Welcome to the EASA AI Days 2025 High-Level Conference

27th – 28th August 2025

EASA
European Union Aviation Safety Agency

Your safety is our mission.

An Agency of the European Union

EASA
European Union Aviation Safety Agency

EASA AI Days 2025
Day 2 – 28th August 2025

Janet Northcote
EASA Head of Communications

Your safety is our mission.

An Agency of the European Union

EASA

European Union Aviation Safety Agency

EASA AI Days 2025

Innovation Partnership with Boeing (BEACON)

‘Investigation of an auto-taxiing AI-based system’

Moderator: Giovanni Cima

BEACON IPC Project Manager

Emily Howard

Boeing

Dragos Margineantu

Boeing

Matthew Jahn

Boeing

Joshua Neighbor

Boeing

Your safety is our mission.

An Agency of the European Union

Collaborative approach with all Stakeholders

AI/ML Guidance

Use Cases

(IPCs, MoUs, first applications)

SESAR 3 projects

Horizon Europe MLEAP (Machine Learning Application Approval)

CEN CENELEC JTC21

ISO/IEC SC42

Working group

WG-114/G-34

PROTECT

IPC = Innovation Partnership Contract
MoU = Memorandum of Understanding

BEACON IPC : EASA Perspective

- **Contract signed in 2023**
  - Assess possible ways to certify a Level 2A AI/ML based feature or system that could be installed on a transport category (CS-25) aircraft.
- **EASA team**
  - Project Manager/Concept of Operations
  - AI Expertise
  - Human Factors Expertise
  - Safety Assessment Expertise
  - Avionics Expertise
- **Objectives**
  - Exercise / test Concept Paper guidance
  - Assess two distinct AI/ML based auto taxi systems: first AI Level 2A (phase 1) and then AI Level 3A (phase 2).

- **Objectives achieved**
- **Main takeaways**
  - First time live demonstration in an IPC
  - Level 3A showed the potential for further exploration

EASA

The BEACON Project:
Boeing & EASA IPC on AI/ML Certification

Matt Jahn – Associate Technical Fellow
Dragos Margineantu – Senior Technical Fellow
Josh Neighbor – Senior AI Technologist
Emily Howard – Principal Senior Technical Fellow

August 28th, 2025

Introduction to The Beacon Project

Presenter

Matt Jahn **Global Safety & Regulatory Affairs** **Associated Technical Fellow** **Engineering Unit Member (E-UM)** 	**Presenter biography:** **Matt is an Associate Technical Fellow on the Autonomous Systems team within the Global Safety & Regulatory Affairs organization. He is also an Engineering Unit Member (E-UM), an FAA designated delegate with over 20 years of experience in certification. He focuses on certification strategy for autonomous systems to normalize their certification under Boeing's ODA and provides certification guidance to teams across the enterprise to help design certifiable systems and drive regulatory advancement of Artificial Intelligence and Machine Learning (AI/ML).** **Matt works with governmental agencies, global regulators, and standards organizations to establish regulatory policies and industry standards. His work includes working with NASA to explore technologies to assure safety of AI, technical exchanges with the FAA on the certification approach for safety assurance of AI/ML in critical airborne systems, certifiable AI roadmap, SAE G-34 committee (Artificial Intelligence in Aviation) and supporting Boeing's eVTOL subsidiary Wisk with their certification efforts.**

Agenda

- EASA perspective on this collaboration
- Introduction to The Beacon Project
- The automated taxi system
- Validation & Verification of AI systems
- The safe runway system and technical discussion
- Human Factors considerations, and Systems-Theoretic Process Analysis
- Wrap-up and next steps
- Q&A

System Demonstration

Boeing Efforts in Regulatory Innovation

WHAT DO WE DO?

Boeing coordinates the enterprise’s regulatory-innovation engagements with global regulatory authorities.

We work with authorities to find solutions that address open industry-wide regulatory challenges associated with innovation and emerging technologies.

We share what we are learning to advance the realization of aerospace innovation globally, in a safe and harmonized way.

HOW WILL WE WORK?

- With a safety mindset
- Collaboratively
- Respectfully
- Transparently
- Inquisitively
- Flexibly

WHAT ARE OUR STRATEGIC OBJECTIVES?

STRENGTHEN
regulatory relationships

INFORM
the regulatory ecosystem

FOSTER
global regulatory alignment

BUILD
Boeing’s regulatory knowledge & capability

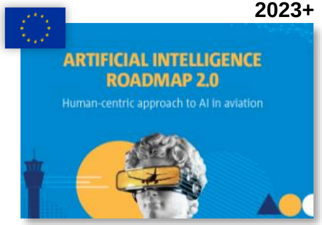
Collaboration Snapshot

2023 +

Advanced Air Mobility Regulatory Pathways

Innovation Partnership Contract undertaking a regulatory gap assessment for self-flying remotely supervised Urban Air Mobility passenger carrying operations.

2023+



Artificial Intelligence Certification Pathways

Innovation Partnership Contract exploring a Boeing use case application to inform future EASA AI Guidance Material.

2019+

UK CAA Airspace Modernisation Goals

UK Sandbox collaboration informing routine Beyond Visual Line of Sight Operations in Class G Airspace via a use case application.

2023+

CASA Airspace Integration

Collaboration exploring routine pathways toward integrated uncrewed operations in Controlled Airspace, informing future Digital Flight Rules.

2022+

Oceania Safety Information Sharing

An industry initiative to design and implement an aviation safety data sharing program in the Oceania region.

ECCN: 7E994

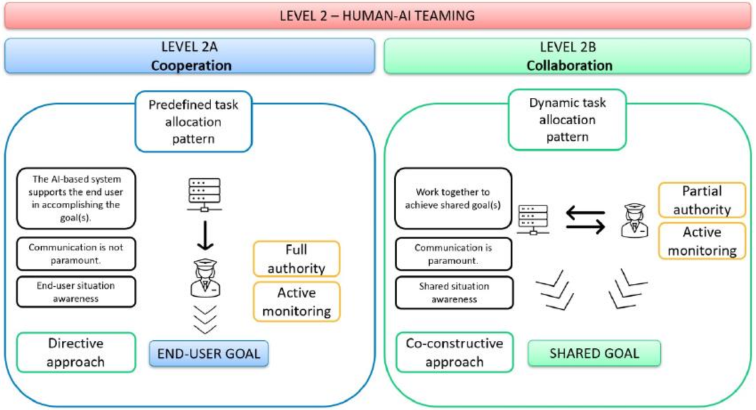
IPC Overview

IPC Overview

The IPC will:

- In collaboration with EASA, establish the regulatory requirements, means of compliance, and V&V strategy for an ML-based system
  - The effort will use the EASA Artificial Intelligence Concept Paper issue 2 as the basis for these requirements and MOCs
- Use Boeing's experimental automated taxi system as the surrogate for the certification process
- Consider both a level 2A (human/machine teaming) system and a level 3A (more autonomous machine) system, per EASA's leveling scheme
- Begin June 2023, and last approximately 18-20 months
  - Expected deliverable: a published report which documents the efforts and findings

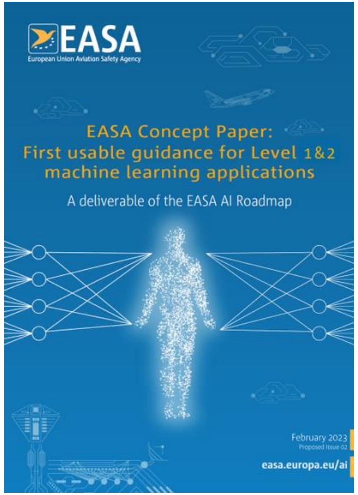
Level 1 AI : assistance to human	Level 2 AI : human-AI teaming	Level 3 AI : advanced automation
•Level 1A: Human augmentation •Level 1B: Human cognitive assistance in decision and action selection	•Level 2A: Human and AI-based system cooperation •Level 2B: Human and AI-based system collaboration	• Level 3A: The AI-based system makes decisions and performs actions, safeguarded by the human. •Level 3B: The AI-based system makes non-supervised decisions and performs non-supervised actions.



IPC Overview

Why did Boeing propose this IPC?

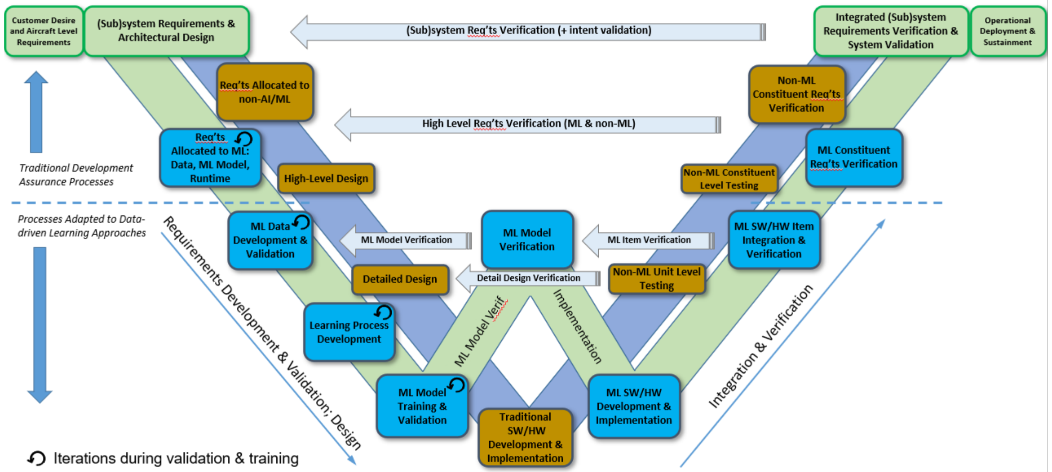
- The exercise of applying the Concept Paper to a level 2A and 3A system, off the critical path of certification, will allow exploration of topics of interest
- The work will also highlight potential areas of refinement for future issues of the Concept Paper
- The IPC will lay the groundwork for future certified AI systems
- It will also allow Boeing to help contribute to and build upon the body of work that has been created in other IPCs
- We hope to leverage the IPC work to facilitate harmonization amongst regulatory agencies and Standards Orgs including EASA, FAA and others.



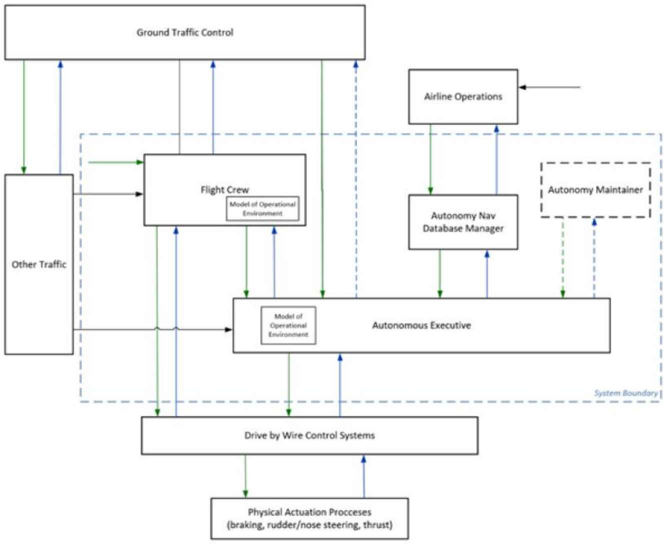
IPC Focus Areas

- Application of Concept Paper Objectives and MOCs
- Validation and Verification approach
- Human Factors, and the use of System-Theoretic Process Analysis (STPA)

AI/ML Development Lifecycle (MLDL) in context with traditional Engineering "V"



High Level Control Structure for existing system



ECCN: 7E994

IPC Results

Concept Paper Objectives Review

Boeing conducted a review of the objectives in the Proposed Issue 2 to:

- Establish which objectives are applicable to the automated taxi system
- Explore if the intent of the objectives were clear and an applicant could show compliance to
- Ensure the set of applicable objectives were sufficient to ensure the safety of an AI system developed to meet them

Major themes of the review:

- The Concept Paper objectives are intended to be an additional layer on top of existing regulations, addressing the specific aspects of ML that are not covered in those existing regulations
- Additional guidance will be needed from the regulator as to what is “good enough” with regard to qualitative requirements
    - And will likely be system specific
- Additional guidance will also be needed as to the expectations around the continuous safety assessment
- Unsafe deskilling
- The list of objectives that are applicable to Level 2A vs 2B may need to be refined
- ORG objectives are at DOA level, not project level

Definition of End User: Is ATC an End User of the automated taxi system?

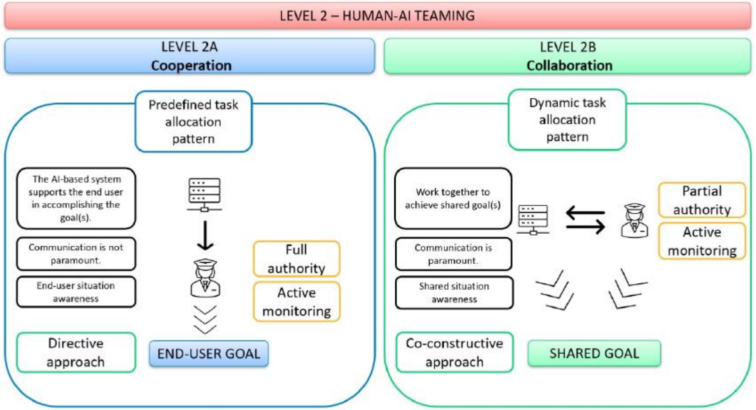
The interaction between the automated taxi system and ATC presented an interesting use case to explore the Concept Paper's definition of an "end user"

- ATC interacts directly with the automated taxi system via radio communication
  - Taxi clearance provided, readback received
- Does this direct interaction lead ATC to be considered an end user, and thus subject to end user related objectives?
- Concept paper definition of end user excludes those who support the product
  - Automated taxi system is designed for seamless integration into existing aviation system, no change to ATC procedures or airport environment
  - Due to this, consensus was reached that ATC supported the product rather than being an end user
- However, consensus was also reached that voice comms related end user objectives should still be applicable to ensure effective communication
- This exercise led to the understanding that the objectives are a good start, but may require tailoring to the specific system

Leveling Discussion

The automated taxi system was determined to be a Level 2A system

- Pre-defined task allocation scheme, no dynamic task allocation
- System functionality includes automated readback of clearance
  - This could remove some “authority” from end user, leading the system to be considered 2B due to partial authority of the end user
- Discussions concluded that because the end user can still overrule the system by disconnecting or transmitting a contravening radio call, they retain full authority
- Area of future exploration: Situational awareness/representation
  - Level 2B requires a shared situational awareness between system and end user, but Level 2A does not
  - However, for the end user to effectively monitor and collaborate with the system, they must be provided with the system’s situational representation (through various types of explainability)
  - This is a suggested area for further research
- Area of future exploration: Impact of information availability on AI level
  - If a system utilizes information not available to the end user, does that affect AI level classification?
  - This is a suggested area for further research (will be explored in Phase 2)



Validation & Verification of AI systems

Presenter

Josh Neighbor **Boeing Artificial Intelligence** **Senior AI Technologist** 	**Presenter biography:** **Josh Neighbor is an AI Technologist at Boeing, where he leads the Verification and Validation for AI (VV4AI) team and serves as a Machine Learning Engineer for the KC-46 Autonomous Air-to-Air Refueling (A3R) project. He is the Principal Investigator for collaborative research with Carnegie Mellon University and the University of Illinois at Urbana-Champaign, focusing on perception robustness and safety contracts for AI-enabled systems. Josh is a named inventor on two granted and four pending patents related to autonomous refueling, and he was recently published in ACM Transactions on Cyber-Physical Systems for his work on certifying robustness of learning-based perception methods.** **With a background spanning AI/ML, computer vision, controls, and autonomous systems, Josh focuses on bridging research and engineering to enable the future certification of safety-critical aerospace AI systems. He holds advanced degrees from UC San Diego and Georgia Tech. Outside of work, Josh enjoys skiing, mountain biking, and hiking.**

Overview – V&V and Assurance

- The automated taxi system
- Introduction & Motivation
- The Challenge: AI in Safety-Critical Aviation Systems
- Regulatory Landscape & Standards
- Verification & Validation (V&V) for AI/ML Systems
- Defining the AI Constituent & operational design domain (ODD)
- Performance Metrics & Risk Management
- V&V Lifecycle
- Human Factors & Explainability
- Case Study: Automated Taxi System
- Key Takeaways & Next Steps
- Q&A

The automated taxi system

Automated taxi system Overview

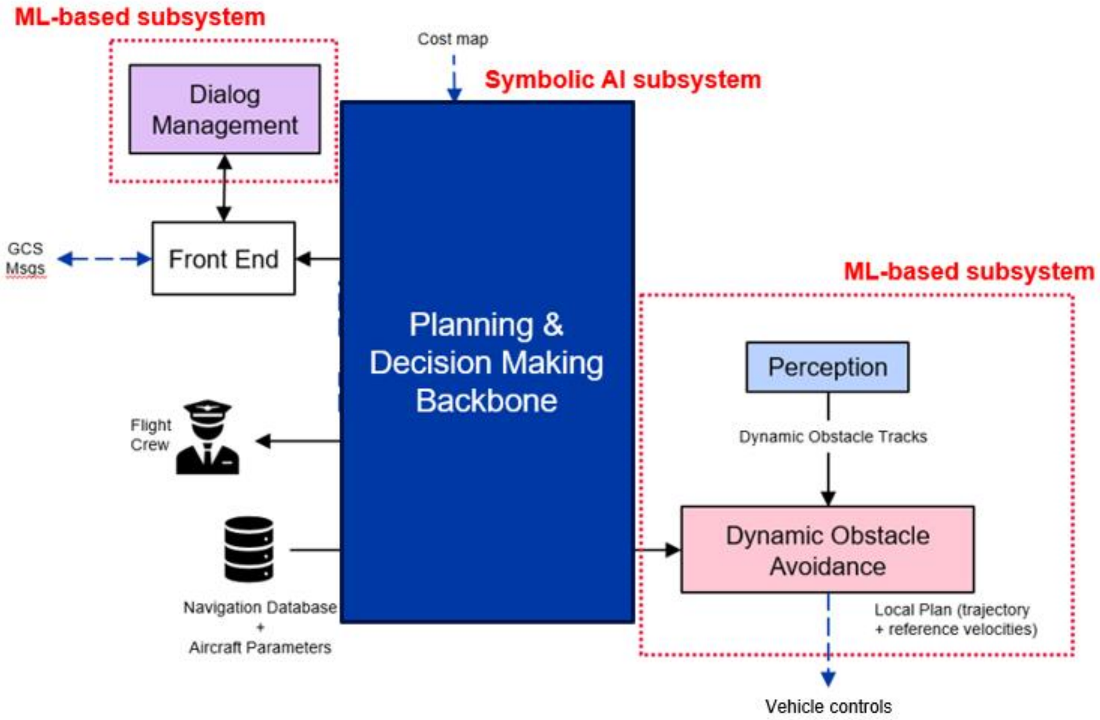
- The automated taxi system is an experimental system being researched by Boeing
- It is capable of:
  - Receiving a taxi clearance via radio
  - Parsing that clearance, planning a taxi route, and providing a readback
  - Executing the taxi plan to autonomously taxi the aircraft from one location to another
  - Using its perception system to localize itself on the airport map
  - Using its perception system to sense, classify, and avoid obstacles
- The flight crew monitors the automated taxi system and retains the ability to override and disconnect the system at any time.

