

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

Sara Kebir^{†*}, Alexandre Veyri ^{†*}, Guillaume Truong^{ †}, Elizabeth Humm^{ ‡}, Nicola Cavagnetto^{ ‡},
Mobyen Uddin Ahmed^{ §}, Shaibal Barua^{ §}, Mir Riyanul Islam^{ §}, Giulia Cartocci^{ ¶},
Pietro Aric ^{ ¶}, Christophe Hurter^{ †}

[†]F d ration ENAC ISAE-SUPAERO ONREA, Universit  de Toulouse, Toulouse, France

[‡]Deep Blue s.r.l., Rome, Italy

[§]School of Innovation, Design and Engineering, M lardalen University, V ster s, Sweden

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

*Corresponding Authors

Abstract—Remote Digital Towers (RDTs) use multimodal feeds and AI to support air traffic controllers. While AI offers vigilance and consistency, trust remains essential to preserve human authority in safety-critical operations. This paper presents: (1) a review of human–AI collaboration in RDTs, (2) a gap analysis of trust, task allocation, and human–machine interface (HMI) design, and (3) integration strategies including explainable AI, adaptive workload management, and progressive trust-building. We propose initial Key Performance Indicators (KPIs) linking psychological trust constructs to operational outcomes. Our contribution clarifies how to deploy AI in RDTs without compromising safety, transparency, or human agency.

Keywords—Air Traffic Control, Explainable Artificial Intelligence, Gap Analysis, Remote Digital Towers, Trust.

I. INTRODUCTION

Air Traffic Management (ATM) and Air Traffic Control (ATC) systems are undergoing a significant transformation driven by digital technologies. Among the most prominent innovations is the concept of Remote Digital Towers (RDTs), which leverage high-resolution cameras, advanced sensors, and networked systems to provide Air Traffic Services (ATS) remotely. This paradigm enables cost-efficiency and continuity of service, particularly for low-traffic or geographically remote airports, while consolidating operations across multiple locations [1]. The integration of Artificial Intelligence (AI) further extends this transformation, with solutions such as the DiTA Digital Tower Assistant illustrating how AI can act as a “digital colleague” to support controllers [2]. However, trust in such systems depends heavily on their demonstrated competence, transparency, and incremental introduction into operations.

While technical feasibility of RDTs has been demonstrated, systematic approaches for calibrating trust in AI support, especially under multi-airport supervision and time-critical conditions, remain under-specified. Existing accounts rarely translate psychological trust constructs into operational Key Performance Indicators (KPIs) that can guide certification, interface design, and deployment. The shift from traditional towers to AI-enabled RDTs introduces considerable socio-technical challenges. Controllers face heightened mental and

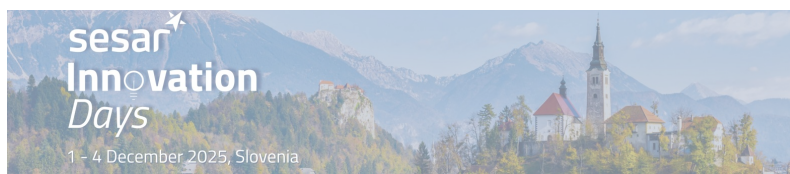
temporal demands, particularly when managing multiple remote towers simultaneously, a situation exacerbated during traffic peaks or unexpected events [3], [4]. Effective deployment of RDTs therefore requires attention not only to technological robustness but also to ergonomic interface design, cognitive workload management, and the establishment of training environments that faithfully replicate real-world operations. Central to the operational acceptance of AI-enabled RDTs is trust. Unlike conventional automation, AI systems frequently operate in adaptive and non-deterministic ways, making their behaviour difficult to predict or explain. When operators do not understand the rationale behind AI recommendations, they are less likely to accept them, even when correct [5]. ATM-specific reviews emphasise explainability, cognitive alignment, and transparency as prerequisites for trusted ATCO-AI systems [6]. Without these, trust, and by extension operational reliability, is compromised.

This paper (i) conducts a comprehensive gap analysis of trust, task allocation, and Human machine Interface (HMI) mechanisms for RDTs; (ii) proposes XAI-driven, human-centric solutions; (iii) outlines a methodological framework for integration and validation in operational contexts; and (iv) establishes initial KPIs to quantify effects on trust, usability, and safety. Our analysis aligns with the EASA AI Roadmap emphasis on transparency and learning assurance [7]–[9] and targets the Single European Sky ATM Research (SESAR) priorities for safe digital tower deployment [10].

II. LITERATURE REVIEW AND BACKGROUND

A. Remote Digital Towers and Artificial Intelligence

Remote Digital Towers (RDTs) allow ATS provision from locations other than the served airport, combining pan-tilt-zoom cameras, infrared sensors, and augmented reality overlays to present a rich airfield picture to ATCOs [1]. Early operations in Sweden (2015) and at Scandinavian Mountains (2019) catalysed broader adoption, with further deployments (e.g., London City) demonstrating scalability and redundancy benefits [11]. Across deployments, however, evaluations consistently surface concerns about peak-workload management,



human-computer interaction, and resilience to digital disruptions, issues that become acute in multiple remote tower settings [12], [13]. Digital transformation initiatives and increasing use of AI for prediction and anomaly detection create strong preconditions for RDT-AI integration [10], but operational acceptance hinges on human-centric design and trusted AI behaviour [14].

