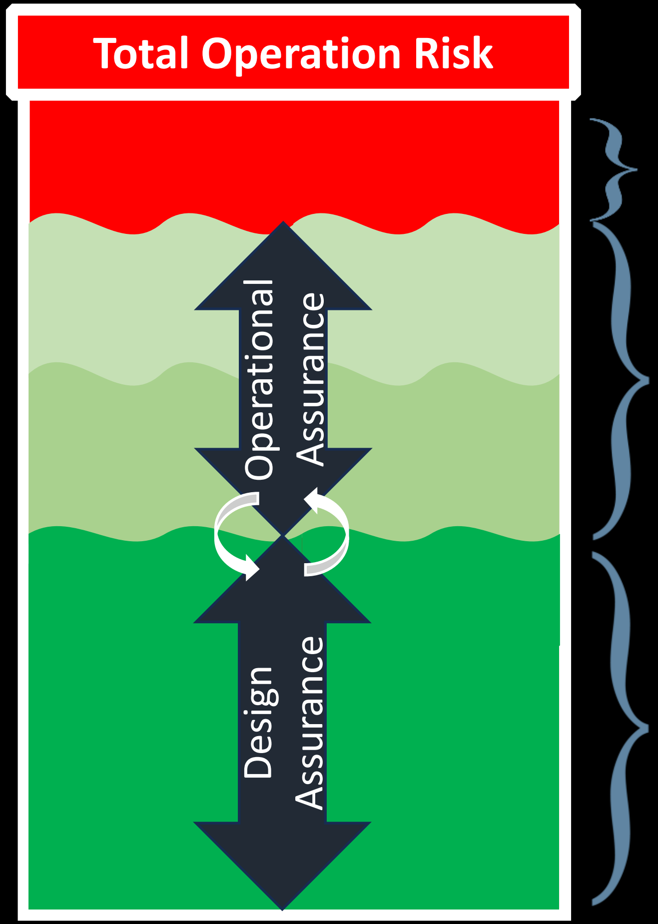
(Explore – Innovate – Discover)

SWS Research – Strategy, Scope, and Technical Challenges



Monitor – Assess – Mitigate

Design & Operational Safety



***Residual / Retained Risk*:**
To be reviewed by the accountable person

Operational Safety:
Industry Good / Best Practice from Reactive and Proactive Safety Systems

***Design Safety*:**
Regulations, Approved Processes and Industry Operating Standards

Safety implementations are context specific, tailored to address hazards and related risks inherent to the operation

What is Assurance of Autonomous Systems?

Traditional systems have a well-established path to certification. It may be expensive and not as efficient as one might like, but it exists. For example, DO-178C for software.

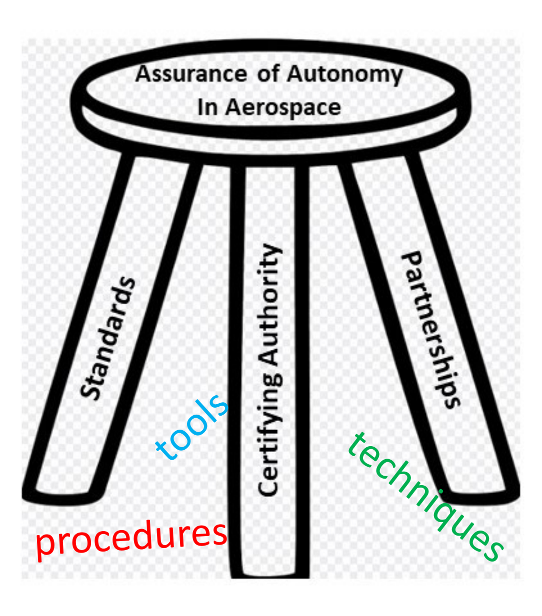
For autonomy systems, and for AI-enabled systems, it is not clear if a path exists. For example, looking at regulations for model-based systems could help. However, there is a consensus that new assurance techniques (to retire risks) are needed, and more importantly, certification guidelines are needed.

That's the <u>focus</u> of this TC4 Technical Challenge

TC4

(3-legged stool)

Approach



- <u>Some of V&V Design Challenges include:</u>

  Refine/create new risk analysis techniques for autonomy-specific risks

  New, and early, design-time assurance techniques
  - Early, to find problems more quickly and reduce V&V costs
  - New, to address data-centric techniques, AI and in particular ML.