Flight Crew Oversight

- The crew will be responsible for:
  - Activation of the system
  - Monitoring the execution (through the provided interfaces)
  - If needed, entering the taxi destination and specific route requirements
  - Monitoring all aspects of the system
  - Overriding the system if abnormal operation or hazards are identified by the crew

- The automated taxi system will provide the flight crew the necessary information in order to monitor the system
  - This information display will be handled by the systems interface with the flight crew

Automated taxi system overview



28

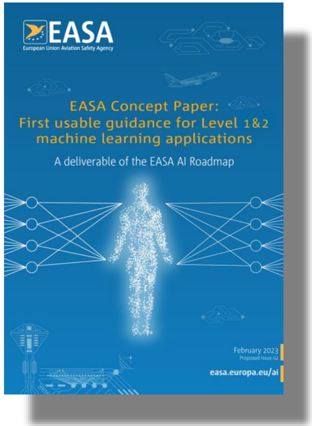
ECCN: 7E994

Introduction & Motivation

- Aviation is one of the safest forms of transportation.
- AI/ML is transforming aviation, promising efficiency, cost savings, and reduced human error.
- *But:* The public and regulators demand that AI-based systems are as safe—or safer—than traditional systems.
- Certification and assurance are critical for trust and adoption.
- *Purpose:* To share some our progress and lessons learned for V&V of safety-critical AI in aviation.

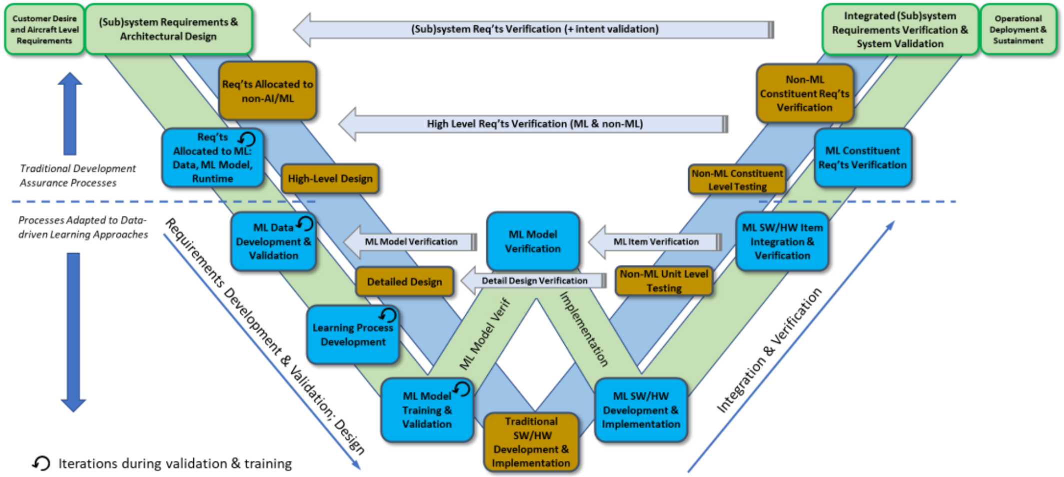
Regulatory Landscape & Standards

- *EASA AI Roadmap & Concept Paper*
  - Levels of AI (1A, 1B, 2A, 2B, 3A, 3B)
  - Human-AI teaming, authority, oversight
- *Key Standards:*
  - ARP4754A (System Development)
  - DO-178C (Software)
  - AIR6988 (AI Statement of Concerns)
  - ISO/IEC TS-4213 (ML Performance Assessment)
- *Ongoing Efforts*: Evolving standards for AI/ML in safety-critical systems

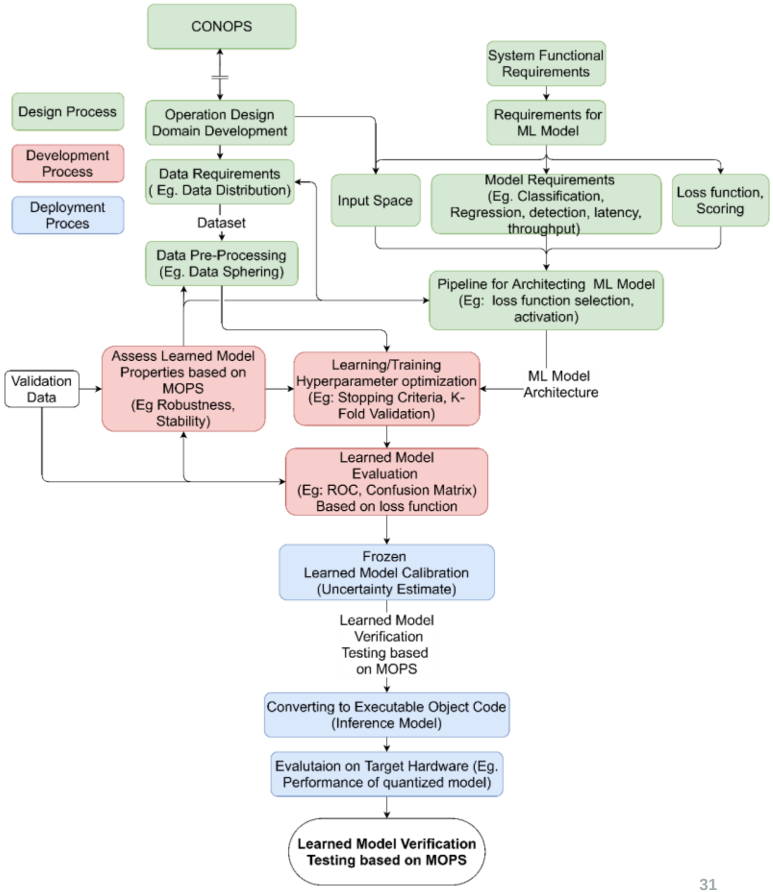


Boeing's Development Lifecycle

AI/ML V-W Development Lifecycle (MLDL) in context with traditional Engineering "V"



Learned Model Development Process



Terminology

There are two types of functions:

Exact: All possible inputs and outputs can be enumerated and verified. These can be certified using traditional methods.

Approximate: Solutions are estimated due to computational intractability, which cannot be exhaustively verified

Tractability describes whether a computational problem can be solved exactly within reasonable (polynomial) time. Many real-world AI problems are intractable, necessitating approximate solutions.

- **Traditional software certification (e.g., DO-178)** is well-suited for exact solutions to tractable problems.
- **AI/ML models** typically provide approximate solutions to intractable problems, requiring new approaches for risk management and certification.

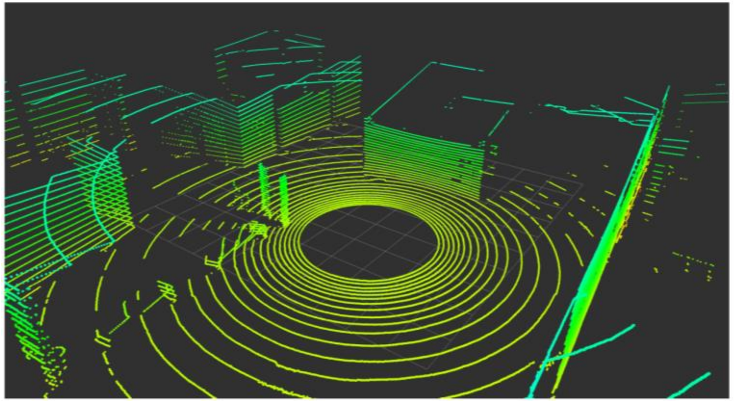
Central Role of Risk Management

What is the probability that the function will produce an output leading to an undesired or hazardous outcome?

Certification efforts must therefore:

- Identify the risks associated with approximate functions, including both known and unknown failure modes.
- Quantify the likelihood and impact of undesired outcomes, using statistical analysis, simulation, and empirical testing.
- Implement design and operational mitigations to reduce risk to an acceptable level, as defined by system safety requirements
- ODD defines the cost part of the risk.

This risk-based approach is consistent with the EASA Concept Paper



How risk management is handled in the Automated taxi

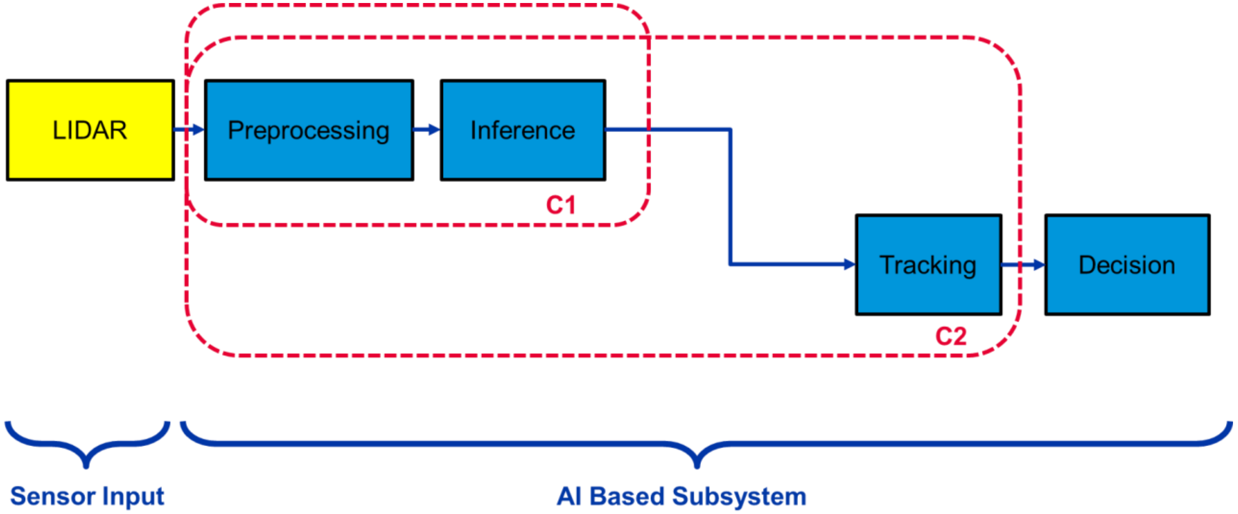
The overarching goal is to minimize the probability that an approximate function will result in a hazardous event, such as a collision.

- **Design-time analysis**
- **Empirical validation**
- **Operational monitoring and mitigation**

- The challenge with the certification of AI Systems is to **accurately** estimate the **risk** for executing approximate functions in conjunction with the actual hardware (primarily sensing and compute) that is employed.

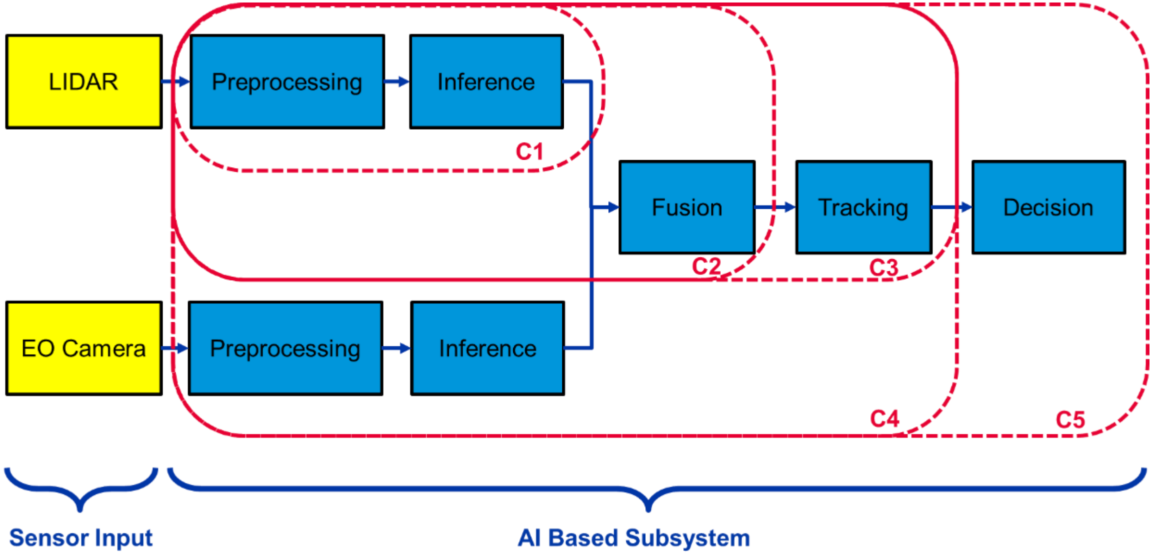
Defining the AI Constituent & ODD

- AI Constituent: The functional boundary of the AI/ML component
  - Impacts requirements, V&V, and certification scope
- ODD: How the AI Constituent impacts the ODD
  - Must be well-defined, validated, and monitored
- Challenges: High-dimensional input spaces, dynamic and rare scenarios



Multimodal perception stack

- Multiple Sensor inputs
- Impacts to ODD
- Support effective V&V and runtime monitoring,
- Facilitate integration with other system elements.



Importance of Defining the ODD

- Safety and Certification Requirements
  - Ensuring Compliance
  - Risk Management
  - Requirement Validation
- Algorithm Performance and Reliability
  - Improves Performance
  - Improves Reliability
- Hazard Identification and Mitigation
  - Predictable Failures
  - Run-Time Safety
- Explainability and Trustworthiness
  - Clear Boundaries
  - Enhanced Trust

Challenges in defining ODD

- Highly dynamic
  - Nonstationary
  - Highly compex
- Safe operation under all specified conditions
- Difficulties collecting sufficient and representative data
- Completeness through scenario-based testing is impossible

AutoTaxi Operating Domain (OD)

Operational Parameters	Relevant Operating Limits
Tarmac – quality / roughness (IRI); Slope	0 to 6 m/km with +/- 2% slope
Dynamic obstacle velocity and direction	20-35 mph in any direction
Material properties / reflectivity	To be determined
Orientation and size of objects	360 deg of freedom around object z-axis

Obstacles	Relevant Operating Limits
Obstacles – Vehicle (dynamic)	Aircraft, ground support equipment, shuttle buses, airport security vehicles, maintenance vehicles, service vehicles, emergency response vehicles
Obstacles – Other (static)	Buildings, terminal, construction areas, airport lighting, airport signage, debris (of a minimum size), runway arresting systems (e.g., barrier, cable, etc.)
Obstacles – Other (dynamic)	Ground crew, personnel, birds, other animals

Environmental Conditions	Relevant Operating Limits
Weather	Operates in clear weather, light to moderate rain, and light snow. Not reliable in heavy rain, heavy snow, or dense fog.
Lighting	Operates in daylight and nighttime conditions with airport tarmac lighting
Time of day (sun position)	Daytime + Nighttime
Time of year (season)	All seasons
Temperature	-20°C to 60°C
Humidity	10% to 90% non-condensing
Altitude	Sea level to 3000 meters
Terrain	Airport

Define AI/ML constituent performance metrics

How should we define performance metrics for Perception AI component?

