AI/ML Assurance

G. Brat, K. Ellis, S. Brandt

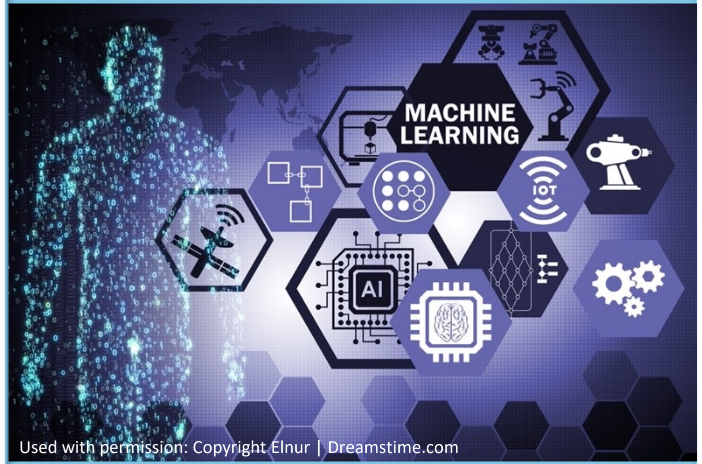
System-Wide Safety Project

SWS Overview

System-Wide Safety

Determining Safety Needs for Aviation Transformation

Developing New and Improved Safety Solutions

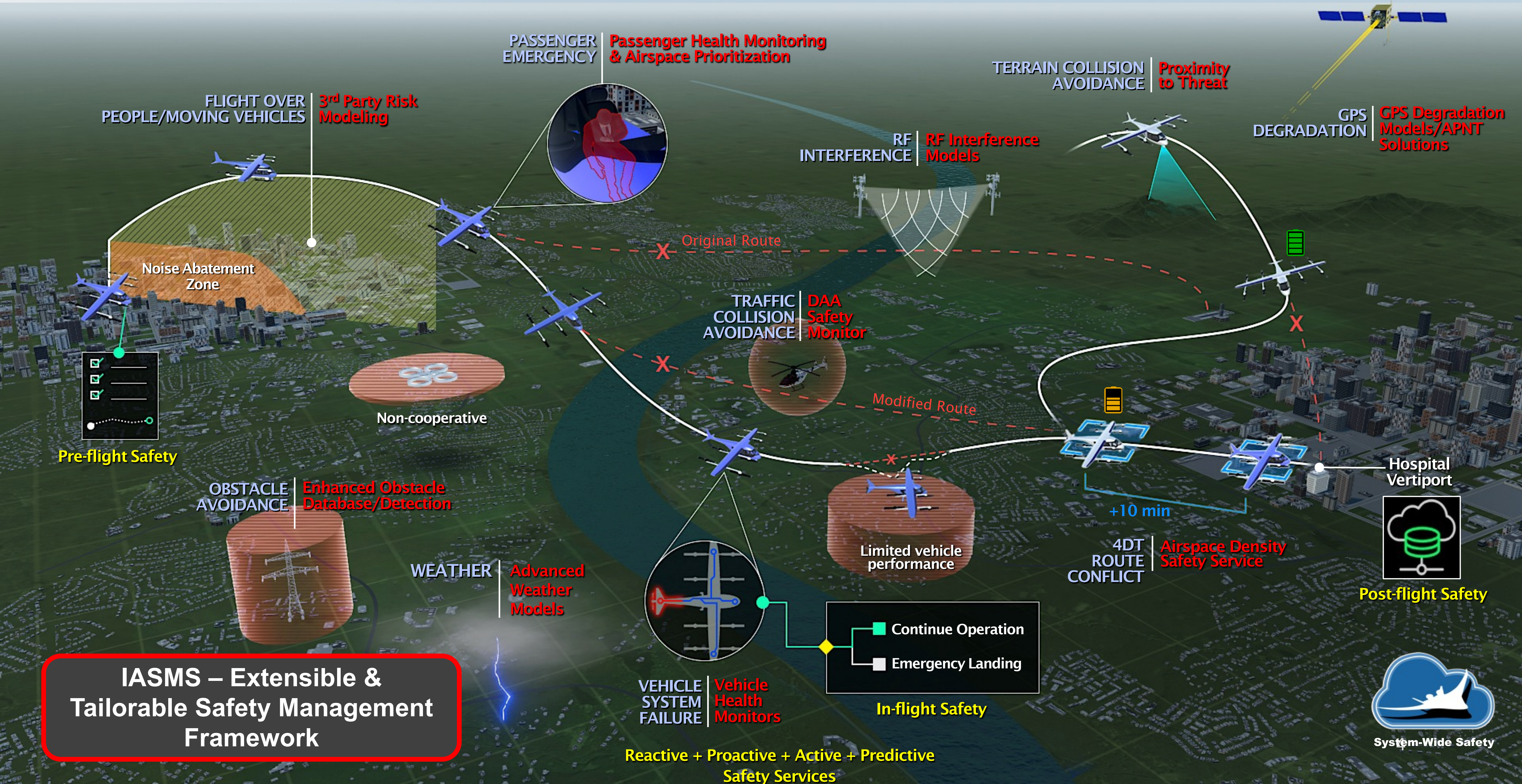


Disseminating Safety Knowledge and Technology

Sustainable aviation transformation through economic, environmental, and safety technology convergence.