  Operational assurance techniques to address new risks

  Develop guidelines for certifying these systems; Start with ML-enabled systems

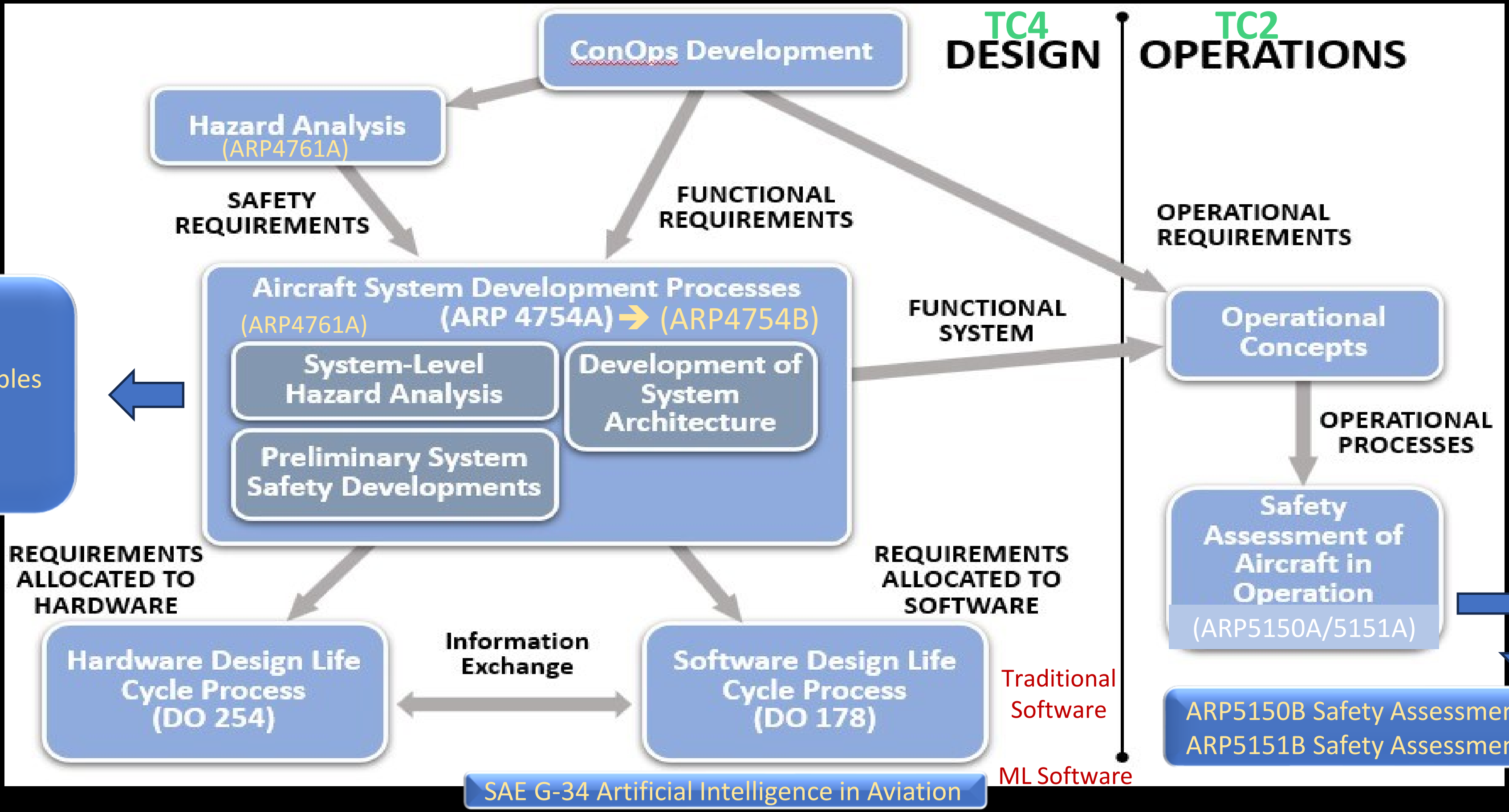
TC4 R&D Impact to Design Standards

SWS

Standards

Standards as of April 2025
(date means published)
ARP4754B (12/23)
ARP4761A (12/23)

SAE AIR7209 Development Assurance Principles for Aerospace Vehicles and Systems (10/22)