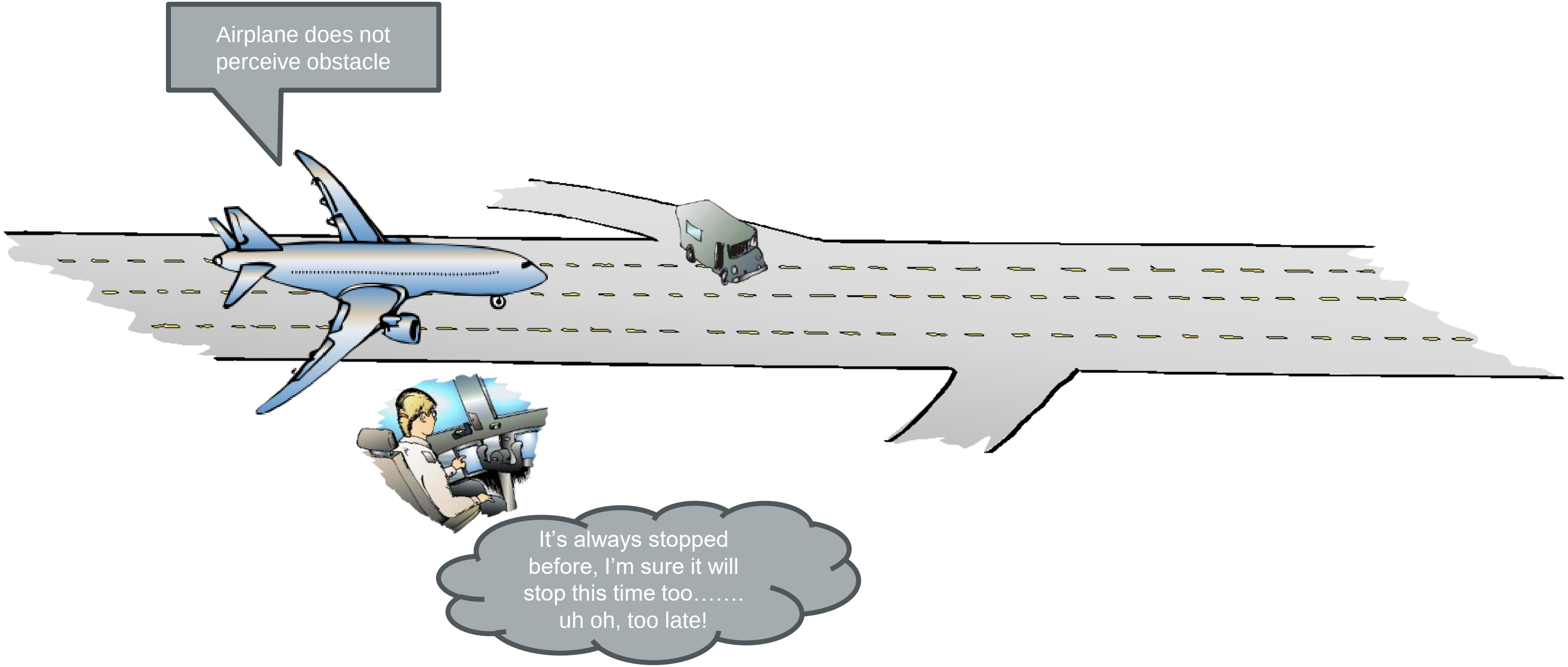
The system needs to Detect, Classify and Track

Detection:
- How far are we from our 'margin' to say there is no object here?
- What would be the distribution?
- How much noise can it take? False positive rate etc.?

Classify / Predict:
- Effects of misclassification
- How much noise can it take? False positive rate etc.?

Track :
- How does noise effect us over time
- How fast can we track something
- What properties would an object need to have to be missed?

Scenario 1: Unannunciated loss of obstacle perception

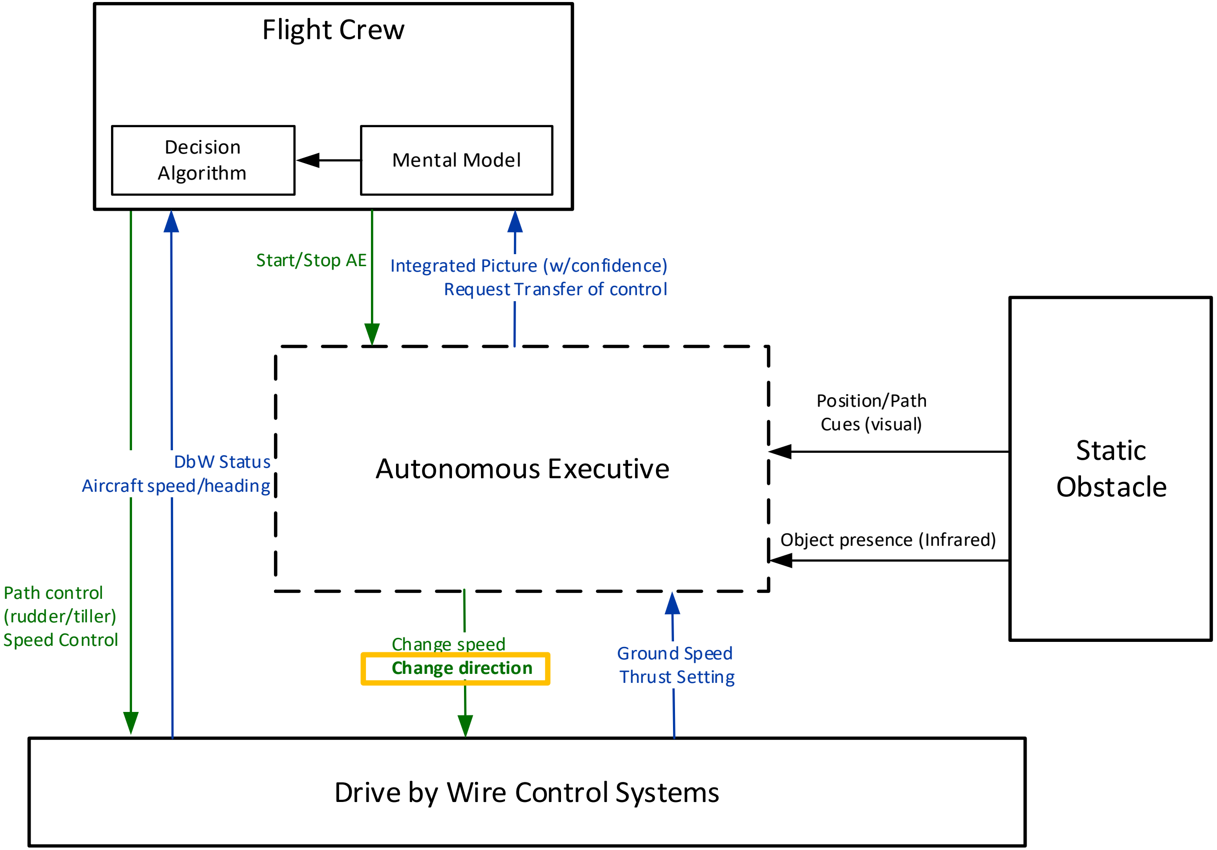


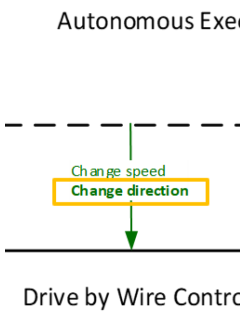
ECCN: 7E994

STPA Hazards and System Constraints

ID	Description	Linked Losses
H1	Airplane departs designated taxiway	L1-L7
H2	Airplane enters unauthorized/uncleared taxiway or runway	L1-L7
H3	Airplane exceeds proximity threshold to relevant objects (airport infrasructure, people, vehicles, other aircraft)	L1-L7
H4	Airplane exceeds expected maximum speed constraints	L1, L2, L3, L5, L6
H5	Airplane refuses to move or moves too slowly (e.g. due to false detection of object)	L1, L2, L3, L5, L6
H6	Autonomy exhibits otherwise safe but unexpected/uncontrollable behavior	L1, L4, L5, L6
H7	Autonomy system creates sequence of decisions which lead to only negative outcomes	L1-L7
H8	Autonomy creates excessive/additional workload for humans?	L1, L5, L6

ID	Description
SC1	Aircraft must remain on paved, active airport surfaces (e.g. taxiway, gate area, fuel apron, deicing station, etc)
SC2	Aircraft must remain on paved, active airport surfaces (e.g. taxiway, gate area, fuel apron, deicing station, etc)
SC3	Aircraft must maintan a minimum proximity threshold from relevant objects (TBD ft from other aircraft, TBD ft from airport infrastructure and vehicles, TBD ft from people)
SC4	In auto-taxi system armed mode, the airplane must not exceed max speed constraints (25 knots straight line and 10 knots turning)
SC5	The auto-taxi system shall taxi at a reasonable speed when there is no obstacle in the way (need to quanitfy reasonable)
SC6	The auto-taxi system must provide adequate feedback (e.g. current goals and algorithm for achieving goals, "Explainability") to the human supervisor in order to maintain situational awareness at all times, without increasing human workload. ~~for them to understand the reasoning behind autonomous behavior.~~
SC7	Decision sequencing must always lead to positive outcome
SC8	Autonomy must never adversely affect or create additional workload for humans


Flight Crew
Decision Algorithm
Mental Model
Start/Stop AE
Integrated Picture (w/confidence)
Request Transfer of control
Autonomous Executive
Position/Path Cues (visual)
Object presence (Infrared)
Static Obstacle
DbW Status
Aircraft speed/heading
Path control (rudder/tiller)
Speed Control
Change speed
Change direction
Ground Speed
Thrust Setting
Drive by Wire Control Systems



UCAs

UCA1: AE does not provide change in direction when a static obstacle is in the aircraft's path [H3]

UCA2: AE Provides change in direction when a static obstacle is not in the aircraft's path [H3]

UCA3: AE provides change in direction too late to avoid obstacle [H3]

UCA4: AE Provides change in direction too soon? [H?]

UCA5: AE provides change in direction to slowly to be effective to fully avoid object. [H3]

UCA6: AE provides change in direction too fast resulting in discomfort for crew/passengers [H4, H8, H6]

R1_UCA1: AE system shall be able to detect, classify, and track object of ? size at ? distance.

R2_UCA5/6: Rate of directional control input must be no less than X and no greater than Y degs/min.

Causal Scenarios:

CS_UCA1: AE believes that change in direction is not necessary due to false negative detection of object due to failure of sensor → SFHA

CS_UCA2: AE Believes the object is no longer there because it is perceiving the object's shadow and cloud cover eliminates the shadow (nonfailure)

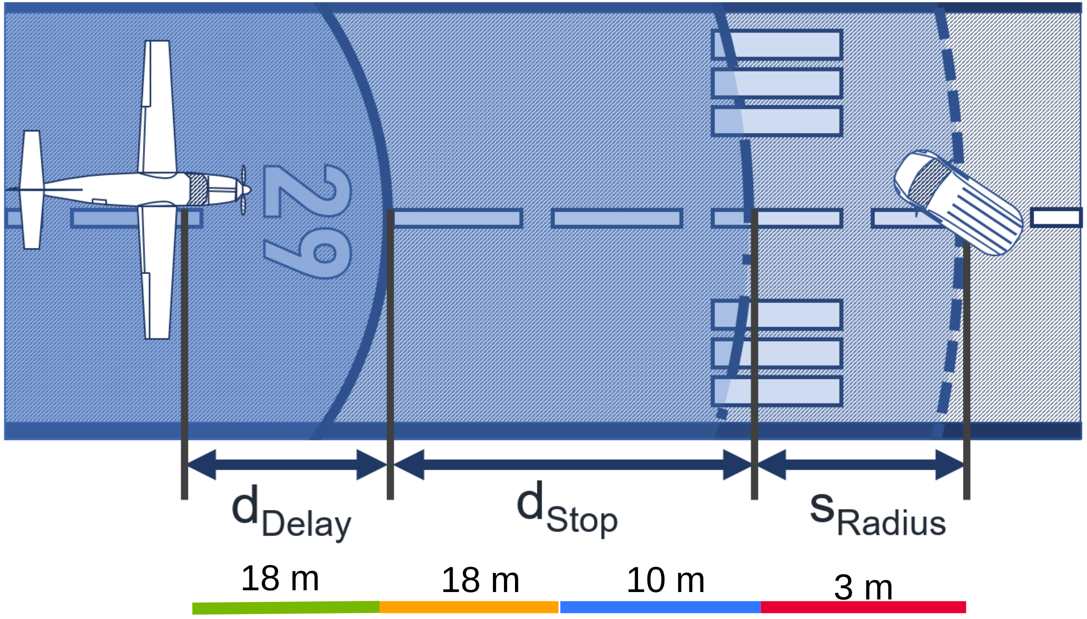
CS_UCA3: AE's detection and classification of object takes too long and the path planner controller has insufficient time to calculate a suitable avoidance solution. (object disappears and reappears)

R1_CS_UCA1: AE must have a situational awareness model readily sharable (HF-01)

R2_CS_UCA2: Flight crew must be provided sufficient visual feedback to represent AE process model to know that the AE has not detected an object. (HF-02)

Example Scenario

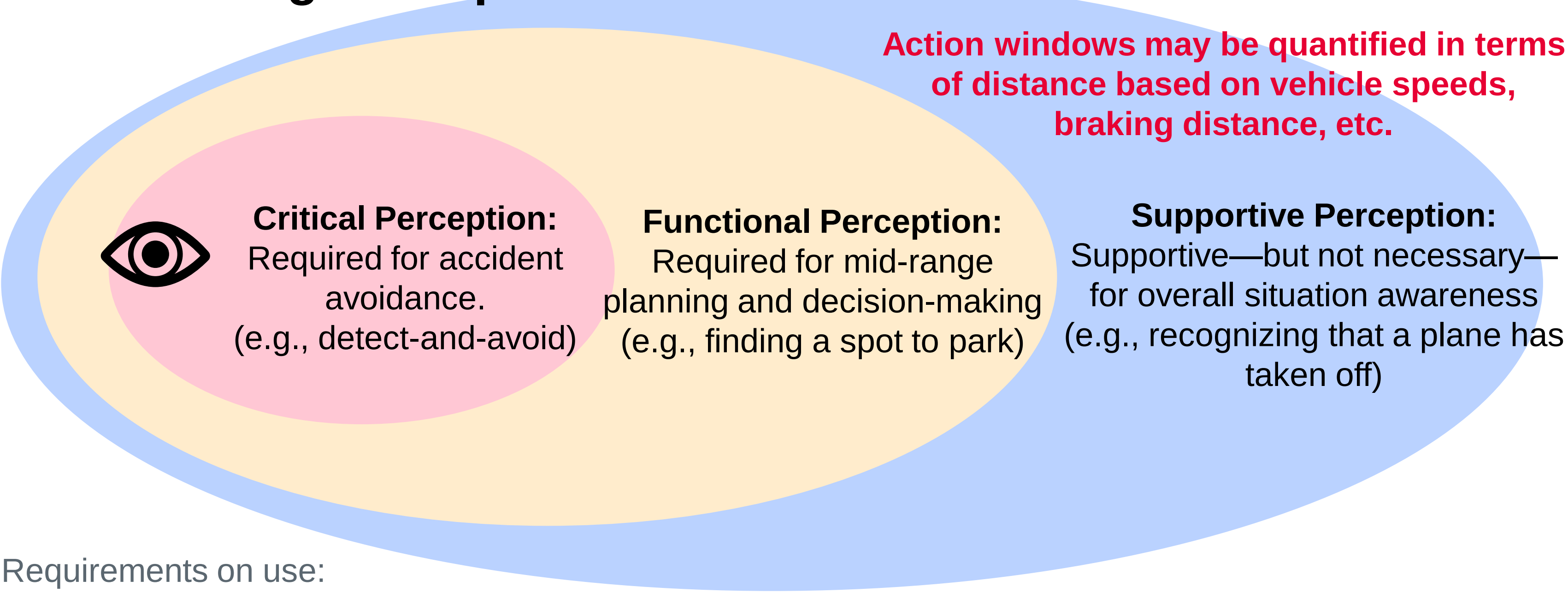
- Aircraft is taxiing toward a stationary vehicle
  - Aircraft max speed on taxiway is 20 mph (8.9 meters per sec)



ECCN: 7E994

Case Study: Auto-Taxi Perception System

Understanding Perception Failures

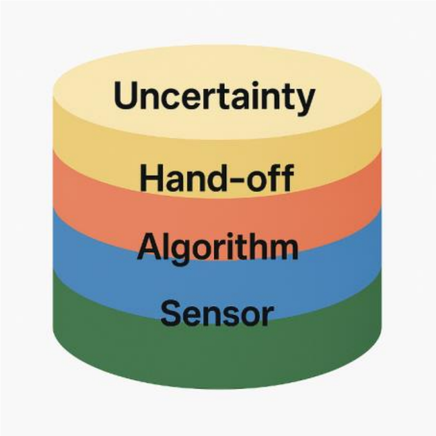


- Requirements on use:
  - Reliability of perception for each type of object in each action window
  - Is the set of objects considered sufficient: Full set of objects which must be considered in each category

ECCN: 7E994

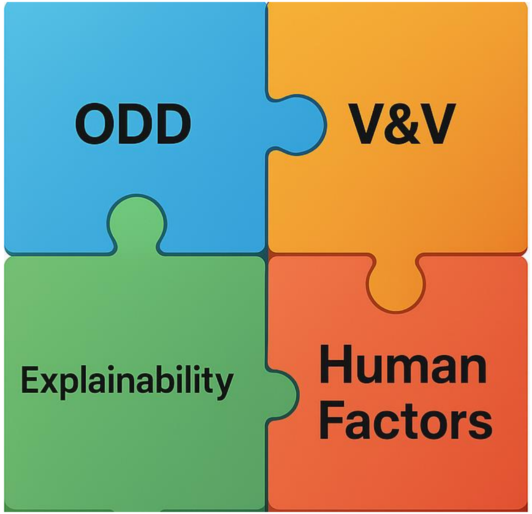
Safety, Uncertainty, and Risk in AI-Enabled Aviation Systems

- **Safety is an emergent property:**
  - Tightly coupled to system design and operational environment
- **Uncertainty arises from:**
  - Sensor limitations
  - Algorithmic performance
  - Handoffs (tracking/decision-making)
- **Key questions:**
  - What does it mean to hit an object and not detect it?
  - What object properties or scenarios lead to undetected hazards?
  - How do we quantify the probability of safe operation given sensor and system properties?
- **Uncertainty is represented by conditional probabilities**
- **Decomposition and analysis:**
  - Drill down into system components and hazard trees
  - Decompose state space to identify and mitigate hazards



Towards Trustworthy AI: V&V, Explainability, and Collaboration

- **V&V for AI/ML requires new, risk-based methods**
- **Operational Design Domain (ODD) and performance metrics are foundational**
- **Human factors and explainability are essential for trust and safety**
  - Clear, actionable explanations for pilots, regulators, and investigators
- **How do we know we've tested enough?**
  - Use statistical analysis, simulation, and empirical testing
- **Collaboration is key:**
  - Industry, regulators, and researchers must work together
- **Next steps:**
  - Evolving standards, continuous improvement, ongoing research



Increasing Safety at Landing with Perception-based Assistants

Technical Overview and Considerations for Certification

Boeing Beacon TEAM

EASA AI Days
Köln, August 28, 2025

ECCN: 7E994

Presenter

Dragos Margineantu, Ph.D. Boeing AI Chief Technologist Senior Technical Fellow 	**Presenter biography:** Dragos Margineantu is a Boeing Senior Technical Fellow and Artificial Intelligence (AI) Chief Technologist who is the technical lead of AI research and engineering in Boeing. His interests include computational methods for robust systems, autonomous commercial flight, anomaly and surprise detection & handling, reasoning under uncertainty, validation and testing of decision systems, cost-sensitive, active, ensemble learning, and inverse reinforcement learning. Dragos was one of the pioneers in research on ensemble learning and cost-sensitive learning and on statistical testing of learned models. At Boeing, he developed machine learning based solutions for autonomous flight, manufacturing, airplane maintenance, airplane performance, surveillance, and security.

