

TRUSTY



SAPIENZA
UNIVERSITÀ DI ROMA

Enhancing Trust in AI for Remote Digital Towers: From Gap Analysis to Integration

Sara Kebir †, **Alexandre Veyrié †**, Guillaume Truong †, Elizabeth Humm ‡, Nicola Cavagnetto ‡, Mobyen Uddin Ahmed §, Shaibal Barua §, Mir Riyanul Islam §, Giulia Cartocci ¶, Pietro Arico' ¶, **Christophe Hurter †**

†Fédération ENAC ISAE-SUPAERO ONREA, Université de Toulouse, Toulouse, France

‡ Deep Blue s.r.l., Rome, Italy

§ School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden

¶Department of Molecular Medicine, Sapienza University of Rome, Rome, Italy



Co-funded by
the European Union

SESAR Innovation Days, 03-12-2025 Slovenia

A paradigm shift in ATM

Remote Digital Towers utilize high-definition cameras, infrared sensors, and secure networks to provide Air Traffic Services from a remote location

- **Cost efficiency:** Consolidates operations for multiple low-traffic airports
- **Enhanced vision:** Augmented reality overlays and night vision capabilities
- **AI integration:** The "Digital Tower Assistant" acts as a cognitive partner



TRUSTY

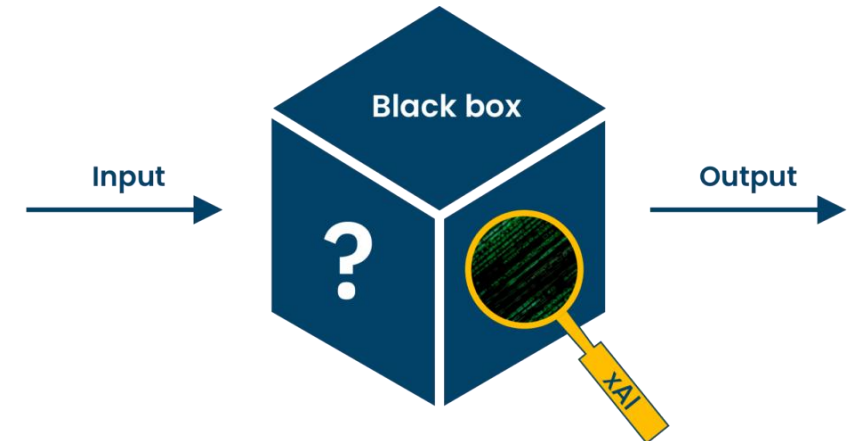
The Operational Challenge

Workload & complexity ¹

- Transitioning to RDTs introduces significant socio-technical hurdles
- Controllers often face heightened mental demands, especially when monitoring multiple airports simultaneously

The trust deficit ^{2,3}

- Unlike traditional automation, AI systems can be opaque ("Black Box")
- Without understanding the why behind an AI decision, controllers may reject valid recommendations, compromising safety



1. Gawade, M., & Zhang, Y. (2016). Synthesis of Remote Air Traffic Control System and Air Traffic Controllers' Perceptions. *Transportation Research Record*, 2600, 49 - 60.
2. Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3), 31-57.
3. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267, 1-38.

TRUSTY

Gap Analysis: Three Critical Areas



Trust & user-centricity

Lack of explainability (XAI) and insufficient human-centric design lead to operator skepticism and under-utilization of AI tools



Task allocation

Ambiguity in responsibilities between AI and humans, particularly during emergency scenarios where AI reliability drops



Situational awareness

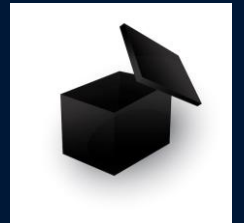
Poor HMI design and lack of data fusion hinder the controller's ability to maintain a comprehensive mental picture

TRUSTY

Gap 1: The "Black Box" Problem

Current RDT AI systems provide operational support but often lack sufficient transparency to explain decision rationale. In safety-critical environments like ATM, this opacity is unacceptable and leads to skepticism and under-utilization ¹