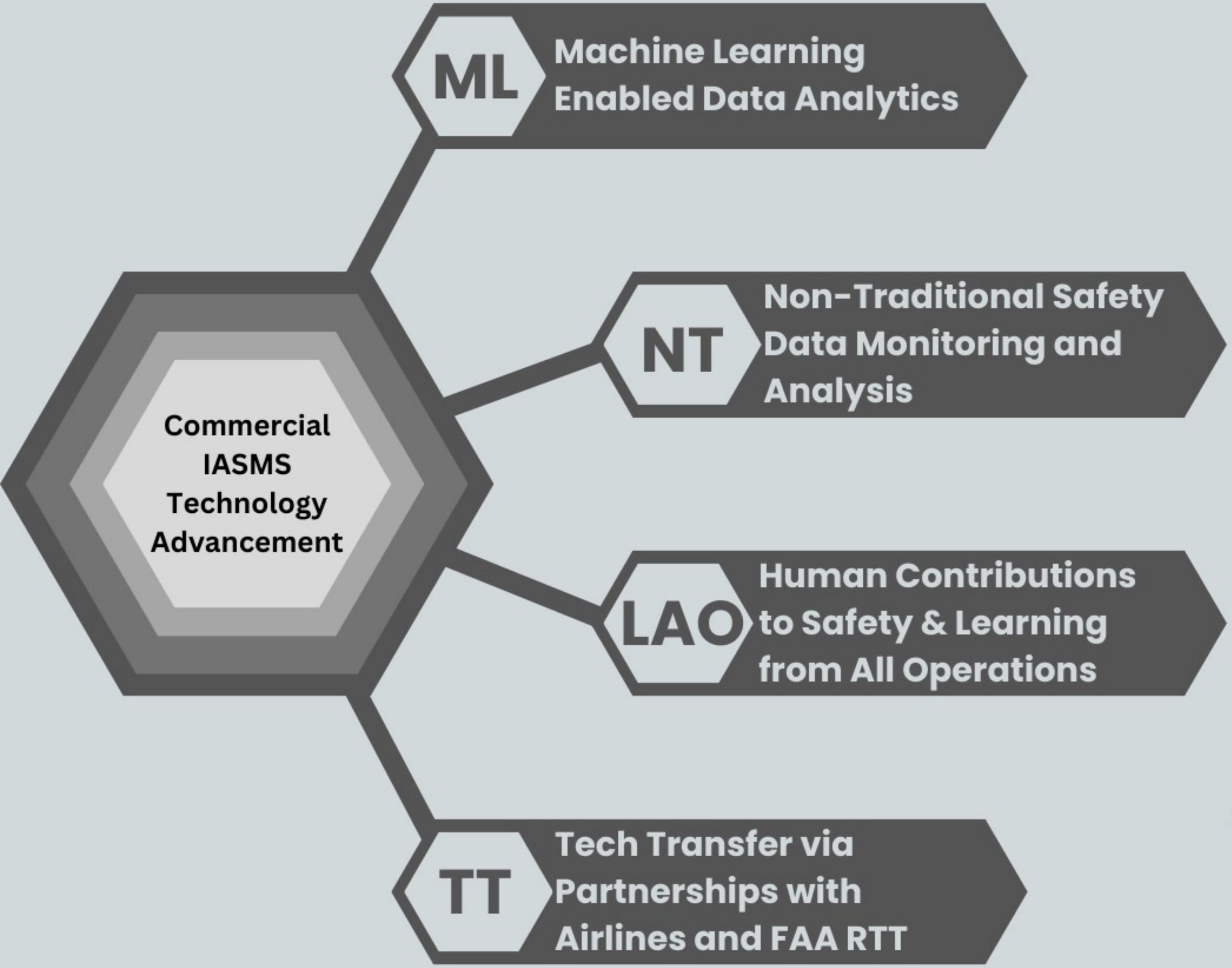
Credit: NASA

In-Time Aviation Safety Management Systems



TC-1

IN-TIME TERMINAL AREA RISK ASSESSMENT



Deliverable: Development of methods to improve air carrier SMS using ML and novel data sources.

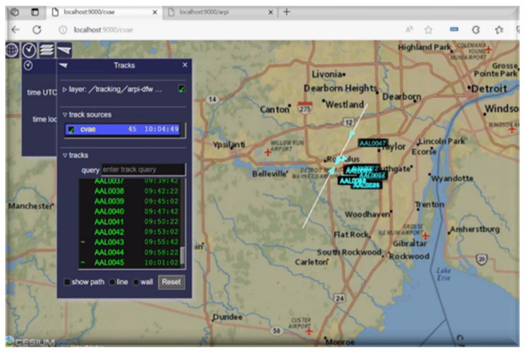
Impact: Airlines SMS has improved ability to predict and mitigate safety threats in-time to prevent accidents and incidents, with new insights on system-wide safety.

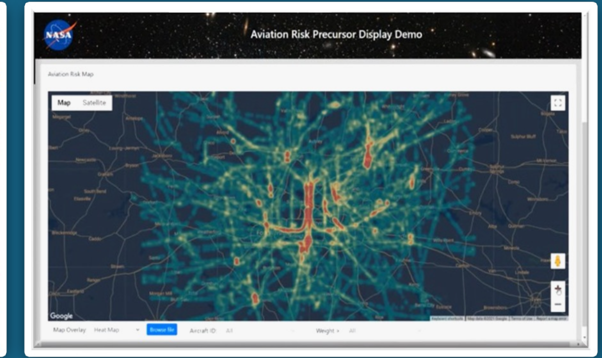
Stakeholders:

TC-1 Terminal Area Risk Assessment Accomplishments

Prototype Safety Dashboard:

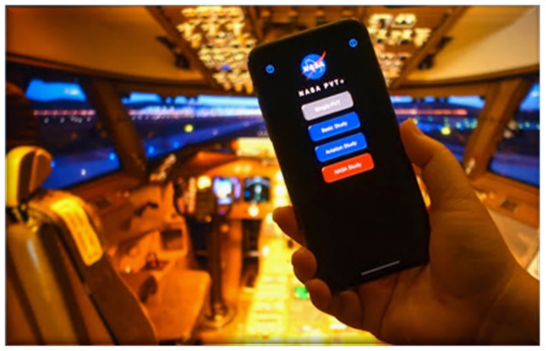
- The Aviation Risk Precursor Identification (ARPI) was developed by Booz Allen Hamilton as a prototype safety dashboard
- Sources System-Wide Information Management (SWIM) data
- Hosted a variety of SWS's anomaly and precursor detection algorithms
  - ➢ Results pointed to neural network which performed 36% better than a baseline predictor
- User interface provides visualizations of risk events for terminal area operations





PVT+ App Being Used on ISS:

- The Psychomotor Vigilance Task+ (PVT+) fatigue monitoring iOS application was developed by SWS to simplify fatigue measurements in the field
  - ➢ Reaction time task and eases logging of sleep schedules
- Now being used by astronauts aboard the International Space Station for a European Space Agency study
- The PVT+ is sensitive enough to detect performance changes due to sleep loss and circadian disruptions
- Currently being used by numerous airlines to assess crew performance in real-world operational scenarios



Directly aligns with ARMD's Strategic Thrust 5 to provide In-Time System-Wide Safety Assurance. In addition, the impact supports strategic industry partnerships, a key ARMD priority. Our partners include:

American Airlines · Boeing · Delta · easyJet.com · Federal Aviation Administration · Flight Safety Foundation · IATA · MITRE · Swiss · United

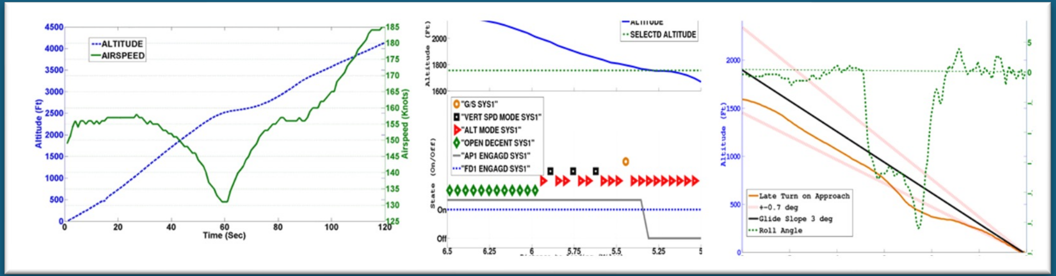
TC-1 Terminal Area Risk Assessment Accomplishments

Successful TC Closeout Event:

- SWS and NARI hosted hybrid event to share work accomplished under TC-1 with the aviation safety community.
- Guest speakers from NASA, FAA and American Airlines shared their perspective on the importance of the work.
- Topics covered included
    - ML/AI risk detection and prediction development and results
    - Research into how to measure human contributions to safety
    - Human fatigue and performance management and prediction
    - Demonstration of safety dashboard with risk prediction algorithms
    - SOTERIA experiment
- Audience of over 200 industry and academia leaders were in attendance