Overview

- Assisting Pilots in case of Runway Incursions
- High level approach and goals
- Data, Knowledge, Unknowns
- Details of system under test
- Multifaceted Understanding / Fusion
- Analysis of our system in real-world conditions

Safe Runway

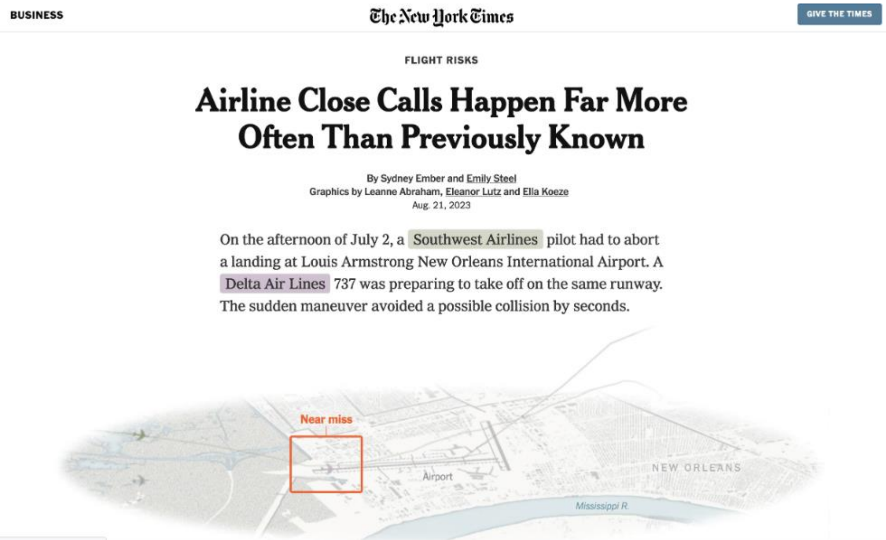
Safe Runway - Beacon TEAM

August 28, 2025

ECCN: 7E994

Runway Incursions

- Data suggests that runway incursions have been happening at an increasing rate
- Over the last 22 years:
  - Average ~700 out of ~16,405,000 flights per year result in a runway incursion requiring pilot action to avoid a potential collision
  - Includes incursions of Class C severity or higher
  - **1 dangerous incursion per 23,204 flights**
    - Approximately $4x10^{-3}$ probability
  - Source: [FAA Runway Incursion Database]


BUSINESS

The New York Times

GIVE THE TIMES

FLIGHT RISKS

Airline Close Calls Happen Far More Often Than Previously Known

By Sydney Ember and Emily Steel
Graphics by Leanne Abraham, Eleanor Lutz and Ella Koeze
Aug. 21, 2023

On the afternoon of July 2, a Southwest Airlines pilot had to abort a landing at Louis Armstrong New Orleans International Airport. A Delta Air Lines 737 was preparing to take off on the same runway. The sudden maneuver avoided a possible collision by seconds.

Source: New York Times / August 21, 2023

How is it Done Today

- **Trained humans (pilots/air traffic controllers)**

- **Assessment made on visual cues, radio comms, and situational understanding**

- **Pilot and air traffic control training employs the "rules of the road" in understanding when an incursion is occurring**

- **Classes of expected objects**
  - Aircraft
  - Utility Vehicles
  - GA Aircraft
  - Humans
  - Animals
  - Other static objects

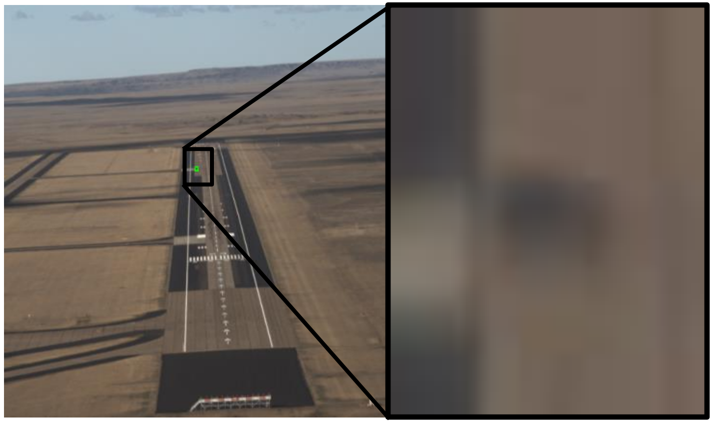
The Runway Incursion Problem

- The Computer Perception Problem

  - Assessment has to be done based at high speeds relative to more standard tasks in computer vision
  - Environment is dynamic
    - Moving platform
    - Moving intruders
    - Changing seasons and airport
  - Objects are detected at far distances (3 km away)
    - Lack of geometry, texture and features commonly used in computer vision
    - Grow quickly in size

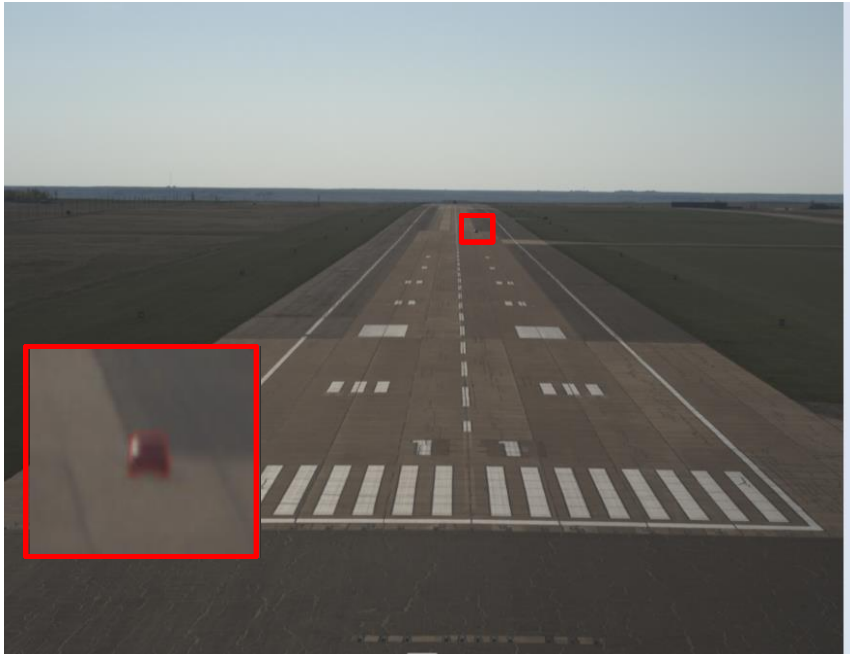
The Runway Incursion Problem

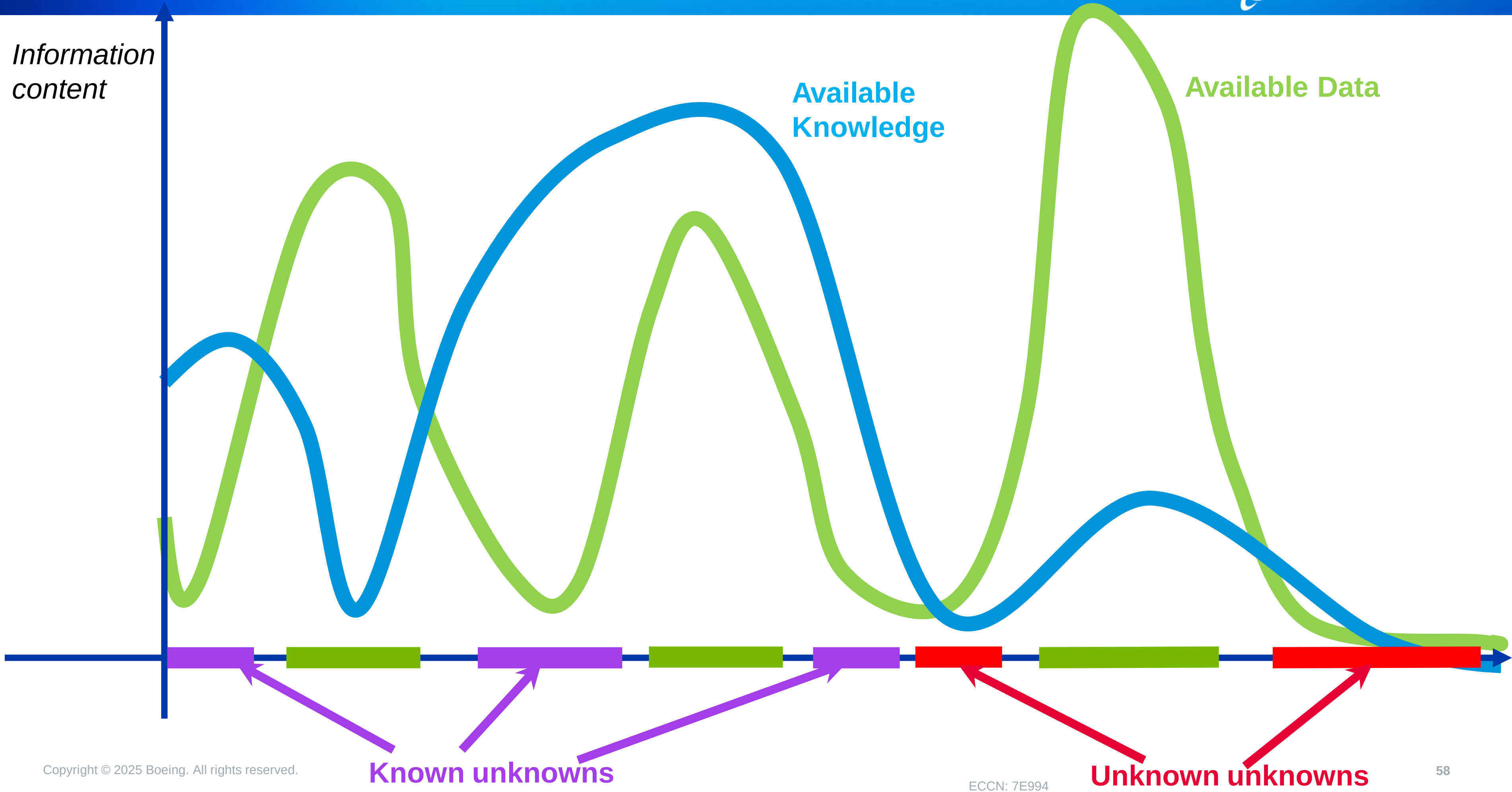
- The Visual Challenge
  - Target object size: ~2x2x2 meter or greater (rough size of an airport utility vehicle)
  - Detection Range - 3km to touchdown
  - On a 12.4 megapixel camera, with a 50mm focal length lens (high optical zoom, narrow field of view)
  - @ 3km
    - Object = ~10x20 pixels => **0.00145% of image**
  - @ 1km
    - Object = ~30x65 pixels => **0.013% of image**



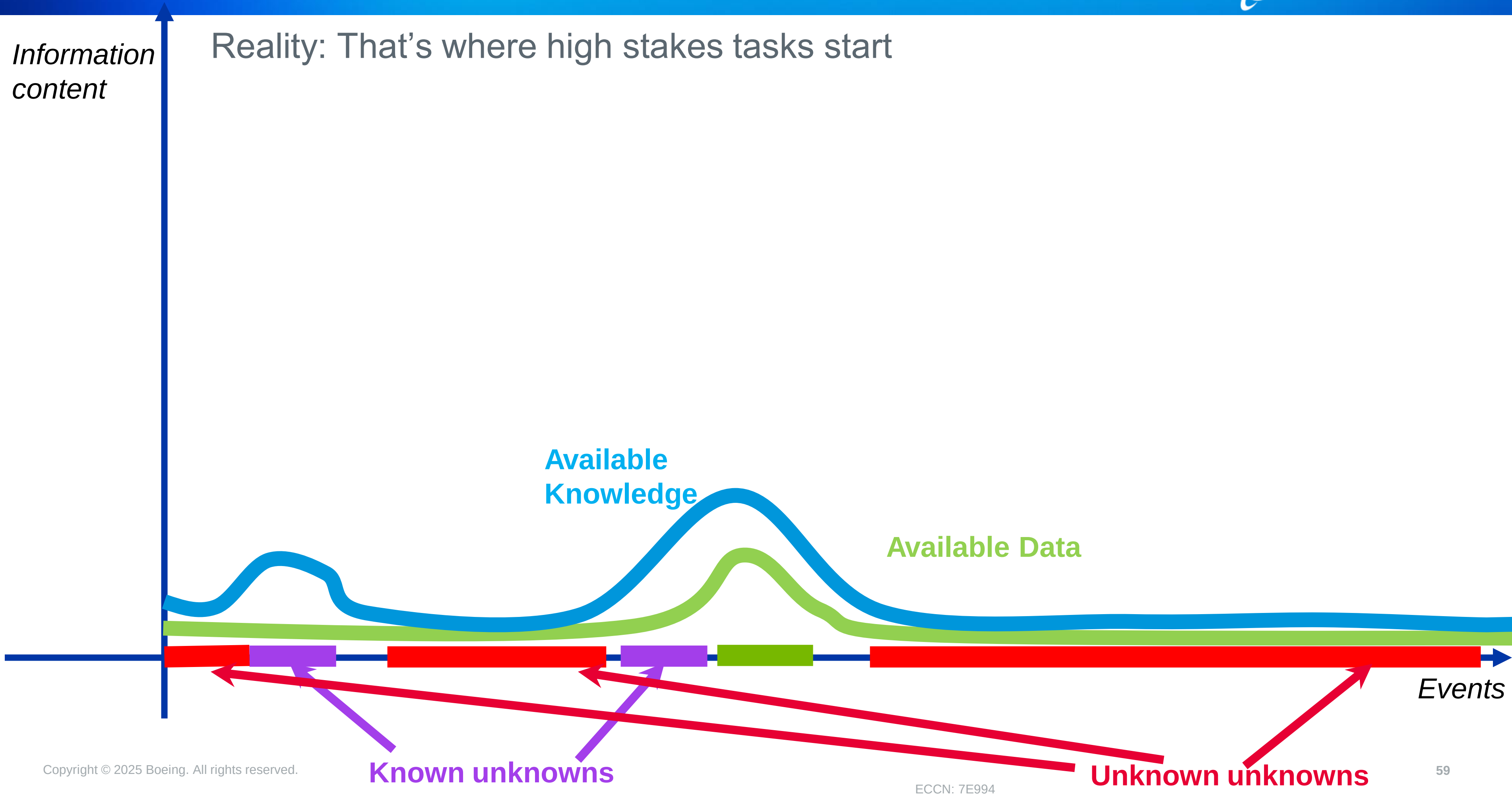
- **Tiny objects:**
  - **lack of geometry**
  - **lack of texture**
  - **lack of color saturation**

Goal




Information content
Available Knowledge
Available Data
Known unknowns
Unknown unknowns