- **No context:** Controllers don't know if a recommendation considers all variables (e.g., weather, fuel, incursion)
- **Static trust:** Systems lack mechanisms to build trust gradually over time through feedback loops
- **Consequence:** Even accurate AI advice is ignored if the controller cannot verify it instantly



TRUSTY

Gap 2: Ambiguous Responsibility

Who is in control? ¹

The division of tasks/roles between human and machine remains unclear, specifically in non-nominal situations

- **Emergency limits:** AI models trained on nominal data may fail during rare emergencies, leaving the controller to take over abruptly
- **Mode confusion:** Controllers may over-rely on AI when they should be vigilant, or interfere when AI is performing optimally
- **Static allocation:** Workload isn't dynamically rebalanced based on traffic density or operator state





Fragmented Data

Operators must synthesize video feeds, radar, and voice comms manually. AI systems often fail to fuse these into a single, coherent operational picture



Information Overload

Poorly designed Human-Machine Interfaces (HMI) clutter the screen with non-essential data, distracting from critical alerts during peak traffic



Latency Risks

Real-time decision making requires zero latency. Heavy data processing without optimized HMIs can introduce dangerous delays in awareness

TRUSTY

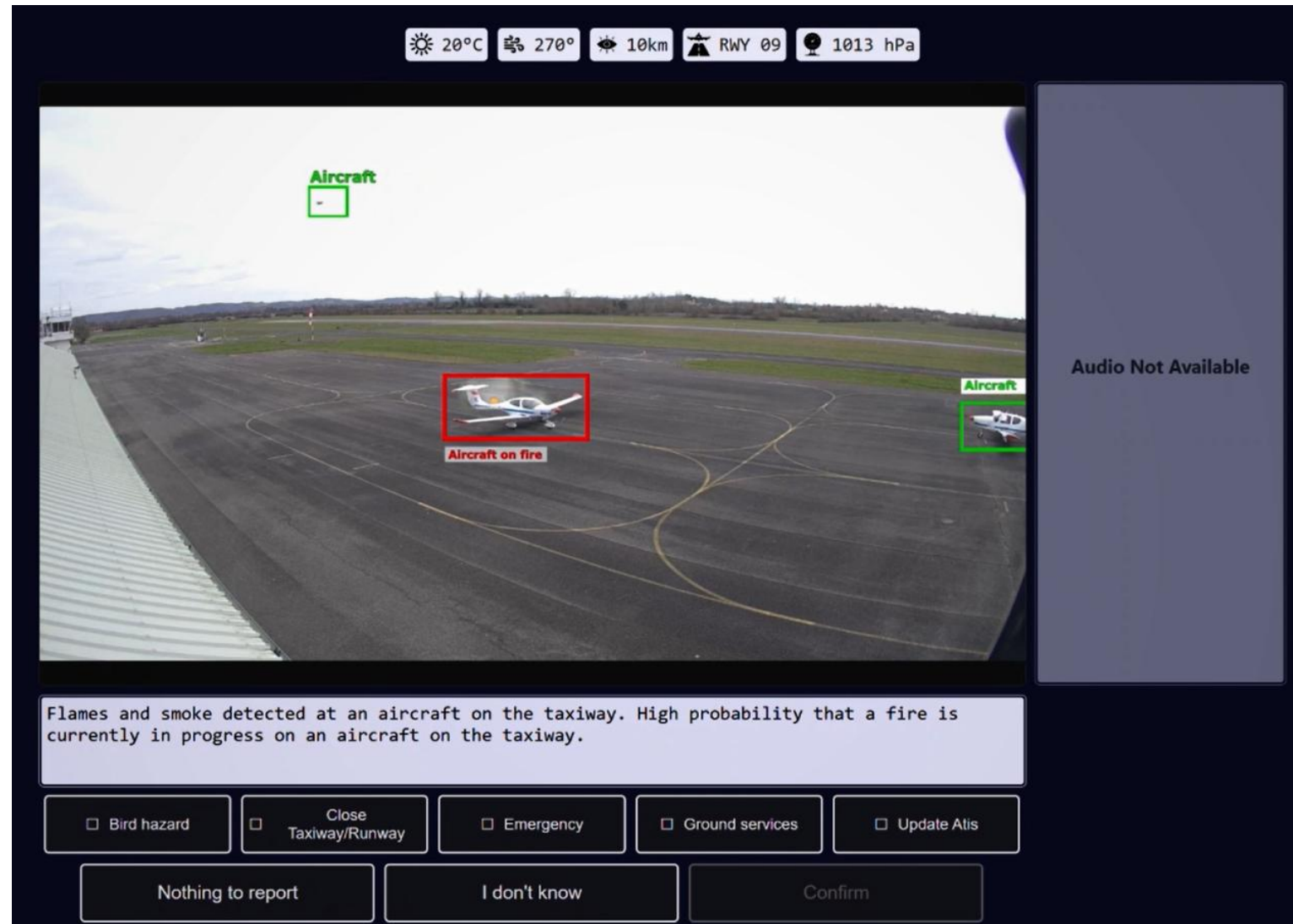
Solution: Explainable AI (XAI)