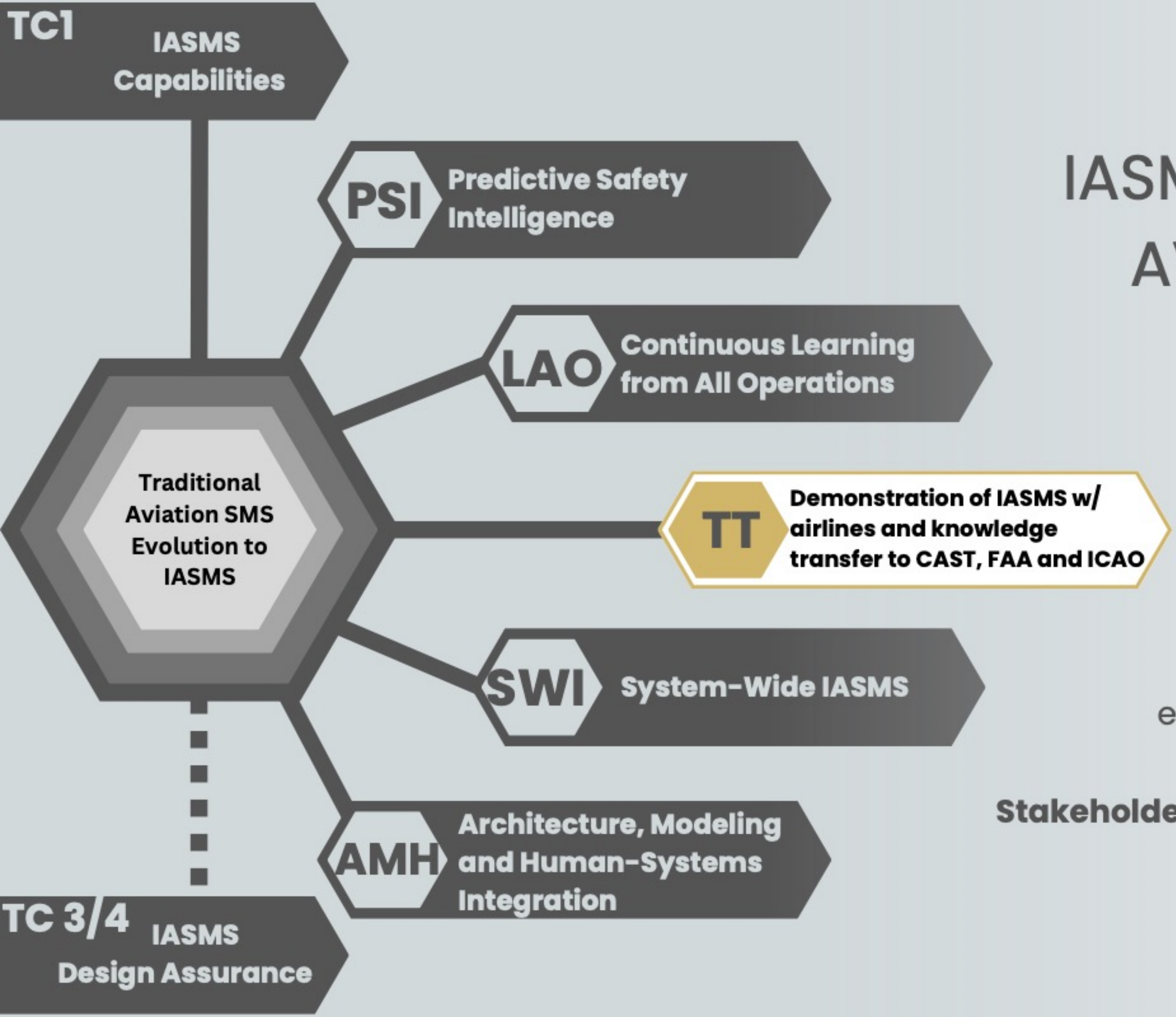
ML Tools Integrated into ASIAS:

- SWS-developed anomaly and precursor detection algorithms are being integrated into the Aviation Safety Information Analysis and Sharing (ASIAS) system
- ASIAS is used by the FAA and airlines to monitor operational safety.
- Early results detected 11 safety events and 35 false alarms
    - Drop in airspeed from slow rotation -> **not regularly detected with existing methods**
    - Possible mode confusion -> **25-35 seconds before speed exceedance**
    - Unstable approach -> **14 seconds before exceedance**



Directly aligns with ARMD's Strategic Thrust 5 to provide In-Time System-Wide Safety Assurance. In addition, the impact supports strategic industry partnerships, a key ARMD priority. Our partners include:

American Airlines, BOEING, DELTA, easyJet.com, Federal Aviation Administration, FLIGHT SAFETY FOUNDATION, IATA, MITRE, SWISS, UNITED



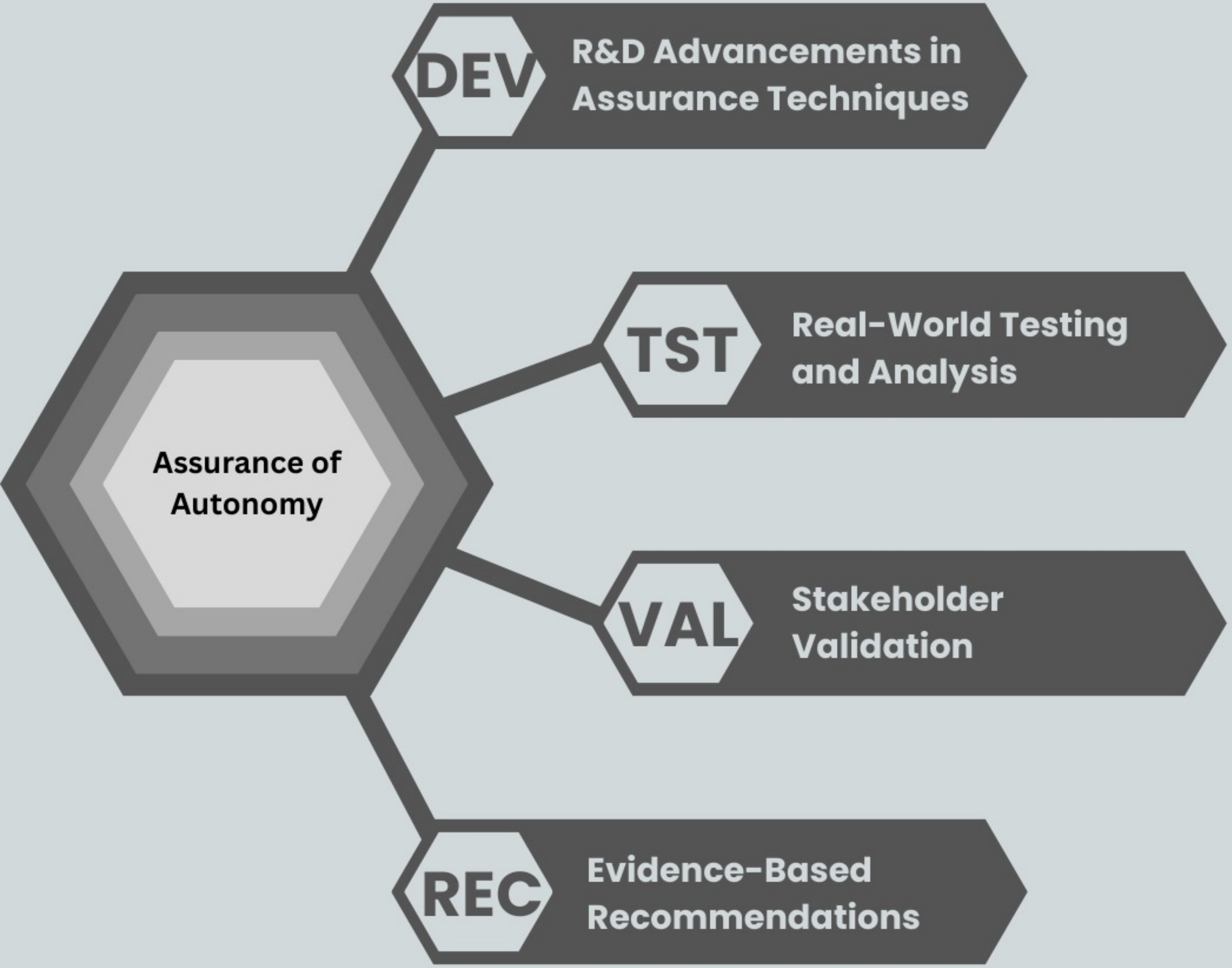
TC-6

IASMS FOR COMMERCIAL AVIATION OPERATIONS

Deliverable: Demonstration of effective strategies and technologies to predict and mitigate safety threats in-time to prevent accidents in an increasingly complex airspace.