Reality: That's where high stakes tasks start



The Thinking about Data & Knowledge

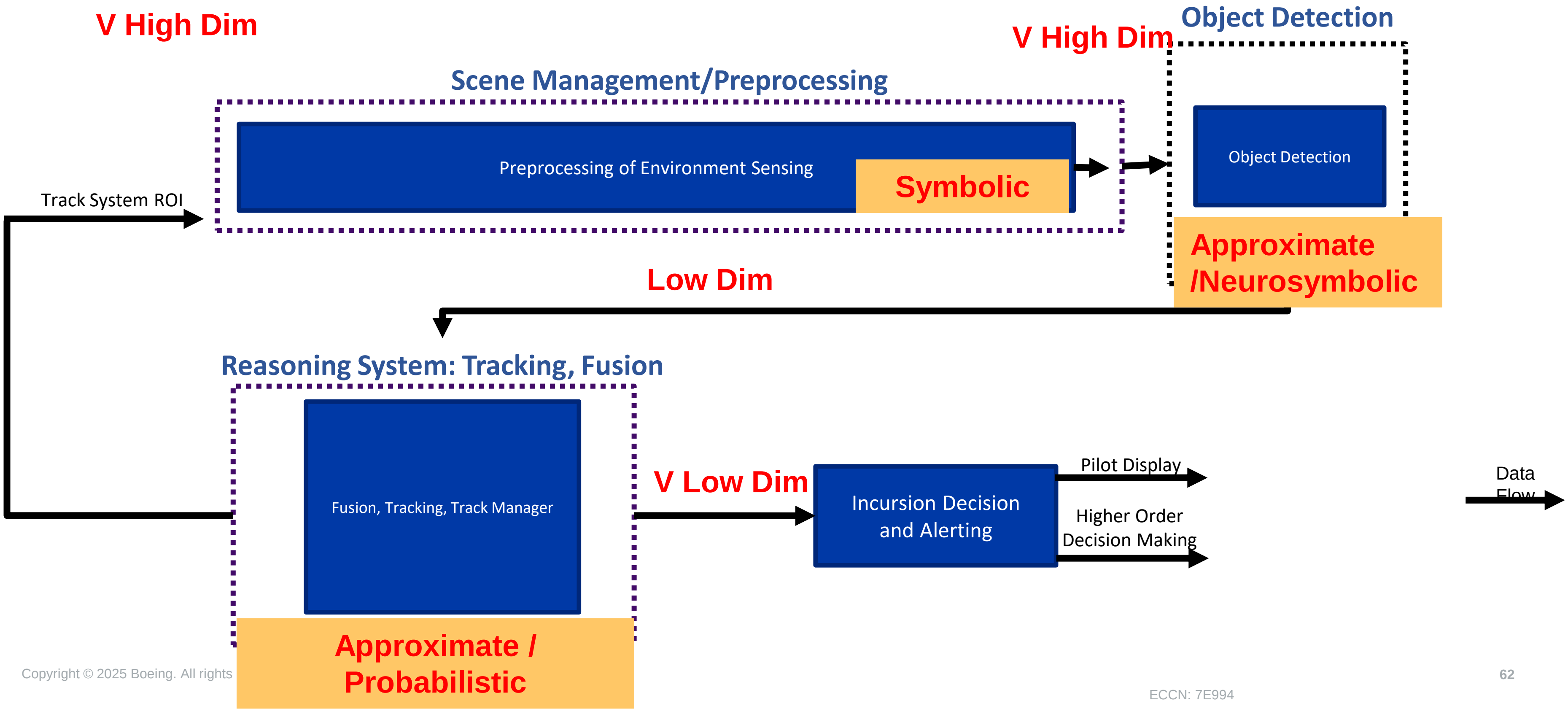
- Data ‘in the wild’ is required (as widely distributed as possible)
- For all high-risk applications, such data is limited and costly
- Strategy for real-world data is required

- *Data needed first for ‘understanding’, validation; then we’ll think about training.*

The Engineer's Challenges

- Identify the dimensions that are informationally relevant for task
  - Sensors
- & COMPUTATIONAL techniques to process and reason about them
- & means to collect, generate, manage DATA
  - For understanding
  - For validation
  - For training
- & means to handle regions of low density data and no data
  - System functionality
  - For analysis and evaluation

Safe Runway Assessment & Alerting System – Architecture



Understand in multiple ways: Redundancy & Increased Dimensionality

- M. Minsky:

“*You don’t really understand something if you only understand it one way*”

- Most current AI systems only perform task one way:
  - Computer vision:
    - Object Appearance ➔ alert
  - ADS-B:
    - Positive cue from Runway ➔ alert
  - Many standard solutions explored so far in most AI-based approaches

Understand one way

Understand another way

Understand yet another way

Key: REASONING /FUSION + Multifaceted Understanding

Overview of Goals

- Goal: safer landings
  - Assist pilots with safety critical task performed during period of high workload

- Desired Qualities
  - ~0 False Negative rate
  - Minimal false alarm rate.
    - Unnecessary go-arounds degrade the operational safety margin of airports
    - Pilots will quickly begin to ignore a system that reports inaccurate information
  - Enhanced robustness with multiple sensing modalities
  - Ownship confirmation of reported traffic conditions

Safe Runway Analysis - Conditions

- Approaches to runway with clearing traffic ahead of landing (a common scenario)
  - Traffic is already on runway when ownship begins approach
  - Traffic exits to taxiway enabling landing / low approach
- Approaches to runway with incursion during approach requiring a go-around
  - Runway is clear when ownship begins approach
  - Traffic enters runway during approach and remains, forcing go-around
- Visually challenging scenarios to both pilots and the safe runway system
  - Traffic is camouflaged on no-contrast surface on runway
  - Runway is incurred, forcing go-around

Sensor Fusion: Robustness + expanding current bounds

- Automatic Dependent Surveillance – Broadcast (ADS-B)
  - Strengths:
    - Very high “detection” range
    - Rich source of situational awareness context
  - Weaknesses:
    - “Detections” often have high uncertainty (i.e. 30 m @ 95 CEP)
    - Highly susceptible to spoofing
    - Not all incursions are equipped / transmitting
- Visual Perception
  - Strengths:
    - Highly accurate cross-range detections
    - Can be trained to detect a wide variety of incursions, including things not covered by ADS-B such as wildlife.
  - Weaknesses:
    - Detection range and field of view are inversely proportional
    - High compute cost and weather degradation

Sensor Fusion Progression

	Visual Only	ADS-B Cueing	Hybrid Fusion
Flight Test Analysis	Yes	Yes	Yes
EASA Demonstration (June 2025)	Yes	No	Yes

- Visual Only
  - Single 4k camera equipped with low distortion 50mm fixed focal length lens
  - Remains under development, but software stack is mature
- ADS-B Cueing
  - Independent tracking of 2D and 3D objects. Feedback loop combines information to provide cues to the camera for slice prioritization
  - Transitional technology step
- Hybrid Fusion
  - Expands ADS-B cueing to provide visual confirmation of self-reported ADS-B detections.
  - Transitional technology step

Visualization

Legend

- ADS-B Track
- Visual Only Track
- Hybrid-fused Track
- Inside RSA
- Outside RSA

Analysis of Results

Overview of Analyzed Test Conditions

@ 50 m AGL

- 59% of the test conditions have incursions
- 41% have no incursions.

- 36% have ADS-B on the vehicles
- 64% do not.

Intruder Distance @ 50 m AGL	% of Test Conditions
No intruder	41%
0-2000	19%
2000-3500	17%
3500-4500	7%
4500+	16%

Overview of Analyzed Test Conditions

Season Test Condition was flown	% Test Conditions
Spring	40%
Summer	8%
Fall	52%

Decisions for Critical Test Conditions *

Intruder Distance @ 50 m AGL	Number of Errors	Est. Probability of Correct Decisions
No intruder	1	0.992
0 - 2000 m	0	0.999
2000 – 3500 m	1	0.982
3500 – 4500 m	1	0.96
+4500 m**	23	0.56

*Correct decision means an alert when there is an intruder, no alert when there is no intruder.
**Distances outside of current targeted operating conditions

Visual Only

False Negatives

(No Alert when intruder on RSA)

Intruder Distance @ 50 m AGL	1-FN Rate	# FN (Visual Only)
0-2000 m	0.999	0
2000-3500 m	0.982	1
3500-4500 m	0.954	1
+4500 m*	0.27	38

False Positives

(Alert when no intruder on RSA)

1-FP Rate	# FP Visual Only
0.993	1

*Distances outside of current targeted operating conditions

Decisions For Test Conditions with ADS-B Data

	System Design Type		
Intruder Distance @ 50 m AGL	Visual Only	ADSB Cueing	Hybrid Fusion
No intruder	0.999	0.999	0.999
0 - 2000 m	0.999	0.999	0.999
2000 – 3500 m	0.965 (1error)	0.965 (1 error)	0.999
3500 – 4500 m	0.999	0.999	0.999
+4500 m*	0.5	0.67	0.92

*Distances outside of current targeted operating conditions

Success at the limits

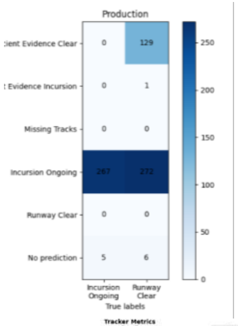
- The visual object detector can track large vehicles at large distances.
- Test conditions show successful and consistent tracking of a fire truck that starts ~13 km from the aircraft.

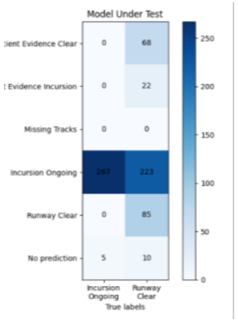

Actual state: Incursion Ongoing
Predicted state: Incursion Ongoing
AGL = 229
12759 m
7581 m
7522 m
5971 m

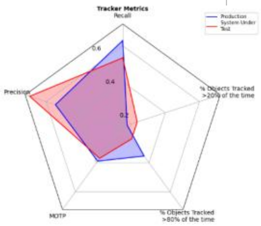

Actual state: Incursion Ongoing
Predicted state: Incursion Ongoing
AGL = 49
7021 m

System Performance and Evaluation

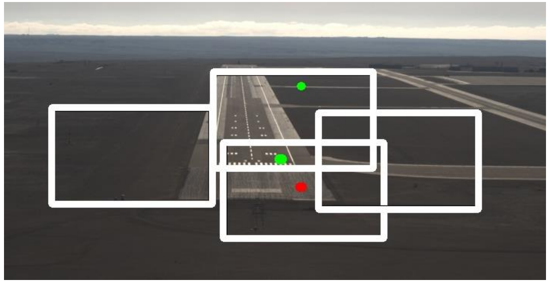
- Online System Evaluation
  - Compute 6 metrics per approach
  - Side-by-side baseline vs test system
  - Qualitative real-time visualization







- Batch System Evaluation
  - Playback sensor data to drive system
  - Compute 56 metrics over all approaches
  - Uncover performance edge-cases
  - Focus development and quantify improvements



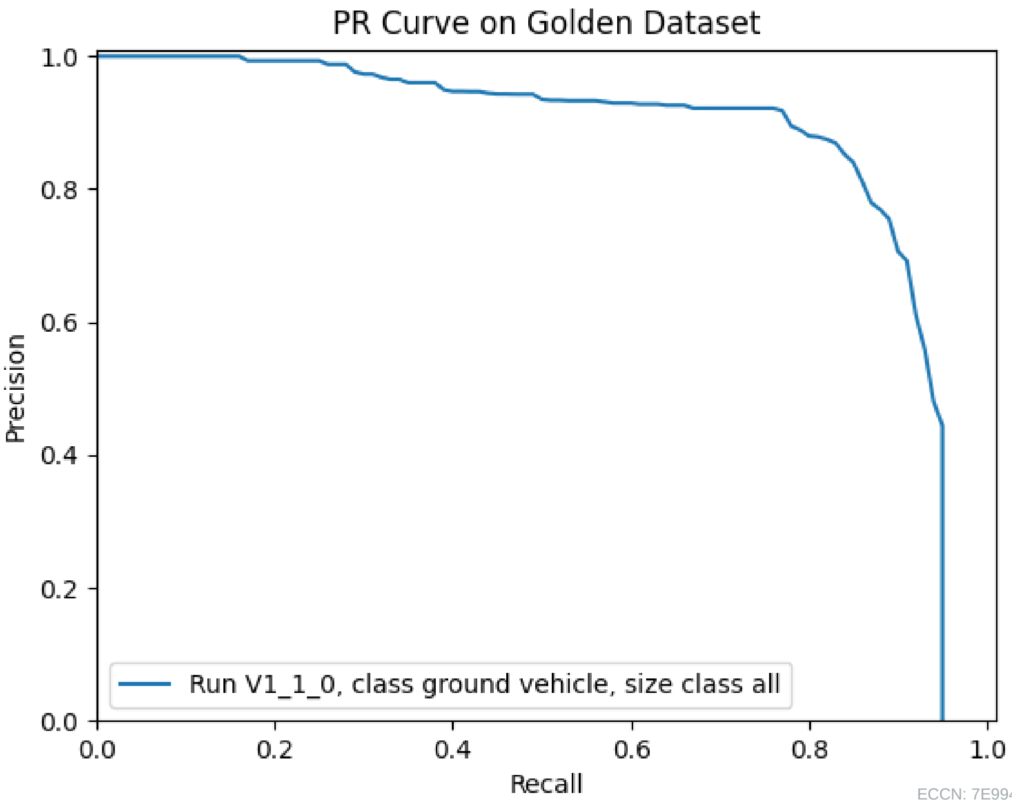
	All	Small	Medium	Large
Recall	0.53	0.38	0.73	0.77
Precision	0.71	0.68	0.67	0.82
Unique Objects	60	39	38	33
Mostly Tracked	19	9	15	17
Partially Tracked	20	16	12	4
Mostly Lost	21	14	11	12
False Positives	1615	852	480	283
Misses	3669	2924	356	389
Switches	37	28	5	4
Fragmentations	218	162	30	24
MOTA	0.31	0.20	0.37	0.59
MOTP	0.45	0.52	0.37	0.42

Specific Metrics of Interest

- Standard multiple object tracking metrics:
  - MOTA (Multiple Object Tracking Accuracy), MOTP (Multiple Object Tracking Performance), etc.
- Custom tracking metrics:
  - Track Quality, Track Continuity, Track Association
- In-depth alerting metrics tracking the alerting system's performance throughout an approach.

The CNN-based Object Detector in isolation

Visual Object Detector's PR Curve for vehicle on the runway lower than 75 m AGL



Detailed Analysis of Failures

Critical Failures – False Positives

- A false positive is when there is no intruder on the runway at 50 m AGL, but a warning occurs.
- Of the all the test conditions without incursion intruder is declared once.

Failure case: May 26, 2022

- False positive is on a distance marker sign
- The grass is particularly green during the time this image was acquired, which points to the need for more diverse detector training data.

Critical Failures – False Negatives

- A false negative is when there is a intruder on the runway at 50 m AGL and the intruder is estimated to be less than 3,500 m, but no incursion is declared.
- Of the all test conditions, of this nature, one (1) was a false negative.

Failure case: Nov.17, 2023

- Intruder distance at 50 m AGL 3340 meters
- Intruder declared at 40 m AGL intruder distance – 3079 meters
- Had ADS-B.
- Fails with visual only and ADS-B cueing, but alerts properly with **full fusion**.

Marginal Failures – False Negatives

- A marginal false negative is when there is an intruder on the runway at 50 m AGL and the intruder is estimated to be between 3,500 m – 4,500 m but no incursion is declared.
- Of the all test conditions, of this nature, one (1) was a false negative.

Failure case: Sept. 29, 2022

- Intruder distance at 50 m AGL 3560 meters
- Intruder declared at 44 m AGL intruder distance – 3378 meters.
- The intruder is found and then lost before 50 m.
- Early case: ADS-B data was not on the intruder for this test condition.