The "Glass Box" approach

To restore trust, AI must be transparent and interpretable. We should develop XAI interfaces that provide context-sensitive justifications for every alert. Operator's trust increases when AI reasoning is understandable

Key features

- **Visual explanations:** Overlays that show why a reroute is suggested (e.g., "Debris on Runway 15L Detected")
- **Confidence metrics:** Displaying the AI's certainty level (e.g., "95% Confidence")
- **Human-in-the-loop:** Allowing controllers to query "What-if" scenarios



TRUSTY

Solution: Adaptive Workload Management

Intelligent balancing

Systems should monitor real-time traffic density and controller biometric data (e.g., eye tracking) to prevent cognitive overload

Dynamic task allocation

During low workload, the system encourages manual tasks to maintain engagement. During peaks, AI autonomously handles routine tasks (e.g., handoffs)

Smart alerting

Moving beyond simple alarms to intelligent information management

- **Contextual filtering:** Suppress minor notifications during critical maneuvers
- **Prioritization:** Highlight only the most urgent conflict detections
- **Predictive overlays:** Project future states to aid "ahead-of-the-curve" planning



Building on the identified gaps, the TRUSTY project adopted a double funnel methodology to identify the main trust-related Key Performance Indicators for AI integration and Human-AI Teaming

Top-down approach

“a-priori”

Scientifically based ideas about the components of trust, originating from a literature analysis

Bottom-up approach

“a- posteriori”

Empirically based findings grounded in a project workshop with stakeholders and end-users