Impact: Improved speed and characterization of system-wide risk identification to augment and evolve existing aviation industry SMS processes.

Stakeholders:



TC-4

COMPLEX, AUTONOMOUS SYSTEMS ASSURANCE

Deliverable: Final evidence and recommendations on process for certification of autonomous components in aerospace systems

Impact: Validated means of certifying complex autonomous systems

Stakeholders:

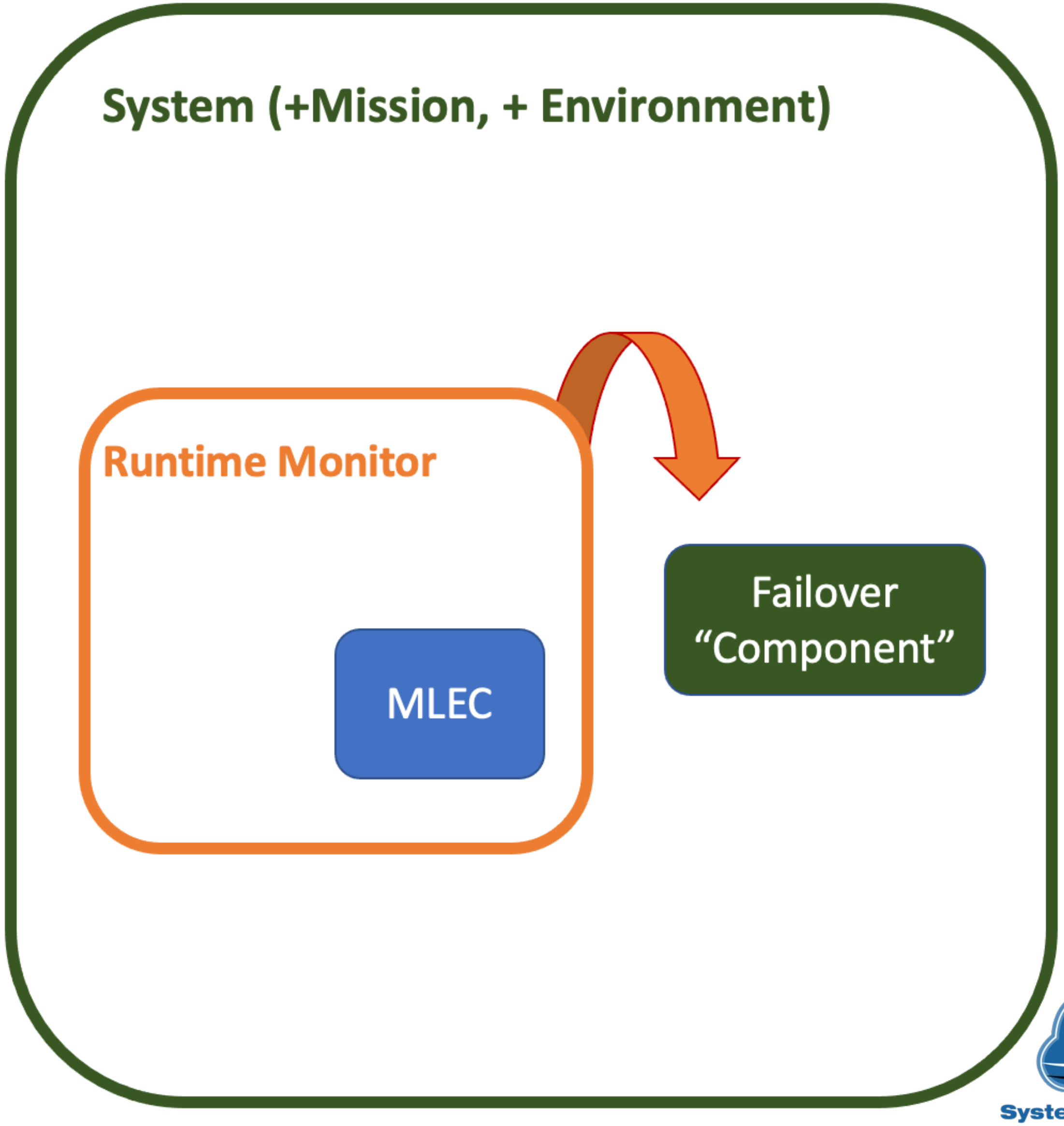
Assurance of Autonomy

GOAL: Demonstrate algorithms for checking safety standards for systems relying on untrusted components for autonomous surface operations and autonomous drone flight operations.

1) **Understand your Machine-Learning Enabled Components**

2) **Enforce Safe Limits Using Rigorously Assured Runtime Monitors**

3) **Demonstrate system-level assurance using high-level safety arguments that take contracts about the mission, environment, component, monitors, and failover plans into account**



System-Wide Safety

TC-4 Recent Accomplishments

*MLEC: Machine Learning Enabled Components

Goal

Demonstrate algorithms for checking safety standards for systems relying on untrusted components (MLEC) for autonomous surface and autonomous drone flight operations.

Feedback received on draft guidance for use of runtime assurance monitoring (FY22 API)
+
Formal analysis of closed-loop systems enabled by MLEC
+
Demonstrated abstraction of MLEC to enable formal verification methods
+
Demonstrated ability to express probabilistic safety requirements

FY23 R&D Advancements in Assurance Techniques

TC-2 IASMS SFC flight tests
+
Taxiway centerline tracking case study (DARPA use case)
+
Rover swarm case study

Applications for Validating Assurance Techniques

Enabled formal verification of closed-loop MLEC systems, which leads to scalability of assurance methods

Scientific Impact

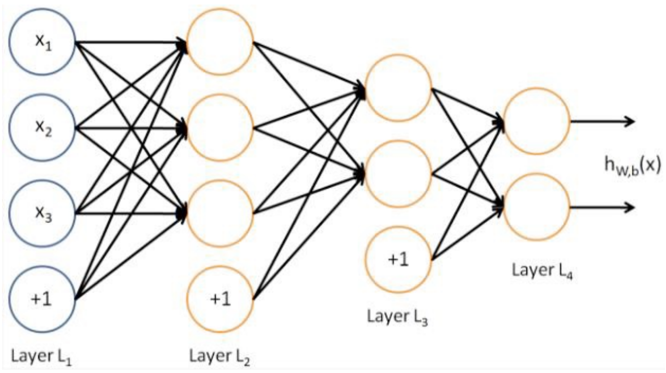
Demonstrable progress towards effective means of assuring complex autonomous systems necessary for certification.

Industry Impact

AI/ML assurance

Why do we need assurance?