Summary & Discussion

- Carefully architected system with components developed using AI methods - employed in the design and implementation of systems that assist pilots in making/improving decisions
- Detecting and Alerting for Incursions at Landing
- System Architecture Overview
- Detecting, Tracking and Fusion
- Analysis on real-world cases

Human Factors considerations, and Systems-Theoretic Process Analysis

Presenter

Emily L. Howard, Ph.D. **Human Factors Engineering** **Principal Senior Technical Fellow** **Enterprise Functional Chief Office** 	**Dr. Emily Howard is a Principal Senior Technical Fellow in Human Factors Engineering, specializing in human information processing, perception and cognition. Her expertise is in systems-of-systems engineering, software architecture and user interface strategies, and workplace systems. She has worked on numerous Boeing products including fighters, bombers, tankers, training systems, handheld radios, and various workstations for autonomous systems, command and control, air and missile defense, homeland security, surface ships and space operations.** **At present, Emily is focused on adapting human factors engineering processes for enterprise deployment, through model-based engineering and technical excellence. She currently chairs the Technical Board for Human Factors Engineering. She is also the enterprise focal for applying Systems-Theoretic Analysis Model and Process (STAMP) for preventing losses. She previously led the Boeing Technical Fellowship as the Chair of the Leadership Team.** **Emily has 2 daughters and a passion for sailing: she has taught dinghy and keelboat classes and has raced for more than 35 years.**

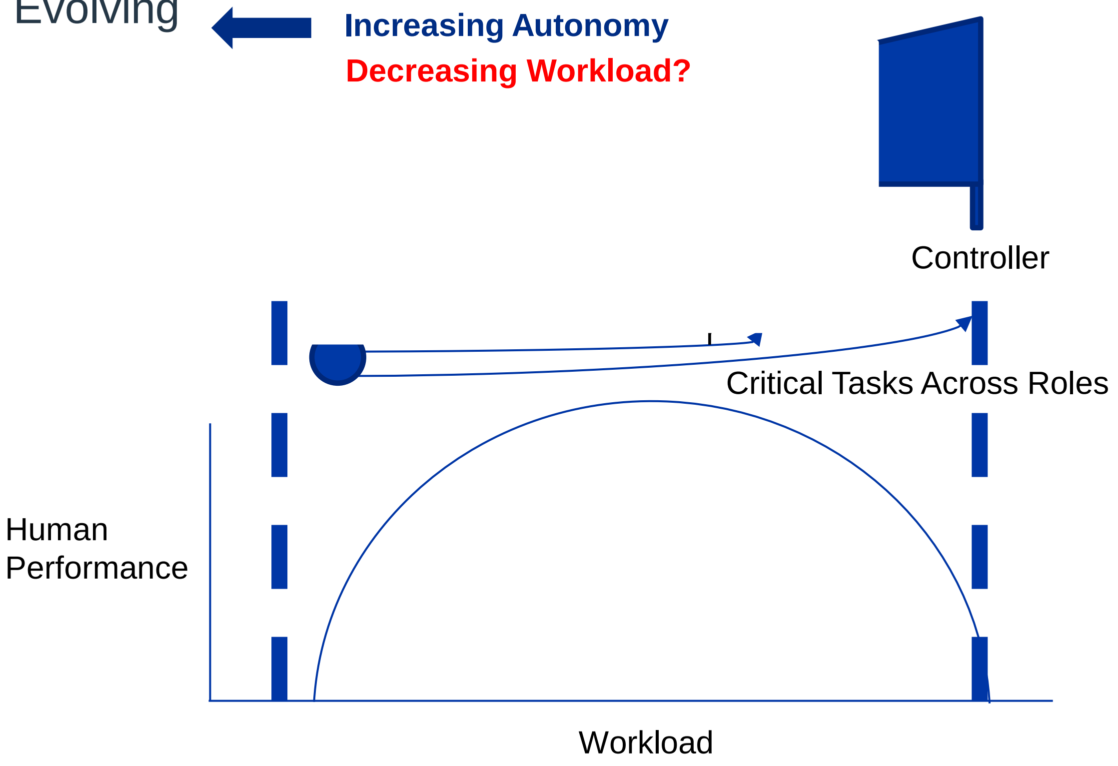
Key Take-aways

- The automated taxi system represents a new concept in for aircraft automation
  - With increasing automation, human roles in aviation are evolving
- Human-centric design must be applied throughout development
  - Important to understand the interaction requirements as part of the human-machine interface
- Systems-Theoretic Process Analysis (STPA) offers an efficient means of analyzing human interaction requirements, along with Safety and Verification analyses
  - Helps to address non-normal, yet possible, situations early in development
  - Can also form a partial basis to show regulatory compliance
- Human interfaces for testing can inform, but do not equate to, human interfaces for operations
- Level 3 AI systems must account for human contributions to system safety and performance

Human-Machine Teaming Roles are Evolving

Human Roles

- Controller: Responsible for performing most vehicle & all mission functions
- Mission Commander: Responsible for performing key mission and vehicle functions
- Supervisor: Responsible for monitoring vehicle & mission functions
- Contingency Management: Responsible only for intervening when autonomous functions are erroneous, inadequate, or unsuitable



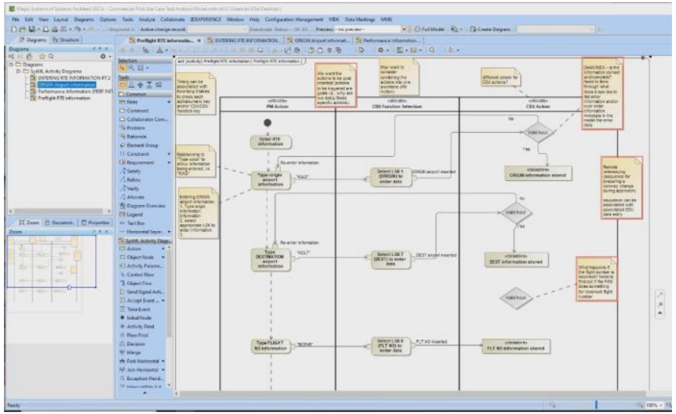
Key Constraint: Workload and human performance are **nonlinearly** related

- **Has implications for interaction design, with opportunities for performance monitoring**

Focus on Contingency Management (Non-normal) Tasks from the Start

Interaction vs Interface Design

- Interaction Design is the determination of:
- Who (considering multiple roles) needs or does
- What tasks (with associated information or action elements)
- When (required or desired) and
- How Well (in terms of precision, timeliness, etc)



Interface Design is informed by Interaction Design, subsequently determining:

- Where (in the space surrounding the human) and
- How display and control features are provided

Interactive Graphics

Pop-Up Windows

Drop-Down Menus

Toolbars

Human response assumptions begin with Interaction Design decisions

Early HFE Involvement is Key to Interaction Design

Flight Crew Oversight

- The crew will be responsible for:
  - Activation of the system
  - Monitoring the execution (through the provided interfaces)
  - If needed, entering the taxi destination and specific route requirements
  - Monitoring all aspects of the system
  - Overriding the system if abnormal operation or hazards are identified by the crew

- The automated taxi system will provide the flight crew the necessary information in order to monitor the system
  - This information display will be handled by the systems interface with the flight crew

Systems-Theoretic Process Analysis

Define the Losses

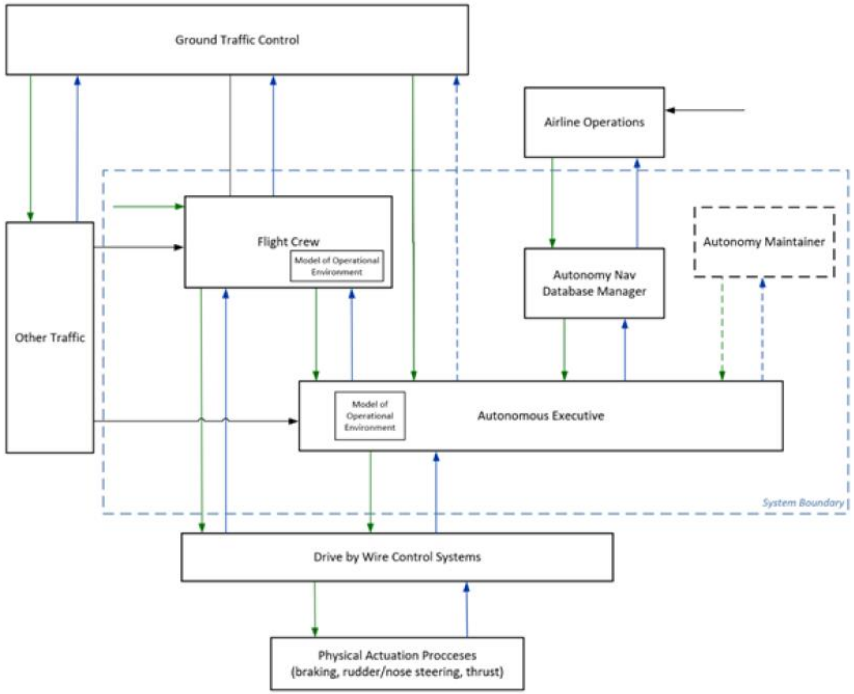
ID	Description	Affected Stakeholders
L1	Disruption to airport or airline operations	Airport authority
L2	Loss of human life, or injury to humans	All
L3	Loss of or damage to aircraft or airport infrastructure	Airport authority, airline, regulator
L4	Loss of passenger comfort	Regulators, airline
L5	Loss of reputation for airplane OEM, aircraft systems, systems supplier, airline, regulatory agency, or ATC agency	Regulators, airline, ATC
L6	Loss of trust in the system	OEM, Airline, Regulator
L7	Loss of revenue to regulatory fines/enforcement actions	Regulators, Airline

Define the Hazards

ID	Description	Linked Losses
H1	Airplane departs designated taxiway	L1-L7
H2	Airplane enters unauthorized/uncleared taxiway or runway	L1-L7
H3	Airplane exceeds proximity threshold to relevant objects (airport infrastructure, people, vehicles, other aircraft)	L1-L7
H4	Airplane exceeds expected maximum speed constraints	L1, L2, L3, L5, L6
H5	Airplane refuses to move or moves too slowly (e.g. due to false detection of object)	L1, L2, L3, L5, L6
H6	Airplane exhibits otherwise safe but unexpected/uncontrollable or inefficient behavior	L1, L4, L5, L6
H7	Autonomy system creates sequence of decisions which lead to only negative outcomes	L1-L7
H8	Autonomy creates excessive/additional workload for humans	L1, L5, L6

Model the Control Structure

High Level Control Structure for existing system



Systems-Theoretic Process Analysis

Identify Unsafe Control Actions (UCA)

Identify Loss Scenarios

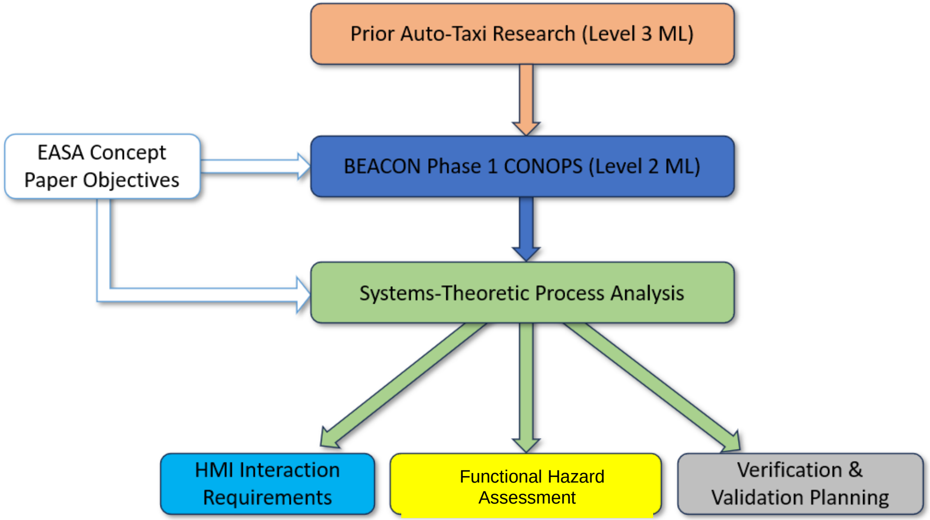
UCA ID Prefix	Controller	Destination Controller	Control Action	Provides	Does not Provide	Out of Sequence: Too Early, Too Late	Stopped too Early, Continued too Long
UCA_FC-AE_05	Flight Crew	Autonomous Executive	Modify Route	FC provides Modify Route command to provide an invalid route [H1]	FC does not provide Modify Route when needed [H1]	FC provides Modify Route too late [H6]	FC provides Modify Route command which does not follow established taxiways [H1]

For more information on STPA, check out the virtual STAMP workshop, Sept 22, 25, 29, & 30:
https://psas.scripts.mit.edu/home/stamp-workshop-information/

Human Factors Considerations

As a research project, BEACON did not apply our traditional Flight Deck design process.

Which enabled us to explore the use of STPA to generate HMI requirements.



Identify Unsafe Control Actions — Identify Loss Scenarios — Define requirements

Control Action	Unsafe Control Action	Causal Scenario	Requirements
Modify Route	Flight Crew provides Modify Route command to provide an invalid route [H1]	Flight Crew has outdated situational awareness of the state of airport NOTAMs.	Design of system must ensure current and complete situational awareness, including current state of airport NOTAMs.

STPA as a Means of Compliance

EASA asked Boeing to propose which objectives may be able to be met via the use of STPA

- Reviewed the objectives in the Concept Paper final issue 2
- The use of STPA can aid system design AND potentially be used to create compliance artifacts

Full Showing
CO-01
CO-02
CO-06
EXP-01
EXP-02
EXP-03
EXP-10
EXP-13
EXP-14
EXP-15

Partial Showing
CO-03
CO-04
ET-07
EXP-08
EXP-09
EXP-11
EXP-12
EXP-19

Partial Showing (cont.)
HF-02
HF-04
HF-07
HF-26
HF-27
HF-28
HF-29
HF-30
HF-31

Remote Operator Oversight of Level 3A AI: Multi-Vehicle Supervisor (MVS)

- The MVS is expected to passively monitor the behaviour and decisions made by the Automated Taxi System through a Human-Machine Interface (HMI) provided via the remote station.
- The MVS intervenes only upon alerting from an onboard monitoring system (Runtime Safety Monitor RSM) that the vehicle is not performing as intended, or any time the MVS observes the vehicle not operating as intended, is in a high risk situation, or otherwise exhibiting unsafe behavior.
- The MVS will have a direct data link with the vehicle, and a means to talk to Air Traffic Ground Control (GC) both via GC frequency on the vehicle or via standard terrestrial communication.
- The MVS will have the means to rapidly build situation awareness of the state of the vehicle and operational context.
  - If the MVS can confirm that the vehicle is functioning correctly, it may be possible to provide the vehicle with onward instructions to resolve the current operational issue and re-engage the automated taxi system.
  - If the vehicle is not performing as expected, the MVS will arrange for safe recovery of the vehicle.

Systems-Theoretic Process Analysis for Level 3 AI Automated Taxi System

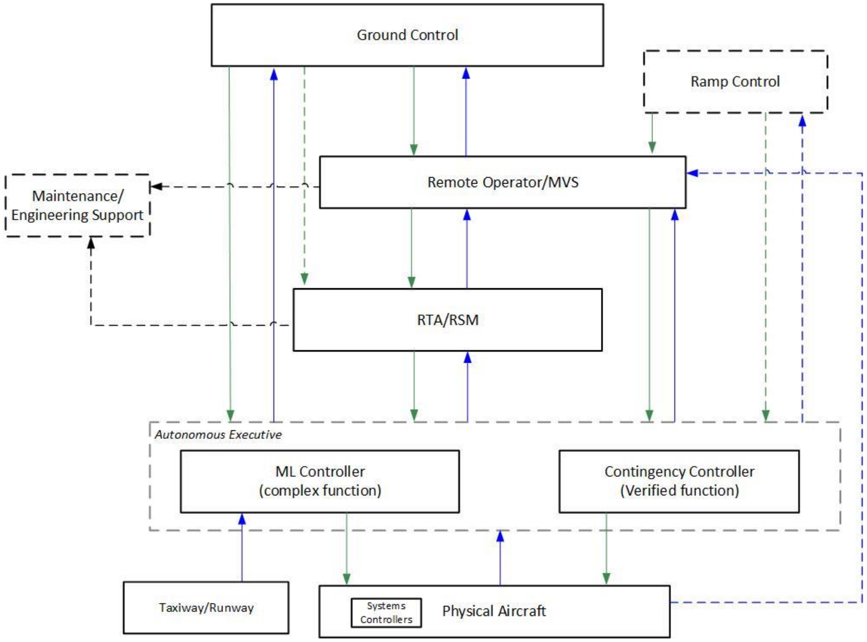
Losses

ID	Description	Affected Stakeholders
L1	Disruption to airport or airline operations	Airport authority
L2	Loss of human life, or injury to humans	Airport authority, airline, regulator, ATC, OEM
L3	Loss of or damage to aircraft or airport infrastructure	Airport authority, airline, regulator
L4	Loss of passenger comfort	Regulators, airline
L5	Loss of reputation for airplane OEM, aircraft systems, systems supplier, airline, regulatory agency, or ATC agency	Regulators, airline, ATC
L6	Loss of trust in automation	OEM, Airline, Regulator, ATC
L7	Loss of revenue to regulatory fines/enforcement actions	Regulators, Airline

Hazards

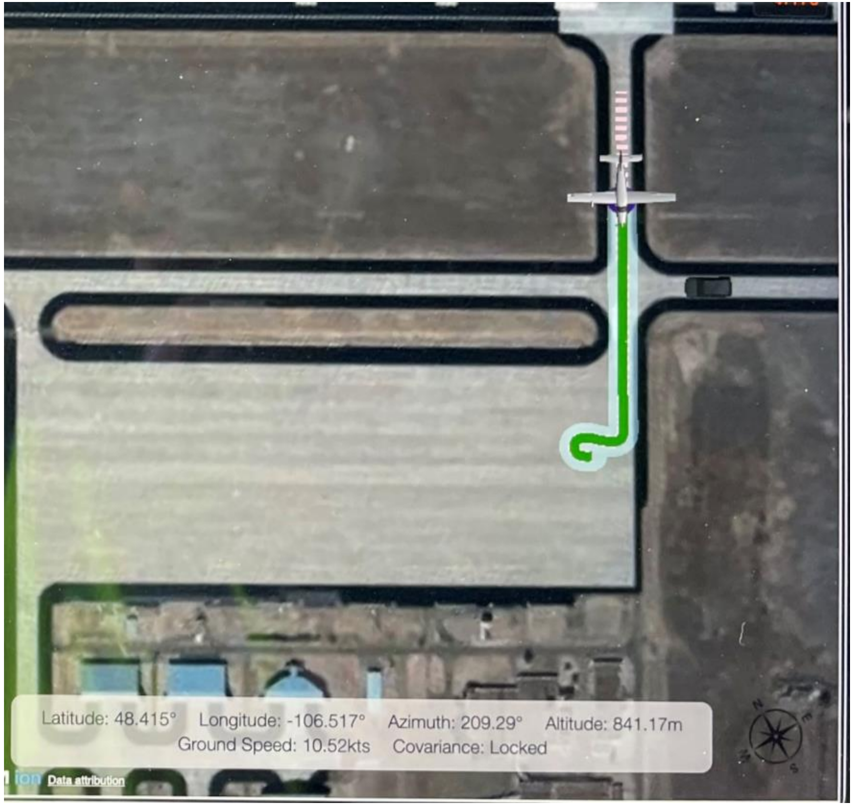
ID	Description	Linked Losses
H1	Aircraft departs designated taxiway (allowable taxiway space)	L1-L7
H2	Aircraft enters unauthorized/uncleared taxiway or runway	L1-L7
H3	Aircraft exceeds proximity threshold to relevant objects (airport infrasructure, people, vehicles, other aircraft)	L1-L7
H4	Aircraft exceeds expected maximum speed constraints	L1, L2, L3, L5, L6
H5	Aircraft refuses to move or moves too slowly (e.g. due to false detection of object)	L1, L2, L3, L5, L6
H6	Aircraft exhibits otherwise safe but unexpected or inefficient behavior	L1, L4, L5, L6
H7	Autonomy creates excessive/additional workload for humans	L1, L5, L6

Control Structure



ECCN: 7E994

Sample HMI from auto-taxi test program



ECCN: 7E994

Proposed objective for Level 3 AI

Objective HF-XX: For AI-based systems that aim to reduce human interaction compared to current or comparable systems, the applicant should design the AI-based system with a detailed understanding and retention (at a minimum) or enhancement (preferable) of all of the human contributions to operational safety and performance.