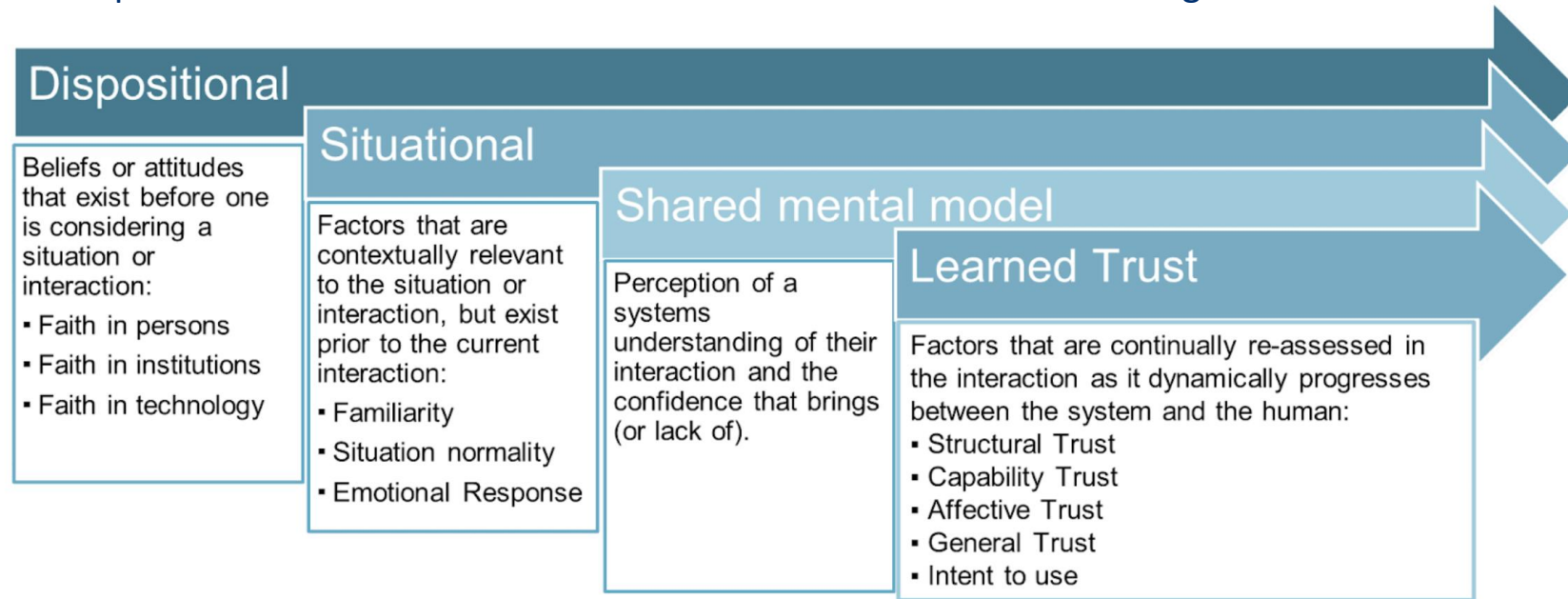
TRUSTY

Trust in RDTs - Top-down Approach

The model highlights four components¹:

- (i) Dispositional trust - attitudes toward people, institutions, and technology
- (ii) Situational trust - familiarity, context, and emotions
- (iii) Shared mental models - alignment between human and system understanding
- (iv) Learned trust - adaptation through experience

These dimensions explain how controllers build and calibrate trust in AI for Remote Digital Towers



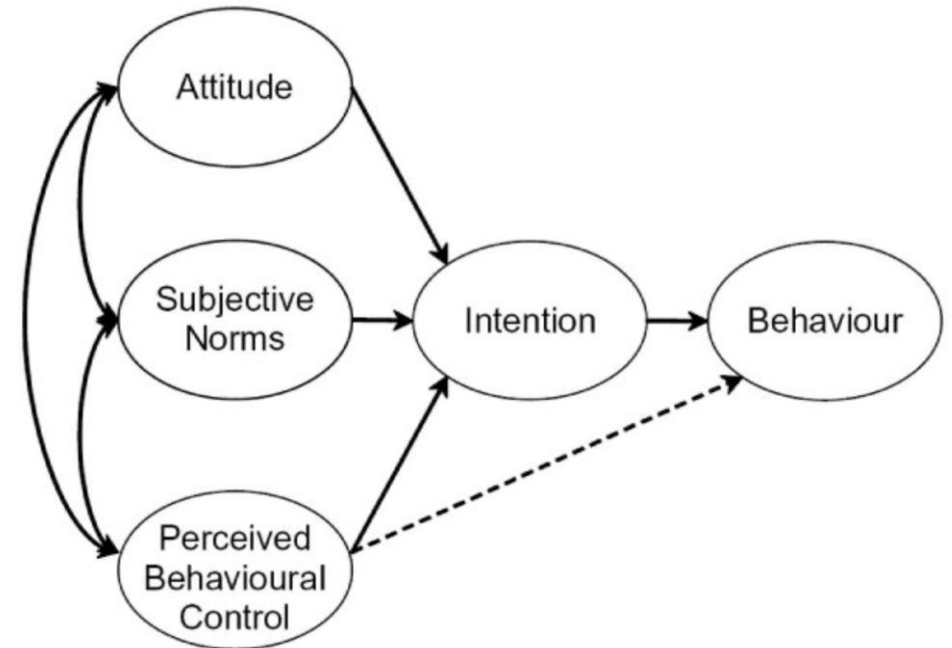
Factors of trust

1. Yosef S. Razin and Karen M. Feigh. 2024. Converging Measures and an Emergent Model: A Meta-Analysis of Human-Machine Trust Questionnaires. ACM Trans. Hum.-Robot Interact. 13, 4, Article 58