- Problem
  - The inability to establish appropriate assurance for AI/ML components leaves us unable to effectively manage their risks and benefits.
    - Drives cost of development uneconomically high
    - Delays adoption of AI/ML at scale in safety critical systems
    - Results in unknown and unmanageable risks
- Goal
  - Discover and define what constitutes sufficient scientific-based evidence to substantiate a safety claim related to an AI/ML component performing a safety-critical function.



AI/ML will not see widespread adoption in safety-critical aviation systems until it is properly assured

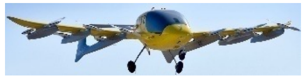
Beyond Autonomy/AAM Use Cases and Functions

BOEING

Aurora
FLIGHT SCIENCES
A Boeing Company

INSITU
A Boeing Company

wisk



and others

Autonomous platforms and infrastructure

Safety assurance technologies

Certification and regulations

Standards and guidance

Verification & Validation

Safety assessment

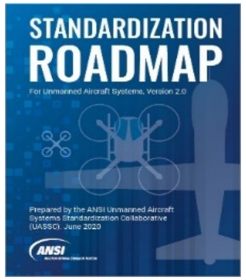
Risk mitigation

Contingency management

	Aircraft Requirements	Pilot Requirements	Airspace Requirements	Types of Operation
Part 107	UAS < 55 lbs.	Part 107 remote pilot certificate with small UAS rating	Airspace waiver or authorization for Class B, C, D, E airspace	VLOS, daytime, Class G, 400 ft., not over people (some regulations subject to waiver)
Section 333	As specified in exemption	Part 61 airman certificate	Blanket COA or Standard COA for specific airspace	UAS > 55 lbs.
Experimental Aircraft	Experimental Special Airworthiness Certificate	Part 61 airman certificate	Standard COA for specific airspace	Research and development, crew training, and market survey
Type Certificated Aircraft	Restricted type or special class certification	Part 61 airman certificate	Part 91 airspace requirements	Specified in operating authorization
Public Aircraft	Self-certification by public agency	Self-certification by public agency	Blanket COA or Standard COA for specific airspace	Public Aircraft Operations (AC 00-1.1A); UAS Test Site operations
Section 336 Model Aircraft	UAS < 55 lbs.*	Community-based organization (CBO) standards	Notification requirement within 5 miles of an airport	Hobby or recreational, VLOS, Section 336 operating rules, CBO standards

*Note: All UAS greater than 0.55 pounds aircraft must be registered (see part 47 and part 48 requirements).

Ron Matusiak, UAS Integration Office Overview, 2020



FEDERAL AVIATION ADMINISTRATION

NASA

EASA

RTCA

SAE INTERNATIONAL

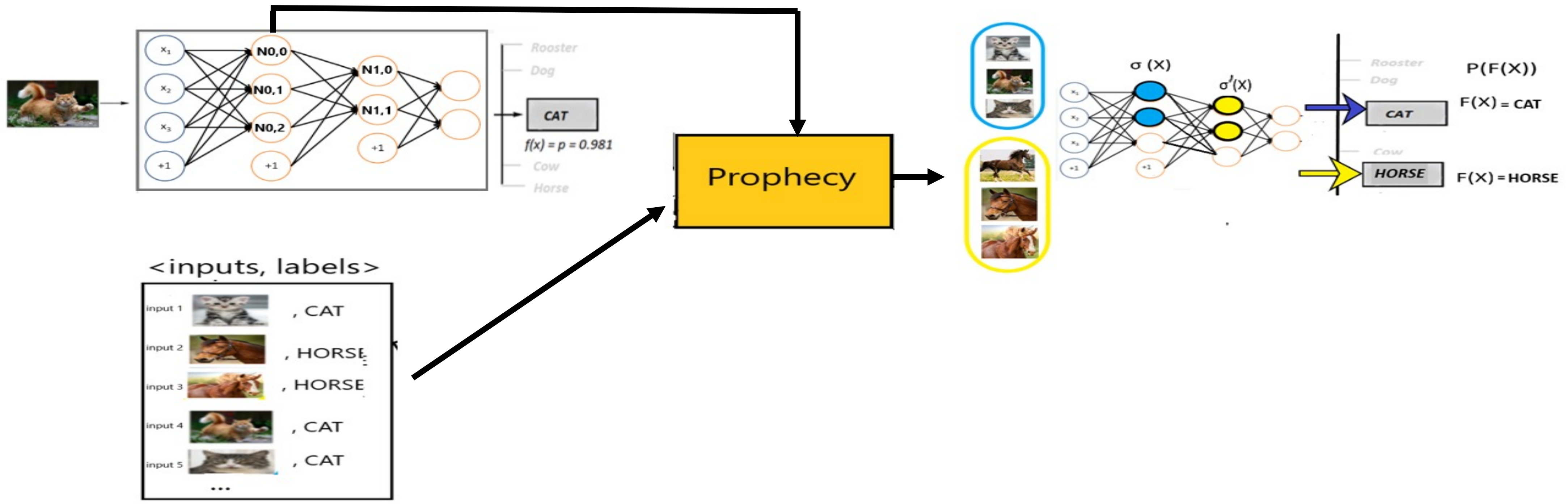
ASTM INTERNATIONAL

AI/ML assurance in SWS

- Regulatory aspects
  - Investigate means for certification
    - Overarching properties
    - Safety cases
- Standard and guidance
  - Participate in many relevant standard committees
- Safety assurance technologies
  - Hazard analysis and requirements
  - Data management
    - ODD
    - Simulation vs real data, model generalization
  - Formal verification of ML components
  - Advanced testing
  - Runtime monitoring
- Testbeds
  - In-house rover-based case studies using ML for vision