ARP5150B Safety Assessment of Transport Airlines in Comm Serv
ARP5151B Safety Assessment of GA Airplanes & Rotorcraft in CS

UL 4600 Evaluation of Autonomous Products, 2nd Ed. (road veh) (10/22)
UL 4601 Evaluation of Autonomous Air Products (UAVs)

Security →	AIR8480 Safety-Security Interactions for Aircraft/System Development
ARP4761A →	AIR9953 Applying Dev Assurance with Model-Based Systems Engineering
ARP4761A →	AIR6913 Using System-Theoretic Process Analysis During Dev and Safety for Civil AC
ARP4754B →	AIR6276 Use of Modeling and Tools for AC Systems Development
ARP4754B →	AIR6218A Constructing a Development Assurance Plan for Integrated Systems
ARP4754B →	AIR7121 Applicability of Existing Dev Assurance & System Safety Practices to UAS (NEATT)
ARP4754B →	AIR7127 Human Considerations for Functional Hazard Assessment (HF)

Certifying Authorities

Safety Wide System Research Transition Team Structure

SWS RTT

Co-Leads:
Misty Davies (NASA)
&
Laura Rodríguez López (FAA-AVP)

Increasing Automation & Complexity, including assurance of AI/ML

Working Group

NASA Co-Leads:
Aaron Dutle and Jon Holbrook

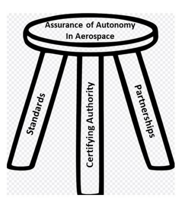
FAA Co- Leads:
Bill Kaliardos and Kathy Abbott
(Trung Pham for Assurance of AI/ML)

Safety Analytics, Risk Synthesis, and Operational Safety

Working Group

NASA Co-Leads:
Michael Vincent and Nikunj Oza

FAA Co-Leads:
Laura Rodríguez López and Robert DeWolf
(Trung Pham for Assurance of AI/ML)



V&V 2045 Roadmap

Partnerships

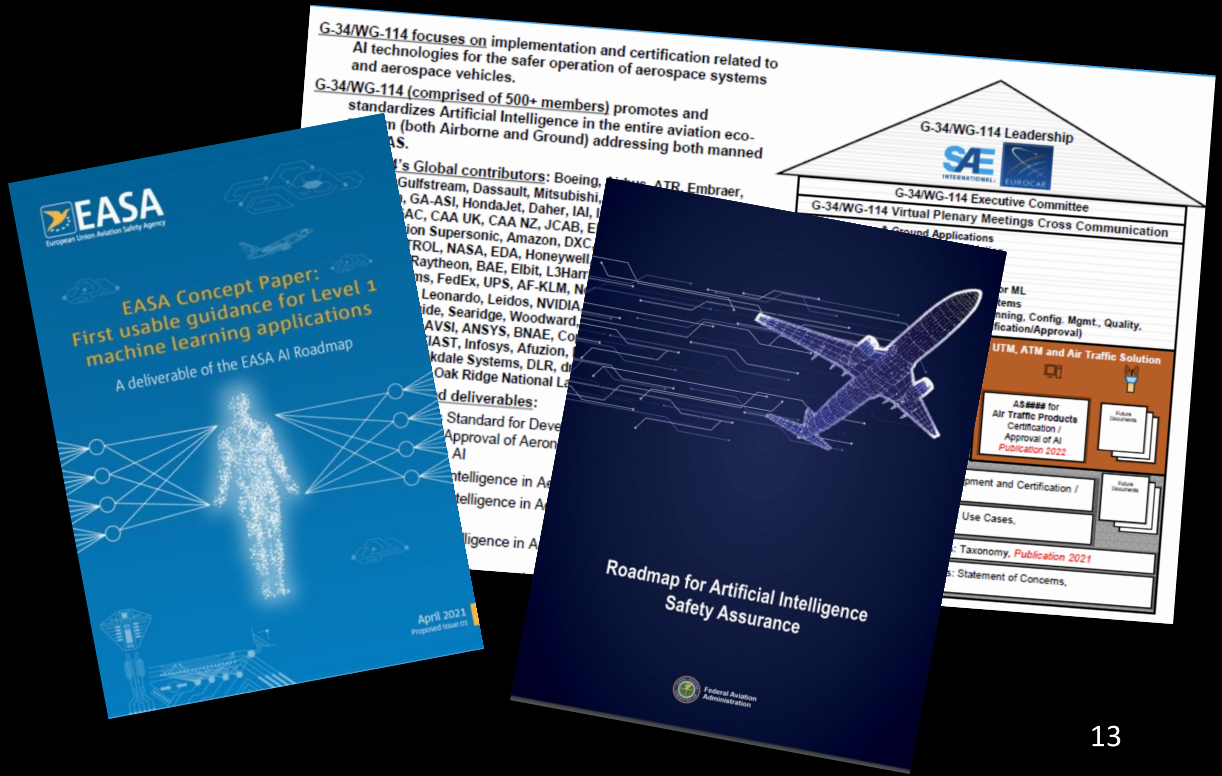
Deliverables – Completed Jan. 2023:

- A list of technologies needed to enable the use of autonomy in aviation
- A mapping of the V&V needs to (possibly Artificial Intelligence-based) technologies and the enabled services in Aviation
- A study of the gaps in the current certification processes and an assessment of potential technologies that could fill those gaps

NASA's roadmap is consistent with, complements, validates, and is validated by other current large-scale assurance of autonomy efforts.

Follow-on Activities:

- ➢ Validation of roadmap through Department of Transportation workshops (industry, academia) and NASA-led V&V workshops (other government agencies)
- ➢ NASA's use of the roadmap to determine highest-priority investment opportunities (e.g. FAA and G-34 use cases)



V&V 2045 Roadmap Results Summary (1 of 2)

Recommendations were split across 3 different time ranges

Near-Term (1-5 years) | Mid-Term (6-10 years) | Long-Term (11-25 years)

The report considers use cases requiring autonomous systems, their possible timeframe of full deployment, and the general categories those use cases fit into.

Near-Term (1-5 years)	Mid-Term (6-10 years)	Long-Term (11-25 years)
Non-Safety-Critical Systems & Functions	Safety-Critical Systems with Safety Pilots	Safety-Critical Systems & Functions
Use Case Example: Environmental Monitoring	**Use Case Example: Forward / VTOL Transitions**	**Use Case Example: Self-Managed Autonomous Flight**

V&V 2045 Roadmap Results Summary (2 of 2)

The report then identifies the assurance evidence necessary to enable those use cases and points out over 200 gaps. The report *prioritizes the top 10 gaps* with the largest impact* if closed sooner rather than later.

Near-Term (1-5 years)

- Open case studies demonstrating how ML-based systems or components can be developed to satisfy the emerging certification guidance (such as what NASA did with DO-333)
- Case studies developed to demonstrate use of RTA in the design of real systems, and satisfaction of specific objectives
- Test generation and coverage metrics should be developed for ML-based systems (neural networks in particular)

Mid-Term (6-10 years)