The multi-factorial trust model is integrated with the Theory of Planned Behaviour (TPB). This framework provides a pathway to quantify trust as an operational performance indicator¹

Mapping Constructs:

- Dispositional Trust → Attitude
- Shared Mental Model → Subjective Norms
- Explainability and Transparency → Perceived Behavioural Control
- Learned Trust → Teaming Behaviour Outcome



Theory of Planned Behaviour (TPB)

A total of **13** experts participated in the workshop held on 25 January 2024, representing a diverse cross-section of the aviation and ATM community and a broad geographical representation across Europe

Seven different countries: Sweden, France, Italy, Spain, Switzerland, Serbia, and Finland

The group included:

- Professionals from air navigation service providers (ENAV, SMATSA, Skyway)
- Industry and system manufacturers (Saab AB, ATM suppliers)
- Professional organisations (IFATCA, ATCU/ANACNA)
- Specialists in domains such as human factors, safety, and AI development

This diversity of backgrounds created a balanced mix of operational, regulatory, industrial, and research perspectives

Dimension	KPI	Description / Measurement Approach
Trust	Explanation Utility Score	Subjective rating (Likert scale) of the clarity and usefulness of AI explanations after tasks/events.
	Reliance Appropriateness Index	Ratio of correct AI advisories accepted minus incorrect advisories accepted, normalised by total advisories.
	Trust Calibration Consistency	Evolution of reliance across time; measures whether operator trust adapts proportionally to AI accuracy.
Situational Awareness	Time-to-Detect Critical Event	Average time (s) for ATCOs to detect conflicts/incursions with vs. without AI support.
	Situational Awareness Rating	Subjective post-scenario assessment of situational awareness in AI-assisted vs. baseline conditions.
Workload & HMI	Unified Operational Picture Consistency	Percentage of scenarios where fused AI outputs align with operator's mental model of the traffic situation.
	NASA-TLX Delta	Change in workload index between AI-assisted and baseline conditions across traffic scenarios.
	Alert Burden Ratio	Number of non-critical alerts per critical alert (lower values indicate more efficient alerting).
	Adaptive HMI Responsiveness	Time (ms) for HMI to adapt displayed information density to traffic or workload changes.
System Performance	False Alarm / Miss Rate	Proportion of incorrect anomaly detections (false positives/negatives) across scenarios.
	Recovery Time from Disruptions	Time (s) for RDT system to resume stable operation after a digital disruption or system degradation.

Key performance indicators (KPIs) for trustworthy AI integration in remote digital towers.

- **Good HAIT** focuses on **resilient socio-technical systems** where controllers and AI jointly sustain safety, efficiency, and scalability
- AI must enhance, not erode, controller expertise, situational awareness, and decision authority
- Building **transparent and resilient RDT systems** depends mainly on **calibrated trust, clear human-AI role allocation, and user-centered interfaces**
- **Trust is the central determinant of successful adoption.** Transparent, interpretable outputs are critical for gradual trust calibration. Implementation should follow a phased approach (e.g., AI begins in advisory capacity, progressing to semi-automation based on proven reliability and controller trust ratings)
- Over-reliance risks "out-of-the-loop" vulnerabilities, while under-trust undermines potential benefits. Thus, clear role definition, for both routine and emergency scenarios, and adaptive task allocation are essential to prevent over- or under-reliance

- Efficient adaptive task allocation should be based on balancing algorithms that leverage real-time data on task complexity, traffic density, and controller status (including biometric indicators)
- Iterative feedback loops should be integrated into the deployment process by embedding mechanisms in the HMI allowing operators to contest AI outputs, enabling the system to learn and reinforcing user confidence

Future Directions:

- Research should refine dynamic trust calibration models and pursue adaptive AI systems capable of tailoring task allocation based on real-time workload/behavioural indicators
- Longitudinal studies are needed to evaluate the sustained effects of XAI on performance, well-being, and trust