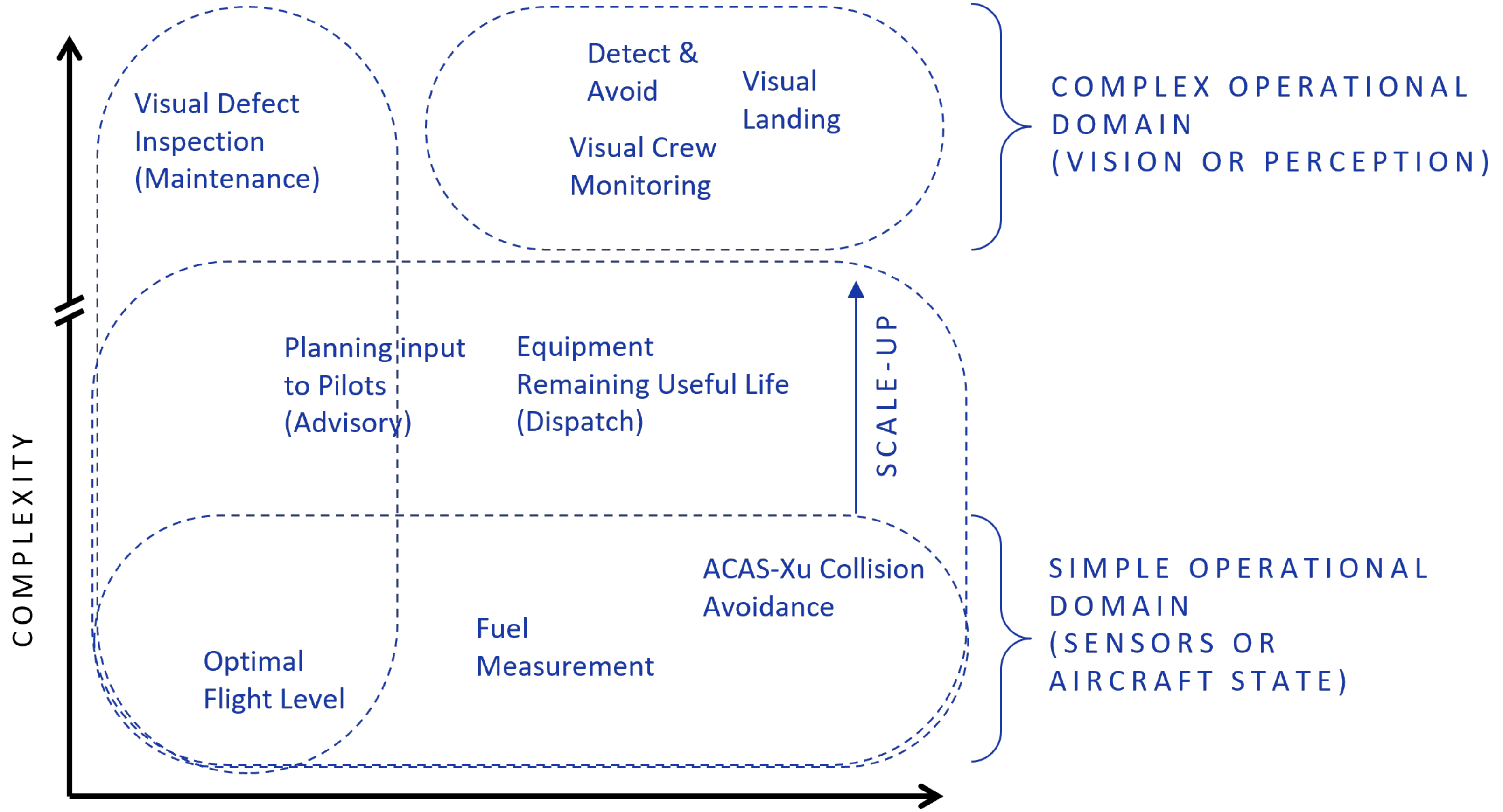
One example: Prophecy

- Decompose the complex DNN model into a *set of simple rules*, amenable to analysis
  - Assume-guarantee type rules are inferred from a trained DNN; $\forall x\ \sigma\ (x) \Rightarrow P(F(x))$
  - P is a property of the network function; functional property
  - σ (X) are formal constraints on neurons at inner layers of the network *(neuron activation patterns)*
  - *Prophecy:* Property Inference for Deep Neural Networks (ASE 2019)



Working through partnerships

- Industry
  - AAM entrants
  - Traditional aviation industry
- Government
  - FAA: collaborating on draft standard for AI/ML
  - DoT: assurance of AI/ML

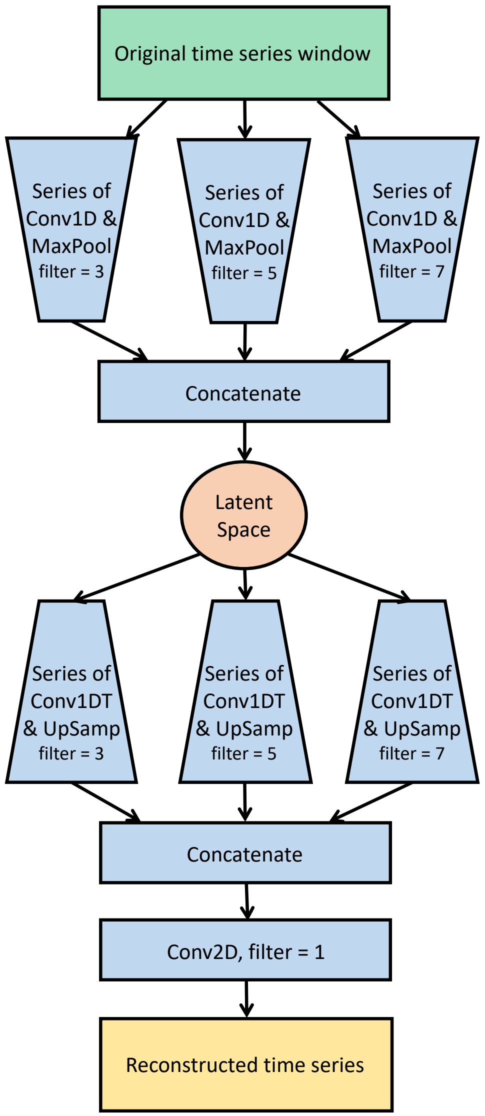


Main deliverables

- FY’22: Draft evidence and recommendations for the robustness of remote operations as a failover plan and the use of run-time monitoring
- FY’23 :
  - Modular framework to evaluate robustness of ML-enabled systems
  - Autonomy V&V 2045 roadmap
- FY’24: : Preliminary certification process for ML-enabled autonomous aerospace systems
- Beyond FY’24: consult with industry and regulatory bodies to establish new cycle of research