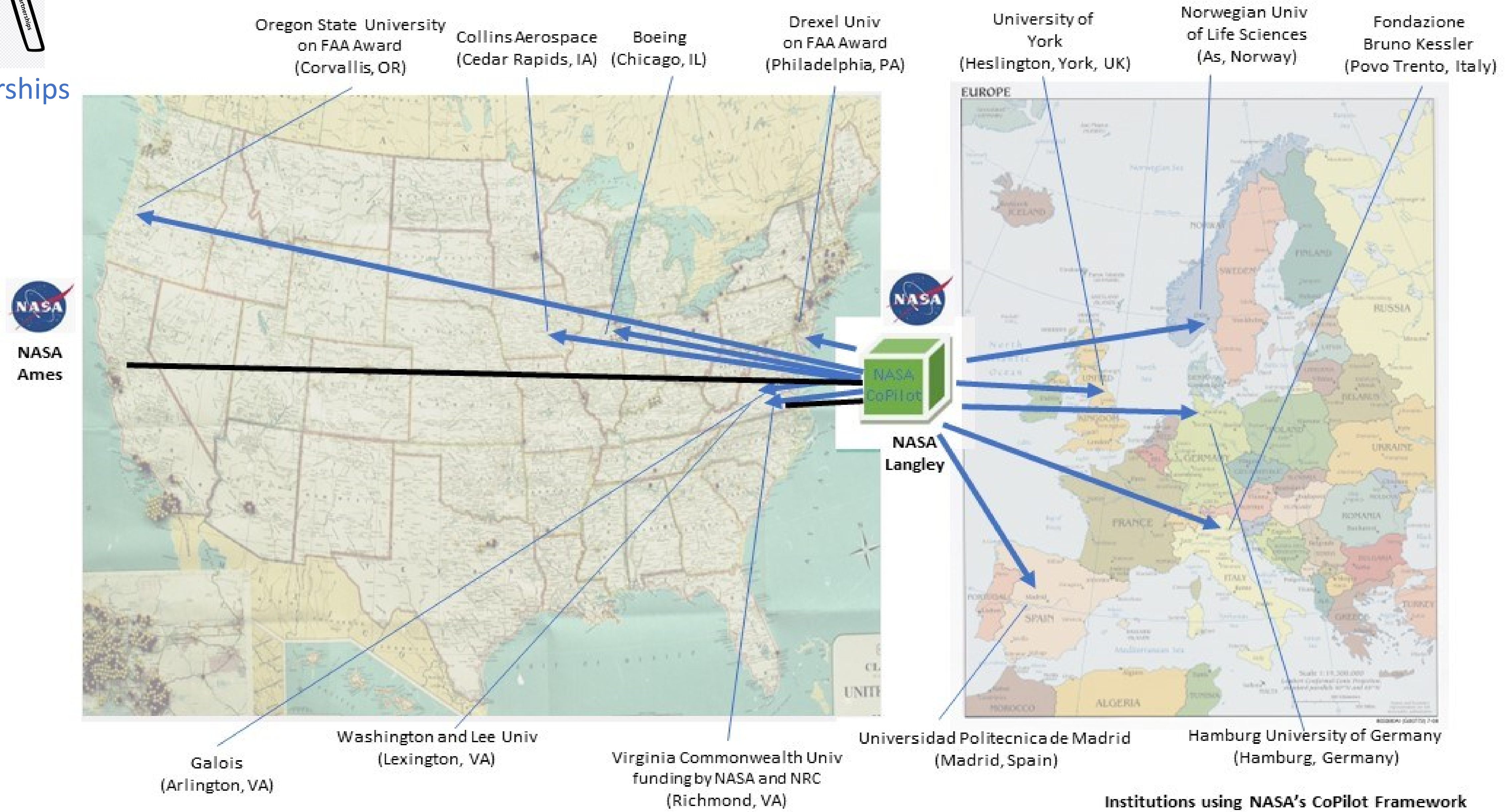
- Demonstrate completeness and representativeness of training and verification datasets with respect to the operational design domain (ODD)
- Methods, languages, and tools for specifying requirements for vision-based systems.
- Demonstrate interactions among monitors and mitigation actions are conflict-free
- Scalability of formal analysis

Long-Term (11-25 years)

- The ability to safely incorporate online learning into an ML-based sensing or control system. (e.g. Dynamic or incremental assurance methods)
- Trustworthy machine-vision systems based on "hybrid AI" research that is easier to assure and to explain
- Methods for assuring the integration of systems with differing levels of autonomy and trustworthiness