- **MOC HF-XX:** A detailed task analysis report of all human activities that are being performed within current (comparable) systems, with traceability to the AI-based system design aspects that perform, or eliminate the need for, those activities.

Wrap-up and next steps

IPC Final Thoughts

- Boeing thanks EASA for partnering with us on this IPC.
- Thank you for the opportunity to share this project with you today
- Boeing wanted to build upon the efforts of past IPCs, and hopes the information shared helps move the entire industry forward.

IPC Closeout

- A public report documenting IPC will be released later this year.

Q&A

EASA

European Union Aviation Safety Agency

EASA AI Days 2025

Coffee break – 10:30 to 11:00

www.sli.do

aidays2025

Passcode: kd7z53

Your safety is our mission.

An Agency of the European Union

EASA
European Union Aviation Safety Agency

EASA AI Days 2025
Panel of discussion: Human oversight in Level 3 AI

Moderator: Ines Berlenga
Project Manager 'Ethics for AI'

Emily Howard
Boeing

Carmen Bruder
DLR

Lee Glazier
Rolls-Royce

Denys Bernard
Airbus

Frances Brazier
TU Delft

Andrew Kilner
EASA

Your safety is our mission.

An Agency of the European Union

Human Oversight of Level 3 AI

Emily Howard – Principal Senior Technical Fellow

August 28th, 2025

"Decades of research have shown that people often struggle to perform this [oversight] role adequately, due to both cognitive limitations (e.g., poor vigilance in monitoring, inappropriate levels of trust) and inadequate design of the technology (e.g., inadequate transparency, system designs that create low engagement levels or bias human decision making)."

National Academies of Sciences Engineering and Medicine 2021. Human-AI Teaming: State-of-the-Art and Research Needs. Washington, DC: National Academies Press.

Overview

- Boeing partners with the Human Factors & Ergonomics Society (HFES)
  - First ever, Platinum Corporate Sponsor, since 2021
  - Selected as a Science Policy Fellow with HFES in 2022
- **This partnership has led to two activities related to Human Oversight of AI**
  - Dissemination of HFES' "Guardrails for Human Use of AI" (video link)
  - Promotion of HFES' "Human Readiness Levels" ANSI Standard for human-centered design

ANSI/HFES 400-2021

TIER 1 ADOPTION NOTICE

ANSI/HFES 400-2021, "Human Readiness Level Scale in the System Development Process" was adopted on 1 May 2025 for use by the Department of Defense (DoD).

This standard defines nine levels of the Human Readiness Level (HRL) scale and provides guidance for their application in the context of systems engineering and human systems integration processes. The HRL scale complements and supplements the existing Technology Readiness Level (TRL) scale to evaluate, track, and communicate the readiness of a technology or system for safe and effective human use.

Implementation guidance not included in the standard will be included in the 2025 update to the DoD Human Systems Integration Guidebook.

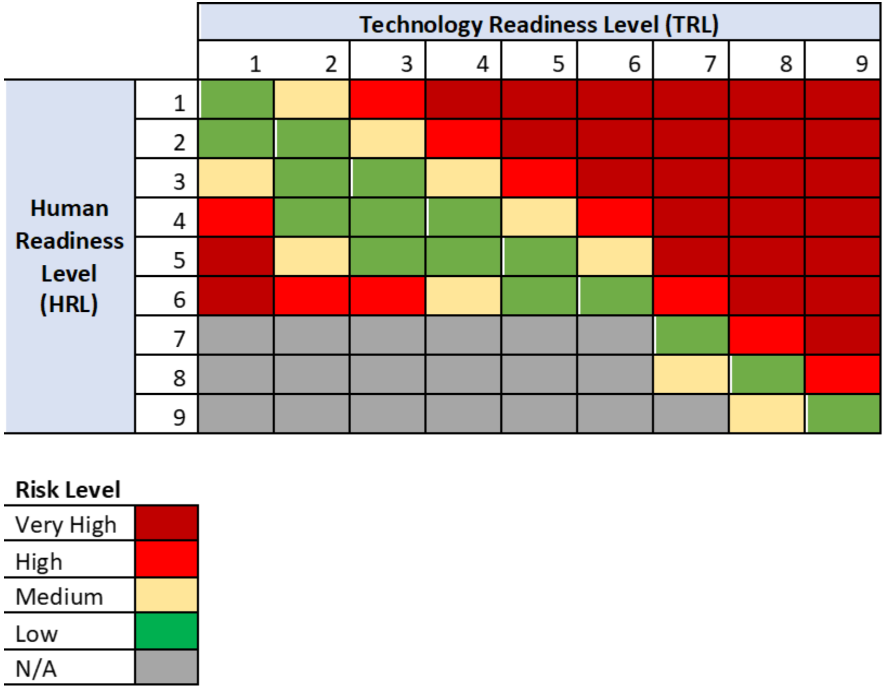
HFES

Systems That Work for **Humans**

ECCN: 7E994

HRL and TRL Scales Compared

Level	HRL	TRL
1	Basic principles for human characteristics, performance, and behavior observed and reported	Basic principles observed and reported
2	Human-centered concepts, applications, and guidelines defined	Technology concept and/or application formulated
3	Human-centered requirements to support human performance and human-technology interactions established	Analytical and experimental critical function and/or characteristic proof of concept
4	Modeling, part-task testing, and trade studies of human systems design concepts and applications completed	Component and/or breadboard validation in laboratory environment
5	Human-centered evaluation of prototypes in mission-relevant part-task simulations completed to inform design	Component and/or breadboard validation in relevant environment
6	Human systems design fully matured and demonstrated in a relevant high-fidelity, simulated environment or actual environment	System/subsystem model or prototype demonstration in a relevant environment
7	Human systems design fully tested and verified in operational environment with system hardware and software and representative users	System prototype demonstration in an operational environment
8	Human systems design fully tested, verified, and approved in mission operations, using completed system hardware and software and representative users	Actual system completed and qualified through test and demonstration
9	System successfully used in operations across the operational envelope with systematic monitoring of human-system performance	Actual system proven through successful mission operations



While the TRL scale focuses on technical maturity, the HRL scale emphasizes the readiness of a developing technology for safe and effective human use. The pace of development must align on both scales

EASA AI Days 2025

Panel of discussion: Human oversight in Level 3 AI

Moderator: Ines Berlenga
Project Manager 'Ethics for AI'

Emily Howard
Boeing

Carmen Bruder
DLR

Lee Glazier
Rolls-Royce

Denys Bernard
Airbus

Frances Brazier
TU Delft

Andrew Kilner
EASA

HUMAN OVERSIGHT IN EASA LEVEL 3A – SPECIFYING THE HUMAN SIDE

Carmen Bruder

Human oversight - human overseer

Source: EU AI Act 2024/1689 (Article 14)

- human oversight

= ensure **meaningful** human control where & when needed to

= to **prevent risks** to health, safety and fundamental rights

- differentiation human centric & human oversight

- scope: ‚**high risk**' AI systems

➢ clarify the requirements for human overseers

  ➢ roles & responsibilities

  ➢ requirements for meaningful oversight

human centric
proactive approach to meet human values and competencies

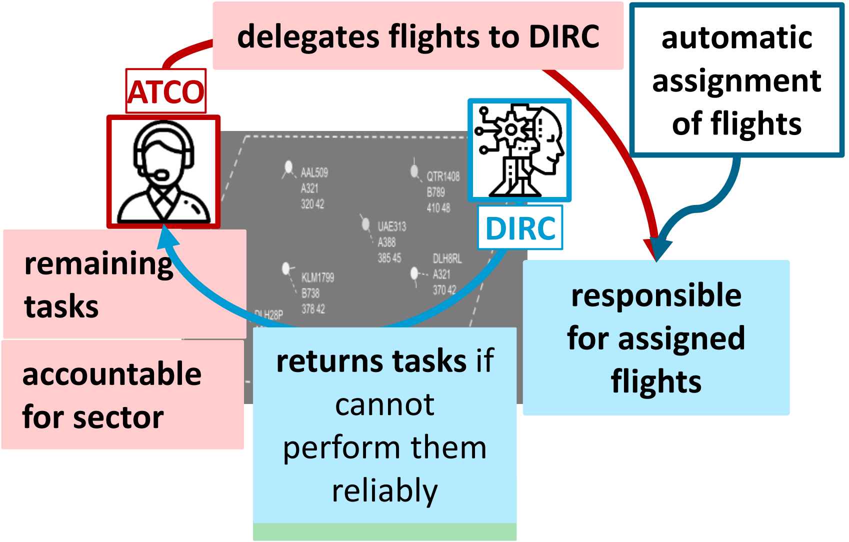
human oversight
reactive measure by assigning the monitoring and preventing of risks of AI to human

high risks systems in aviation (examples):
critical infrastructure management and operation, personnel management or education & training

DLR project LOKI – two high risk ATC use cases

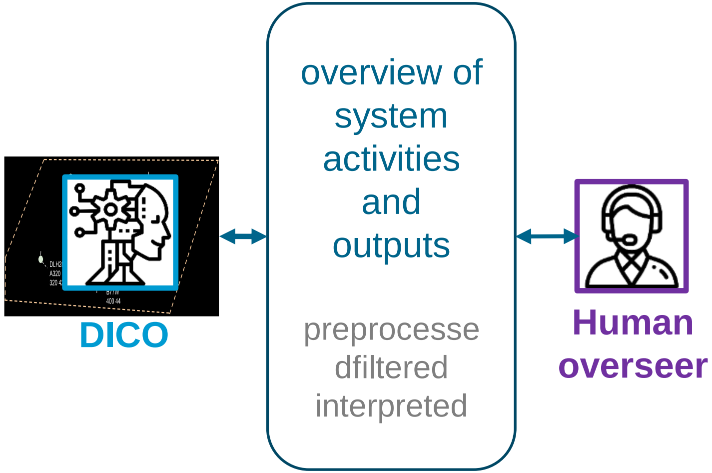
LOKI 2B ATC use case

- collaboration between human ATCO & digital controller DIRC



LOKI 3A ATC use case

- Human overseeing of digital controller DICO



Clarifying the role and tasks of human overseer of AI systems

Sources: Enquist (2023), Sterz et al. (2024)

DLR

human overseer OF high risks systems

Form of oversight

HITL (human-in-the-loop) = capability for active human intervention during overall decision cycle

1-2

HIC (human-in-command) = capability to oversee the overall activity of the AI system

3A

Phase of oversight

operation: decide and intervene directly

operation: oversee overall activity

control (recorded) activities

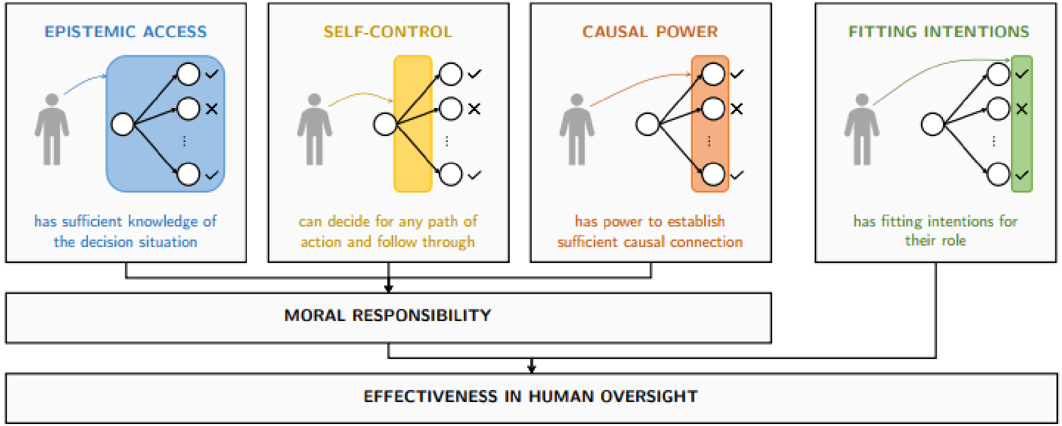
- requirements for all forms and phases: fully understand the capacities and limitations of AI system and be aware of possible biases (e.g. complacency, automation bias)

LOKI

Relevant research on human overseers requirements

Sterz et al. (2024), Bruder & Hasse (2020), Langer et al. (2025), Barzantny et al. (submitted), Rieger et al. (2025)

- humans‘ needs to effectively oversee high risk AI systems (see below)
- capacities, limitations and measurements of human monitoring behavior
- human biases & AI biases
- train and develop humans‘ expertise to oversee high risk AI systems
- how to integrate behavior of AI system in the human information process



The four conditions of effectiveness in human oversight and their relation to moral responsibility by Sterz et al. (2024)

Cited literature

- Barzantny, C., Bruder, C. & Hasse, C. (submitted). Navigating the Skies: A narrative review of human monitoring in avitation. Applied Ergonomics.
- Bruder, C. & Hasse, C. (2020). What the eyes reveal: Investigating the detection of automation failures. Applied Ergonomics, 82, article 102967. Elsevier. ISSN 0003-6870. (https://doi.org/10.1016/j.apergo.2019.102967
- European Data Protection Supervisor. (2025). *AI Act Regulation (EU) 2024/1689 : Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance)*. Publications Office of the European Union. https://data.europa.eu/doi/10.2804/4225375
- Enqvist, L. (2023). 'Human oversight' in the EU artificial intelligence act: what, when and by whom? *Law, Innovation and Technology*, *15*(2), 508–535. https://doi.org/10.1080/17579961.2023.2245683
- Langer, M., Lazar, V. & Baum, K. (2025). How to Test for Compliance with Human Oversight Requirements in AI Regulation? CHI'25 Conference on Human Factors in Computing System. https://doi.org/10.48550/arXiv.2504.03300
- Rieger, T., Onnasch, L, Roesler, E. & Manzey, D. (2025). Why Highly Reliable Decision Support Systems Often Lead to Suboptimal Performance and What We Can Do About it. IEEE Transactions on Human-Machine Systems, doi: 10.1109/THMS.2025.3584662
- Sterz, S., Baum, K., Biewer, S. Hermanns, H., Lauber Rönsberg, A., Meinel, P. & Langer, M. (2024). On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives. In ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT '24), June 3–6, 2024, Rio de Janeiro, Brazil. ACM, New York, NY, USA, https://doi.org/10.1145/3630106.3659051

Impressum

Title: Human oversight in EASA Level 3A – Specifying the Human Side
Date: 28. August 2025
Authors: Carmen Bruder
Institute: ME – Aviation and Space Psychology
Bildcredits: DLR

EASA
European Union Aviation Safety Agency

EASA AI Days 2025

Panel of discussion: Human oversight in Level 3 AI

Moderator: Ines Berlenga
Project Manager 'Ethics for AI'

Emily Howard
Boeing

Carmen Bruder
DLR

Lee Glazier
Rolls-Royce

Denys Bernard
Airbus

Frances Brazier
TU Delft

Andrew Kilner
EASA

Your safety is our mission.

An Agency of the European Union

HUMAN OVERSIGHT IN LEVEL 3 AI

Lee Glazier

HUMAN OVERSIGHT IN LEVEL 3 AI – ADVANCED AUTOMATION

Level 3A - The AI-based system performs decisions and actions that are **overridable by the human**

- Requires:-
  - The human must **know** that the AI system **needs** to be overridden
    - In RR, our Safety Critical AI QA framework has 5 mandated functions to detect 'malfunction' of the AI system
      - These include detection at output level, and at 'AI system level'
      - They have been derived from a certified safety critical software system
  - The human **must be able to override** the AI system
    - Override can range from correcting the AI output, to disabling the AI system
      - If the AI system is part of a critical service, then a safe/default system needs to be available to safely replace the AI system.