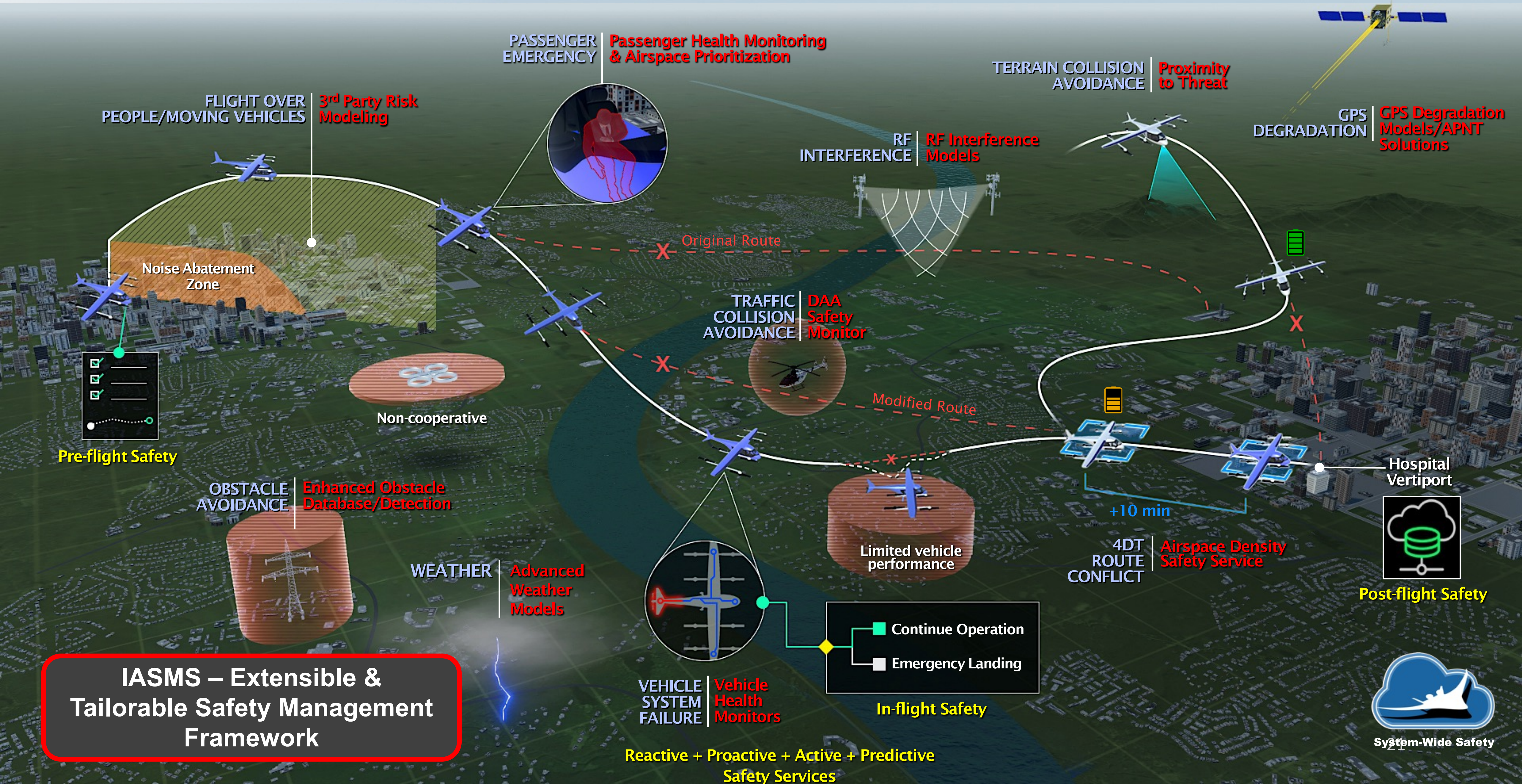
Other ML work in SWS

- Using ML technology to mine aviation incident reports and identify precursor patterns to incidents
- Identified patterns can be used during runtime monitoring
- **Convolutional Variational Auto-encoder:**
    - Using convolutional layers (instead of recurrent) to speed up the training process.
    - Using multiple filter sizes to capture local and global temporal dependence in the time series.

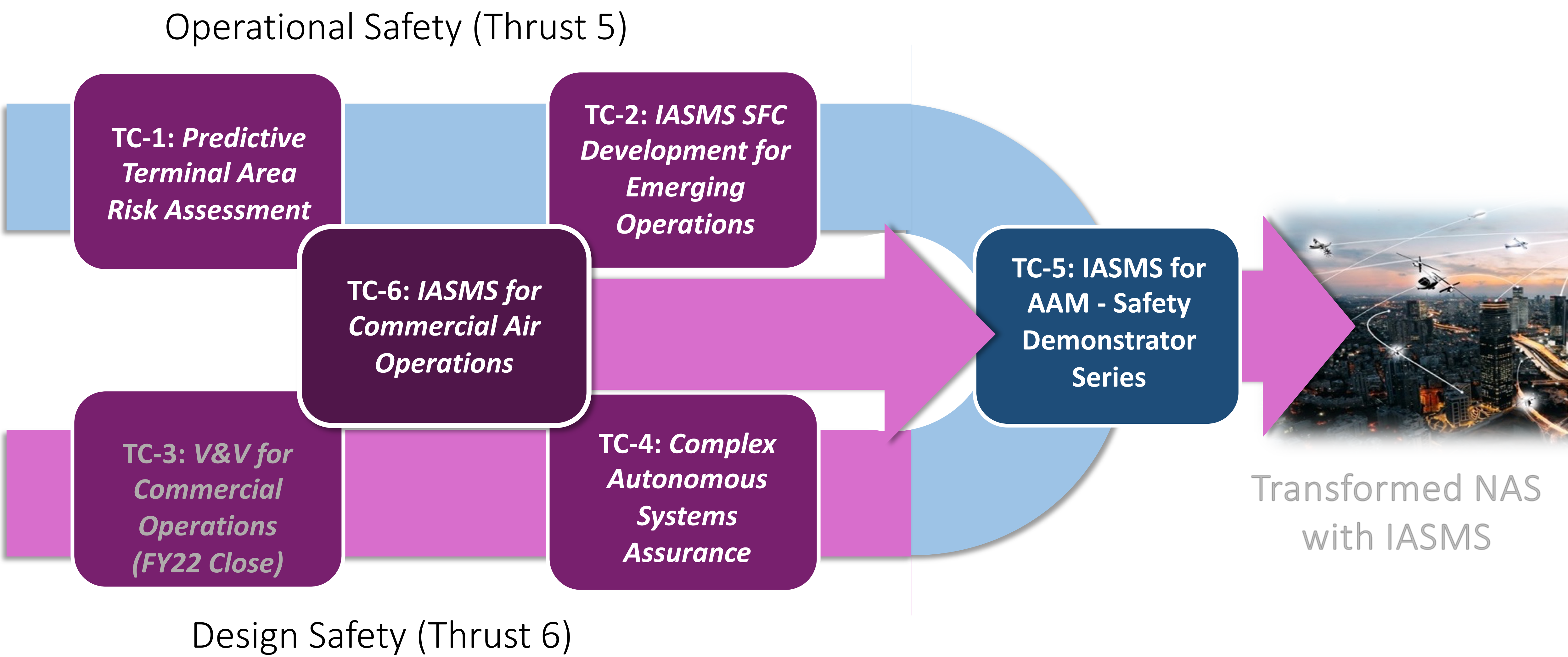


SWS Wildfire work

In-Time Aviation Safety Management Systems



SWS Project Research Portfolio



Safety Demonstrator Scheduled Progression

FY 23 — Wildland Firefighting

- ! HIGH
  Rural and partially evacuated area
- Complexity: LOW-MODERATE
  Intensive HMI and lack of commercial flights
- ? LOW-MODERATE
  Unknown location of fire; poor visibility

Environment:
Low Visibility, Smoke...

Vehicle & Mission:
sUAS, mid-size UAS/
Short Range

Human Role:
High

FY 27 — Hurricane Relief and Recovery

- ! MED
  Partially evacuated area
- Complexity: MODERATE
  Numerous agencies coordinating multiple relief efforts
- ? MODERATE-HIGH
  Unknown state of terrain; poor infrastructure

Environment:
Low Visibility, RF/EMF Hazards, Poor Weather…

Vehicle & Mission:
sUAS, mid-size UAS, large UAS/
Multiple Days

Human Role:
Medium

FY 29 — Emergency Medical

- ! LOW
  Urban area
- Complexity: MODERATE
  Regularly scheduled commercial flights
- ? MODERATE
  All weather operations

Environment:
Urban Airspace, RF/EMF Hazards…

Vehicle & Mission:
sUAS, mid-size UAS, large UAS/
Short to Long Range

Human Role:
Low

FY 32 — Urban Disaster Relief

- ! LOW
  Urban area
- Complexity: HIGH
  - - - - -
- ? HIGH
  - - - - -

Environment:
Degraded Infrastructure, RF/EMF Hazards…

Vehicle & Mission:
sUAS, mid-size UAS, large UAS/
Multiple Days

Human Role:
Multiple Simultaneous HMI paradigms

! Risk Tolerance   Complexity   ? Uncertainty

SWS: ODIN-Fire

Online Data Integration (ODIN)-Fire

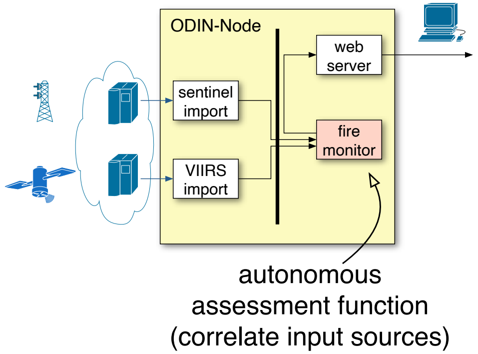
- Monitor for adverse conditions and enhance situational awareness