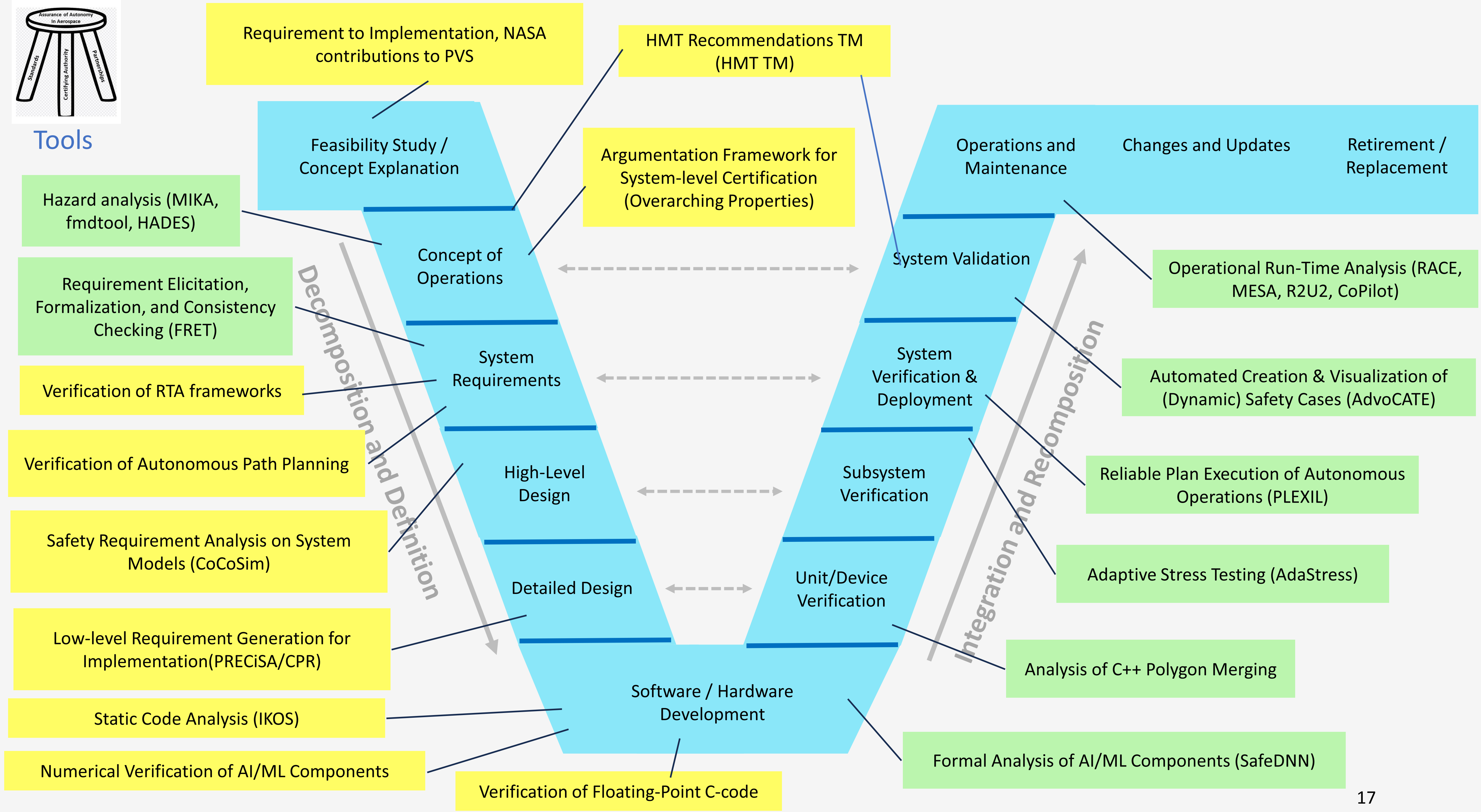
*Out of ~200 identified gaps

Partnerships



Institutions using NASA's CoPilot Framework

Tools



Assurance of Autonomy In Aerospace
Standards
Certifying Authority
Partnerships
Requirement to Implementation, NASA contributions to PVS
HMT Recommendations TM (HMT TM)
Feasibility Study / Concept Explanation
Argumentation Framework for System-level Certification (Overarching Properties)
Hazard analysis (MIKA, fmdtool, HADES)
Concept of Operations
Requirement Elicitation, Formalization, and Consistency Checking (FRET)
System Requirements
Verification of RTA frameworks
Verification of Autonomous Path Planning
High-Level Design
Safety Requirement Analysis on System Models (CoCoSim)
Detailed Design
Low-level Requirement Generation for Implementation(PRECiSA/CPR)
Software / Hardware Development
Static Code Analysis (IKOS)
Numerical Verification of AI/ML Components
Verification of Floating-Point C-code
Decomposition and Definition
Operations and Maintenance
Changes and Updates
Retirement / Replacement
System Validation
Operational Run-Time Analysis (RACE, MESA, R2U2, CoPilot)
System Verification & Deployment
Automated Creation & Visualization of (Dynamic) Safety Cases (AdvoCATE)
Reliable Plan Execution of Autonomous Operations (PLEXIL)
Subsystem Verification
Adaptive Stress Testing (AdaStress)
Unit/Device Verification
Analysis of C++ Polygon Merging
Formal Analysis of AI/ML Components (SafeDNN)
Integration and Recomposition

Overview of Selected Design Assurance Research

- Assuring Increasingly Autonomous Systems for Human Machine Teaming in Novel Roles
- Run Time Assurance
- Verification Technology Topics
- Verification of Autonomous Systems
- Design Assurance Collaborators Viewpoints
- Future Work