Level 3B – The AI-based system performs non-overridable decisions and actions

- Requires:-
  - Such a system needs to have monitoring to detect if it's malfunctioning and be able to override the system as necessary and appropriate
    - In RR, our fully autonomous AI QA framework has the same 5 functions mandated to detect 'malfunction' of the AI system, and pro-actively override as necessary and appropriate
      - If the AI system is part of a critical service, then a safe/default system needs to be available to safely replace the AI system.

Level 3 AI: advanced automation

- Level 3A: The AI-based system performs decisions and actions that are overridable by the human.
- Level 3B: The AI-based system performs non-overridable decisions and actions (e.g. to support safety upon loss of human oversight).

EASA
European Union Aviation Safety Agency

EASA AI Days 2025

Panel of discussion: Human oversight in Level 3 AI

Moderator: Ines Berlenga
Project Manager 'Ethics for AI'

Emily Howard
Boeing

Carmen Bruder
DLR

Lee Glazier
Rolls-Royce

Denys Bernard
Airbus

Frances Brazier
TU Delft

Andrew Kilner
EASA

Your safety is our mission.

An Agency of the European Union

Highlights

Is WG114/SG8 addressing level 3A systems?

Not yet SG8 is a subgroup of WG114 which focuses on the impact of AI on human factors. It is currently completing a diagnostic task, to determine if additional guideline is needed. The current analysis is focused on 1B/2A/2B use cases.

But the **method** can be (and will be) applied to 3A system: The starting point is a list of possible **HF concerns** falling in five categories: **human performances**, human system **interaction**, the **concept of work**, **operations**, **organization**. Then a set of well chosen use cases are then assessed, to determine whether HF issues would arise, and if existing methods and design interface/communication would be sufficient to avoid or mitigate those issues.

What could be the differences between level 2 and level 3A?

We don't expect level 3A system in the short term, and they will probably have **hybrid architecture** (AI + non AI, e.g. for alerting)

The main concerns of level 2 systems is human system interaction, to enable a fluent human system coordination, which will not be the case for level 3A. HF concerns for level 3A will be related to **human cognitive performances** (situation and system awareness, predictability) and work concept (accountabilities)

We must consider for level 3A systems the same questions that we address usually when **increasing the level of automation**: How will task sharing and responsabilities change? **Pilot cognitive capability to take over** to avoid hazard escalation? what would be the impact on pilot training?

From user perspective, the system the systems needs to remain sufficiently **interpretable** and **predictable**

The importance of explainability: Will the pilot accept to forget the system until an alert is raised?

System explainations will have to satisfy diverse **cognitive needs**:

- in case of alert: grasp the **attention** of the pilot, without causing overstress; update **situation awareness**: **understand** the alert, grasp the context, **decide**.

- in normal life: **justification** of what the system is doing (**trust**, **appropriate reliance**), support **prediction**

If explainability is delivered through a conversational interface, we inherit part of the concerns of level 2 systems regarding language

EASA AI Days 2025

Panel of discussion: Human oversight in Level 3 AI

Moderator: Ines Berlenga
Project Manager 'Ethics for AI'

Emily Howard
Boeing

Carmen Bruder
DLR

Lee Glazier
Rolls-Royce

Denys Bernard
Airbus

Frances Brazier
TU Delft

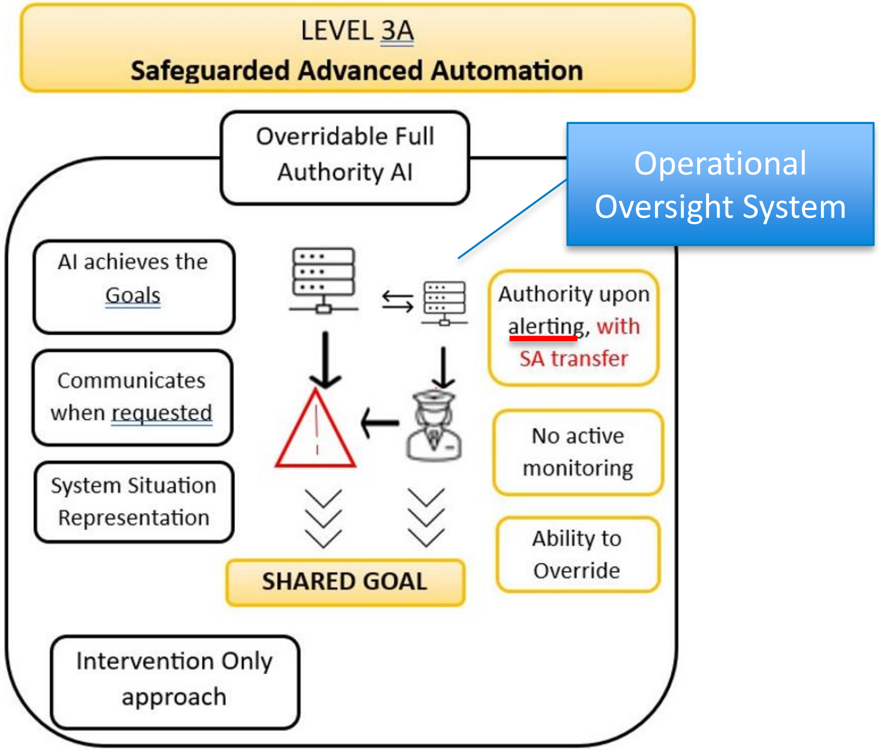
Andrew Kilner
EASA

EASA AI Days 2025
EASA Perspective on Level 3 AI

Andrew Kilner
EASA Human Factors Expert

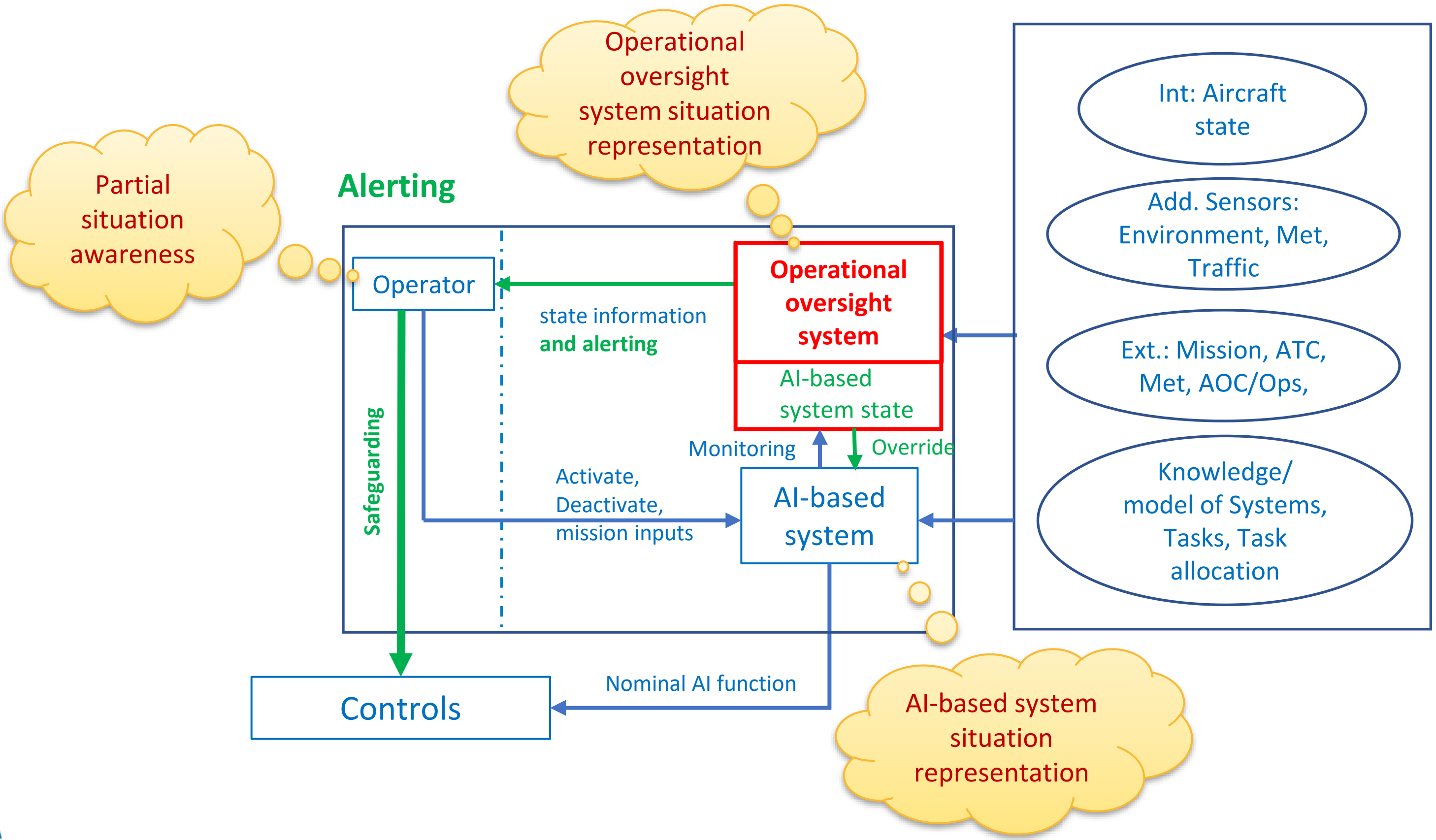
Level 3 AI in Concept Paper ‘Proposed Issue 03’

- **Level 3 AI** concerns “**advanced automation**”: the use of systems that, under specified conditions, **functions without human intervention**
- It is not to be mixed with “**autonomy**” which characterises a **system that is capable of modifying its intended domain of use or goal without external intervention, control or oversight,** and for high-risk systems is not compatible with EU AI Act Article 14!
- In collaboration with its Scientific Committee AI Task Force, EASA developed the following concept involving an independent **“Operational Oversight System”**, in line with EU AI Act Article 14 on “human oversight”.



SA = Situation Awareness

Level 3A – delegated oversight concept



EASA
European Union Aviation Safety Agency

EASA AI Days 2025
Innovation Partnership with Lufthansa Group

Investigation of an LLM-based troubleshooting assistant



Urszula Szamotulska
Lufthansa Group

Martin Ricklin
Lufthansa Group

Marcel Bauer
Lufthansa Group

Your safety is our mission.

An Agency of the European Union

LUFTHANSA GROUP

EASA AI DAYS 2025

TechOps AI Use Case

IPC Lufthansa Group

28.08.2025, Cologne

U. Szamotulska, M. Ricklin, M. Bauer

Public

About us

Urszula **Szamotulska**

Senior Director, Innovation & Regulation, Lufthansa Group

Martin **Ricklin**
Senior Manager, Digitalization & Application Strategy, Lufthansa Group

Marcel **Bauer**

Expert, Regulation & Authority Liaison, Lufthansa Group

Agenda

Airlines and Regulators: Shaping the Future of AI Together

AI Use Cases in Airline TechOps: Boosting Productivity and Reliability

From Concept to Compliance: Risk, Regulation, and the Future of Human-AI-Assistant


LUFTHANSA GROUP

Common AI-Applications for Airline Technical Operations – our goal is to drive innovation through practical use cases

AI-Assisted technical troubleshooting

Predictive and Condition-based maintenance

Intelligent optimization of work packages

Digital twins and simulation technologies

LUFTHANSA GROUP

Technical Fault

Identification

Get overview of the situation

Trouble-shooting

Collect information and consult various knowledge-databases

Rectification

Initiate corrective actions

Ready for Operation

AI-generated image

LUFTHANSA GROUP

Technical Fault

Identification

Trouble-shooting

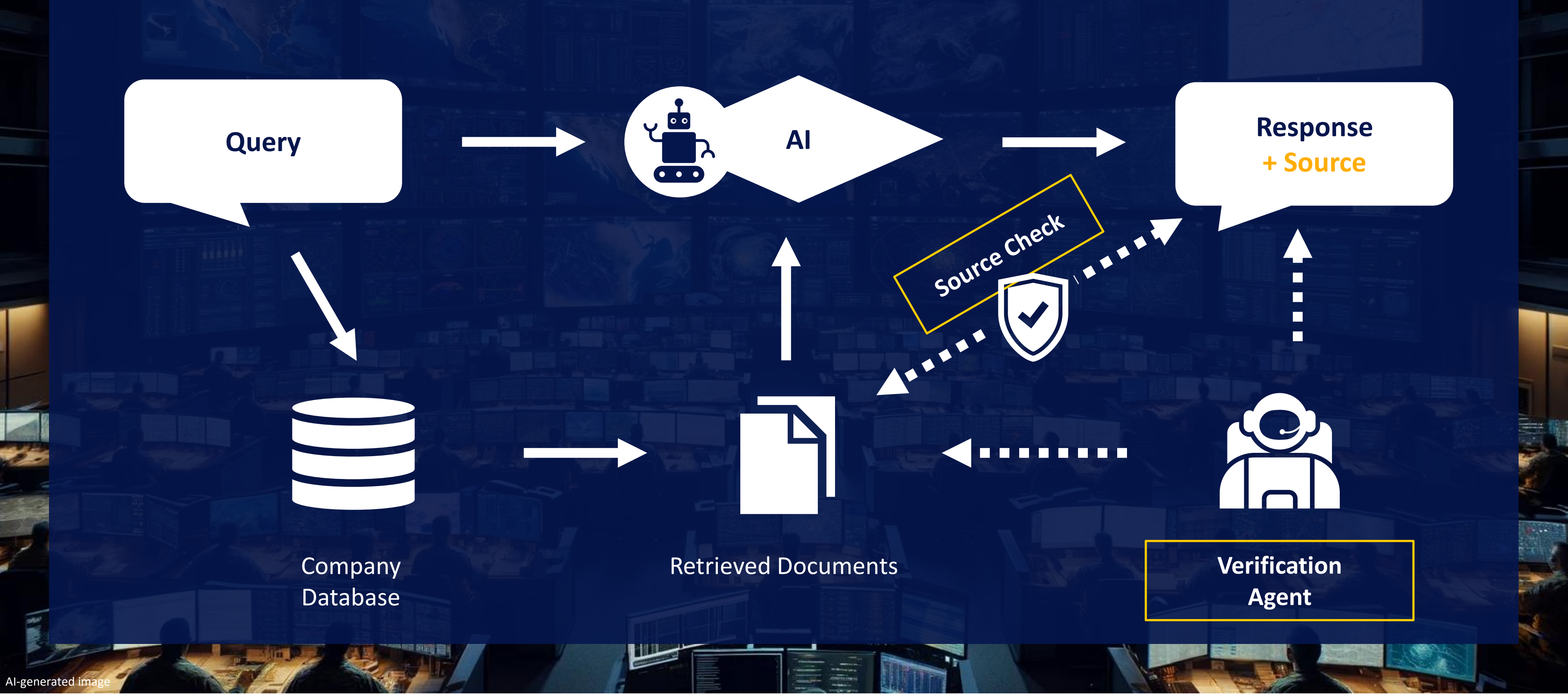
Rectification

GenAI Assistant

Ready for Operation

AI-generated image

Generative AI (GenAI) chatbot using Retrieval Augmented Generation (RAG)



LUFTHANSA GROUP

From Concept to Compliance: Risk, Regulation, and the Future of Human-AI-Assistant

Concept of Operation

Airline Technical Operations

- Line Maintenance (Part-145)
- Troubleshooting Support (Part-145 / Part-CAMO)
- Systems Engineering (Part-CAMO)

AI Classification

Human Assistance – support to decision making

- GenAI Chatbot acts as assistant, not decision-maker -> technician remains in control
- Provides smart search across maintenance history and manuals to surface relevant highlights

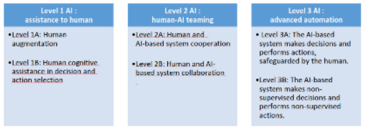


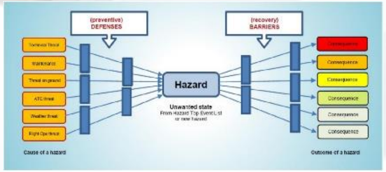
Figure 5 — Classification of AI applications

Ref. EASA Concept Paper Issue 02

Operational Risk Evaluation

Applied Bow-Tie Model for structured risk assessment

- Identified threats, barriers, and mitigations for safe AI use in TechOps
- Risk Assessment = cornerstone of change management and safe introduction of AI tools



LUFTHANSA GROUP

Exploring how GenAI fits within today's Regulatory Frameworks

Where we are today

- No AMC/GM yet for applied GenAI in Part-CAMO / Part-145 approved organizations
- IPC with EASA to co-develop safe pathways
- EASA Concept Paper Issue 02 as anchor → GenAI adaptation

Approach

- Mapping of the trustworthy AI building blocks
- Classification → Level 1B (human assistance): smart search + history data; technician in control

Challenges

- Assessment outcome: OD/ODD boundaries, operational explainability, adaptation of the EASA Concept Paper for GenAI applications
- Use identified challenges to develop new guidance

Future challenges

- Building AI competence across all aviation stakeholders
- Establishing practical validation frameworks for AI-based operational tools within approved organizations
- Clarifying responsibility and defining a common organizational approach

LUFTHANSA GROUP

LUFTHANSA GROUP

Thank you for your attention

EASA

European Union Aviation Safety Agency

EASA AI Days 2025

Lunch break – 13:00 to 14:00

www.sli.do

 # aidays2025

Passcode: kd7z53

Your safety is our mission.

An Agency of the European Union