TRUSTY

Thank you for your attention

alexandre.veyrie@enac.fr

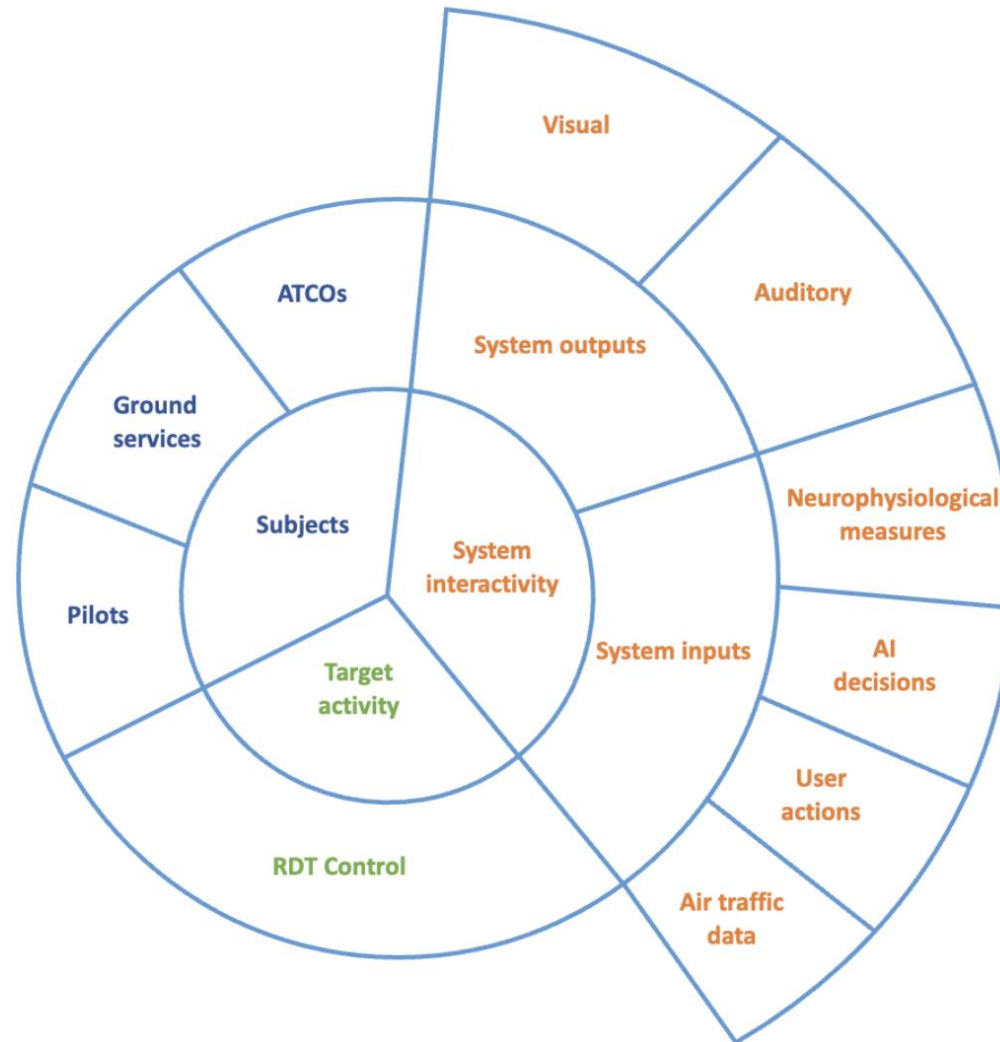
christophe.hurter@enac.fr

mobyen.uddin.ahmed@mdu.se

elizabeth.humm@dblue.it

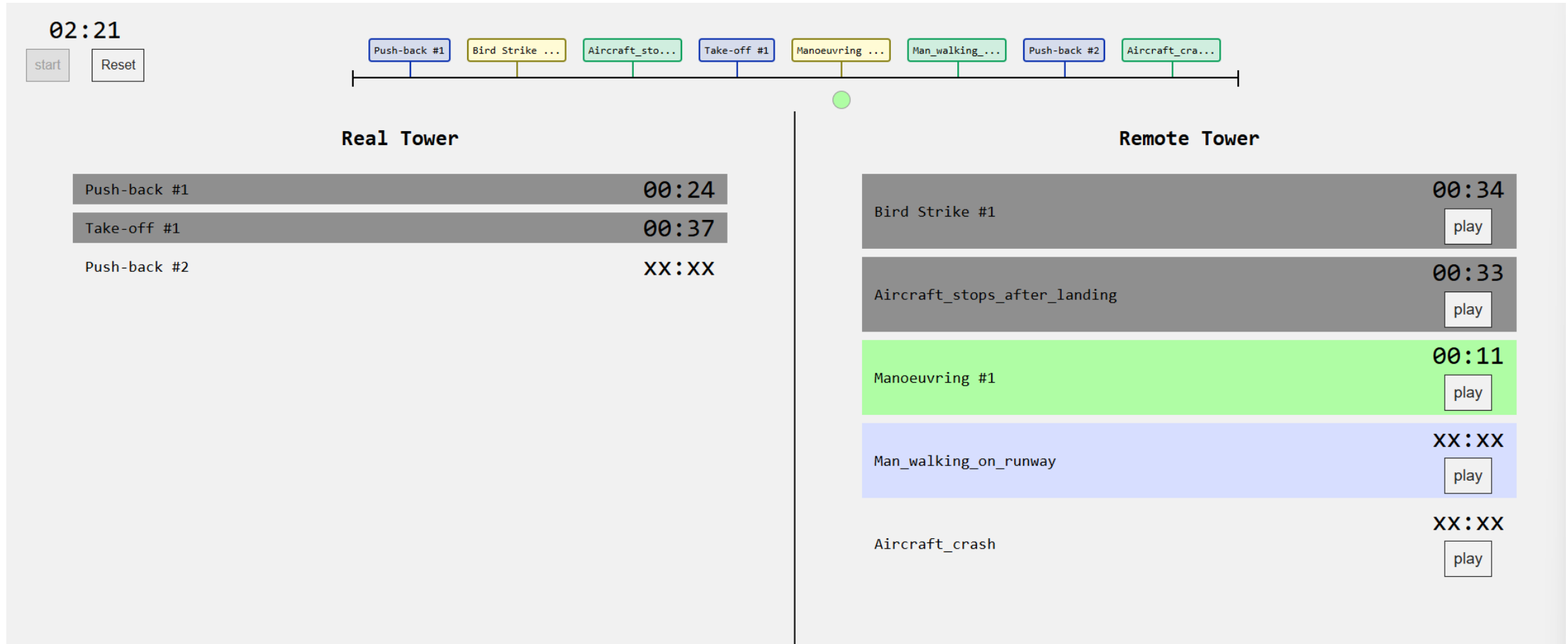
TRUSTY

Multi-Modal Approach



TRUSTY

Human-Machine Interfaces



sesar
JOINT UNDERTAKING

20°C
 270°
 10km
 RWY 09
 1013 hPa

Audio Not Available

Flames and smoke detected at an aircraft on the taxiway. High probability that a fire is currently in progress on an aircraft on the taxiway.

☐ Bird hazard

☐ Close Taxiway/Runway

☐ Emergency

☐ Ground services

☐ Update Atis

Nothing to report

I don't know

Confirm

sesar
JOINT UNDERTAKING

Sun

20°C

Wind

270°

Eye

10km

RWY 09

Pressure

1013 hPa

Audio Not Available

Critical situation. A human figure appears to be moving on the runway. High probability that a man is walking on the runway. An aircraft is currently transferring for lining up. Close the runway 30L. Contact ground services.

☐ Bird hazard

☐ Close Taxiway/Runway

☐ Emergency

☐ Ground services

☐ Update Atis

Nothing to report

I don't know

Confirm

☀ 20°C 🌧 270° 👁 10km ⬆ RWY 09 🌑 101°
Pour quitter le mode plein écran, appuyez de manière prolongée sur Échap

09:12 - Muret Ground, Delta Lima two three five, DR400, **Bird strike during climb. Critical Issue.** Request immediate return to **airdrom.**

Critical situation. An aircraft appears to have stopped on the runway after landing. High probability that the aircraft has technical issue. Close the runway 30L. Contact ground services and the pilot aircraft.