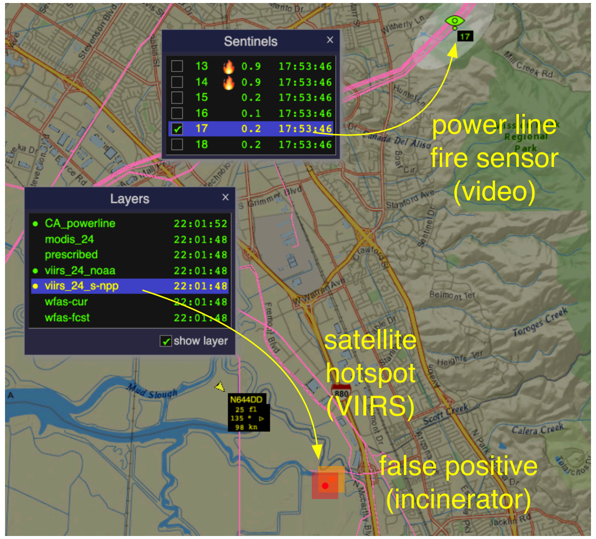
Approach

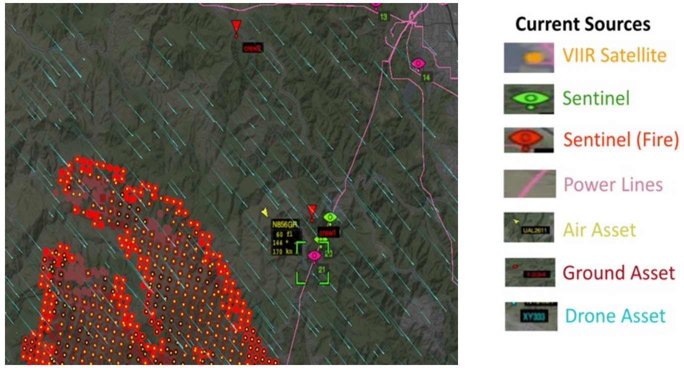
- Configurable dynamic display of multiple data sources
  - Open-source tool
  - Data sources such as fire/smoke detectors on powerlines, satellite heat data, 3D buildings, terrain, air traffic assets
  - Modeling capabilities such as fire spreading models, weather
  - Data age indication on single display
- Uniform architecture – many application types, scalable cross-platform
- Extensible component library

Impact

- Developed registry of Wildfire Fighting Data Sources
- SAA with Delphire
- Collaboration with USFS Rocky Mountain Research Station
  - FireLabs WindNinja micro grid wind forecasts







- Powerline sensors (Delphire), VIIR Satellite, Sentinel, live ADS-B Data (and other sources) integrated into ODIN
- Integrated weather data into ODIN via collaboration with WindNinja/USFS Firelabs

SWS: MIKA-fmdtools Dashboard

Relevant Dataset Identification

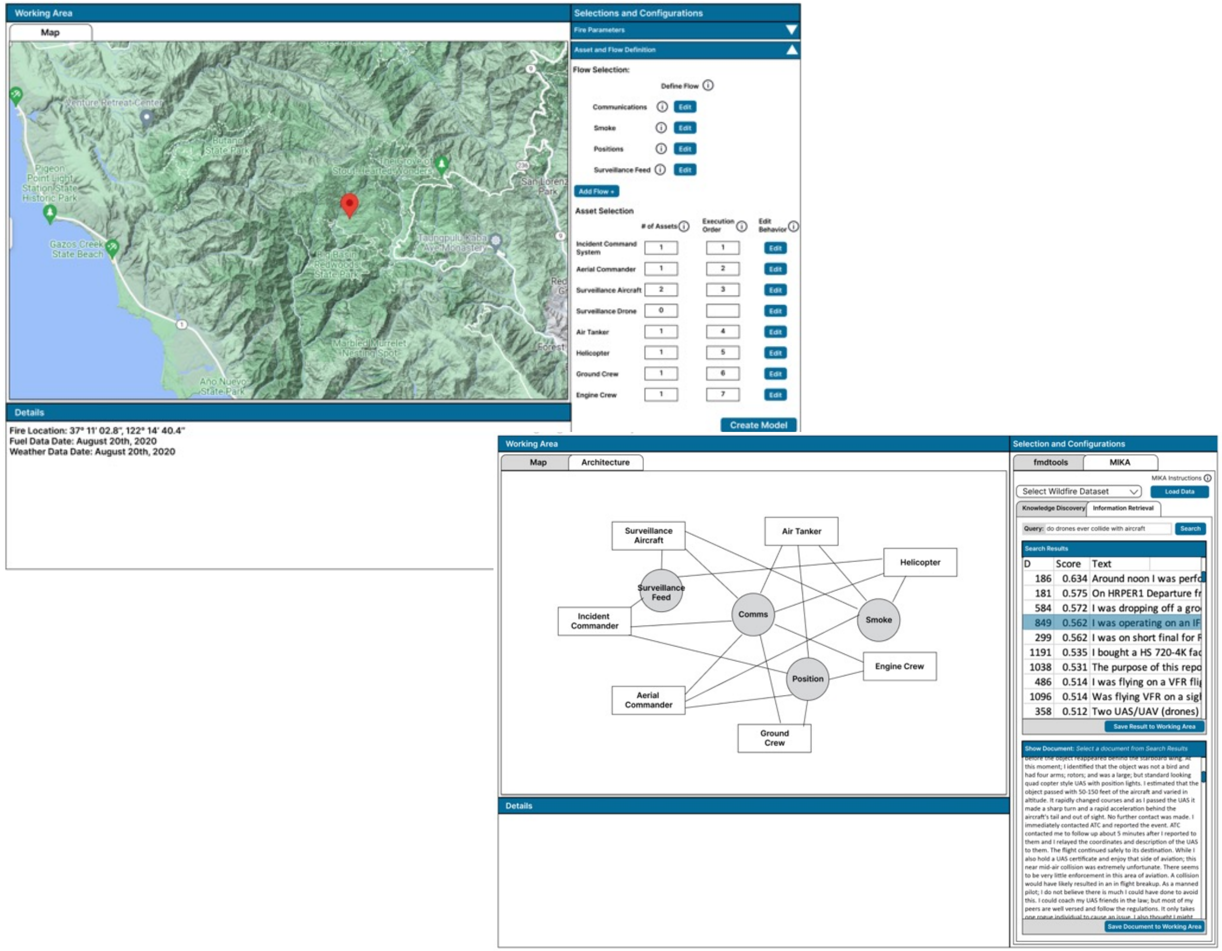
- Identified data sources of interest to operational goals throughout the wildfire management cycle
- Examined wildland firefighting accident and incident databases to assess their comprehensiveness

Approach

- Main data categories include
    - Fuels and ignition sources, climate and weather, detection and tracking, emissions and air quality, infrastructure and ecosystem impacts, object tracking
- MIKA-fmdtools Dashboard
    - Simulation functional model and scenarios
    - Test mitigation strategies
    - Inform modeling with MIKA lessons learned

Impact

- Demonstrated integrated prototype of MIKA-fmdtools Dashboard for end-to-end risk and resiliency analysis in early design
- MIKA approved for open-source release
- fdmtools 2.0 alpha released



MIKA-fmdtools: Manager for Intelligent Knowledge Access - Fault Model Design Tools