❑ Bird hazard

☑ Close Taxiway/Runway

❑ Emergency

☑ Ground services

❑ Update Atis

Confirm

Nothing to report

TRUSTY

Levels of Automation - EASA

ATCOs will be **supported up to a Level 1-** corresponding to **EASA 2A level**. Full authority must be kept in terms of decision of action to be implemented.

EASA AI level 1B – Human assistance

Dynamic trajectory optimization: suggest trajectories that optimize sequencing while considering aircraft trajectory, profiles and speeds, ensuring alignment with the appropriate sequence of arrivals.

DEFINITION	EASA AI level	PERCEPTION Information acquisition and exchange	ANALYSIS Information analysis	DECISION Decision and action selection	EXECUTION Action implementation	Authority of the human operator
LEVEL 0 LOW AUTOMATION Automation gathers and exchanges data. It analyses and prepares all available information for the human operator. The human operator takes all decisions and implements them (with or without execution support).	1A	●	●		◐	 FULL
LEVEL 1 DECISION SUPPORT Automation supports the human operator in action selection by providing a solution space and/or multiple options. The human operator implements the actions (with or without execution support).	1B	●	●	◐	◐	 FULL
LEVEL 2 RESOLUTION SUPPORT Automation proposes the optimal solution in the solution space. The human operator validates the optimal solution or comes up with a different solution. Automation implements the actions when due and if safe. Automation acts under direction.	2A	●	●	◐	●	 FULL
LEVEL 3 CONDITIONAL AUTOMATION Automation selects the optimal solution and implements the respective actions when due and if safe. The human operator supervises automation and overrides or improves decisions that are not deemed appropriate. Automation acts under human supervision.	2B	●	●	●	●	 PARTIAL
LEVEL 4 CONFINED AUTOMATION Automation takes all decisions and implements all actions silently within the confines of a predefined scope. Automation requests the human operator to supervise its operation if outside the predefined scope. Any human intervention results in a reversion to Level 3. Automation acts under human safeguarding.	3A	●	●	●	●	 LIMITED

TRUSTY

Metric	Mapped TRUSTY KPIs	How it Connects	TRUSTY KPI Dimension
Trust Scale Scores (Trust in Establishment)	Trust Calibration Consistency; Reliance Appropriateness Index	Organisational trust influences how reliance is adjusted to AI performance.	Trust
Trust Scale Scores (Personal Trust in Technology)	Trust Calibration Consistency; Reliance Appropriateness Index	Personal tech attitudes shape initial and ongoing reliance patterns.	Trust
Knowledge Retention Rates	Explanation Utility Score; Unified Operational Picture Consistency	Better system understanding increases explanation usefulness and mental model alignment.	Trust / SA
Simulation Performance Scores	Reliance Appropriateness Index; Time-to-Detect Critical Event; Situational Awareness Rating	Scenario performance shows trust-in-use and improvement in detection & SA.	Trust / SA
Frequency of XAI Use in Critical Situations	Trust Calibration Consistency; Reliance Appropriateness Index	Reveals under-trust or over-trust depending on frequency under criticality.	Trust
Incident Outcome Statistics	False Alarm / Miss Rate; Recovery Time from Disruptions; UOP Consistency	Incident outcomes indicate system error patterns, resilience and SA alignment.	System Perf / SA
Feedback Frequency & Quality	Explanation Utility Score; Trust Calibration Consistency; Adaptive HMI Responsiveness	Rich feedback improves XAI explanations, trust evolution, and adaptive interface behaviour.	Trust / Workload-HMI
Trust Trajectory Curves	Trust Calibration Consistency; Reliance Appropriateness Index	Longitudinal trust change directly feeds trust KPIs.	Trust
Accuracy Rates (System Output vs Reality)	False Alarm / Miss Rate; Unified Operational Picture Consistency	Direct measure of XAI operational correctness and SA alignment.	System Perf / SA
Predictability Scores	Trust Calibration Consistency; Reliance Appropriateness Index	Stable behaviour fosters proportional reliance.	Trust
Usage Frequency Metrics	Reliance Appropriateness Index; NASA-TLX Delta; Alert Burden Ratio	Shows over- or under-reliance and its workload impact.	Trust / Workload-HMI