Artificial intelligence technologies are increasingly integral to RDT systems, enhancing both operational efficiency and safety. Applications span routine perception tasks such as object recognition, surface movement monitoring, and weather assessment, as well as more advanced functions including conflict detection, predictive decision support, and anomaly identification. Machine learning models trained on large operational datasets enable these systems to generalize from historical patterns while adapting to evolving contexts.

Operational deployments illustrate the maturity of AI-enabled RDT solutions. Indra’s Advanced Remote Tower at Budapest Airport integrates AI-driven 4K visual coverage, automated flight tracking, and anomaly detection. Frequentis has implemented RDT platforms at German airports (e.g., Leipzig and Erfurt-Weimar), leveraging AI for high-definition video feeds and advanced analytics. Similarly, London Heathrow’s Virtual Control Facility, developed by NATS, employs AI to maintain operational continuity under reduced visibility.

Thus, AI-based monitoring has been used to detect runway incursions, unauthorized drone activity, and foreign object debris, often outperforming humans in speed and consistency. This facilitates timely interventions and supports multi-airport operations from a single facility, a particularly valuable feature for regional airports. By automating routine tasks such as aircraft identification, tracking, and communication, AI alleviates cognitive demands and allows controllers to concentrate on complex decision-making.

Despite these benefits, several challenges remain. A persistent issue is data imbalance: models are trained on abundant nominal data but face sparse abnormal or emergency scenarios, limiting reliability in critical events. This motivates the use of synthetic data and targeted scenario generation. Moreover, many high-performing models are opaque, complicating error tracing and root cause analysis. For RDT contexts, effective AI support must therefore be paired with human-machine interfaces that reveal why the system recommends an action and how confident it is, without delaying interventions. Ensuring robustness beyond the nominal envelope, resilience against cyber threats, and regulatory compliance further underscores the need for explainable and trustworthy AI integration into RDT operations.

B. Importance of Explainable Artificial Intelligence (XAI)

Explainable AI (XAI) addresses the “black-box” challenge of many AI systems by making outputs more transparent and interpretable for human users. In RDTs and ATM, explainability is essential to foster trust in safety-critical contexts, where accountability and situational awareness are paramount. XAI provides context-sensitive justifications for AI-generated

alerts, predictions, and recommendations, ensuring alignment with human mental models and supporting informed decision-making. Empirical studies confirm that operator trust increases when AI reasoning is understandable and complements human judgment [5], [15]. Consequently, XAI is a foundational enabler of human-centric AI integration in aviation.

Operational demonstrations underscore XAI’s value by showing how explainable outputs can help controllers validate AI recommendations, improve situational awareness, and enhance decision support. This is further reinforced by regulatory perspectives: The European Union Aviation Safety Agency (EASA) identifies transparency as central to trustworthy AI in aviation, comprising three interdependent dimensions: (i) explainability, which ensures operators can understand and validate AI decisions; (ii) learning assurance, requiring rigorous training, validation, and testing to guarantee reliable model performance within defined operational domains; and (iii) ethics-based assessment, which evaluates systems against principles such as fairness, accountability, and respect for human rights. Together, these measures promote both technical robustness and societal responsibility in AI deployment.

Explainability supports safety-critical operations, regulatory oversight, and stakeholder acceptance. By revealing the rationale and limitations of AI outputs, XAI enables error detection, builds operator trust, and enhances human-AI collaboration.

Despite significant progress, explainable AI in ATM continues to face important challenges. Complex models such as deep neural networks remain inherently opaque, often requiring post-hoc methods like SHAP and LIME to approximate their decision processes [16]. Explainability also involves trade-offs, as more interpretable models may reduce predictive accuracy. A further difficulty is scalability, since explanations must be delivered in real time to meet operational constraints without compromising safety. To mitigate these issues, various strategies have been proposed, including model-agnostic techniques that attribute predictions to input features, visualization methods such as overlays, bounding boxes, and heat maps that make reasoning accessible within operational displays, simplified proxy models that approximate complex systems while remaining interpretable, and interactive explanations that allow “why” and “what-if” queries to enhance operator understanding and engagement [17].

C. Human-AI Teaming

Human-AI Teaming (HAIT) in ATM and ATC is not merely about introducing automation but about enabling resilient socio-technical systems where controllers and AI jointly sustain safety, efficiency, and scalability. In these high-stakes environments, AI tools must enhance rather than erode controller expertise, situational awareness (SA), and decision authority [18]–[20]. Controllers must trust that AI-generated advisories, ranging from trajectory predictions to conflict detection and workload balancing, are not only accurate but also transparent, recoverable, and explainable. Trust calibration is therefore central: over-reliance risks out-of-the-loop vulnerabilities, while under-trust undermines the potential benefits of adaptive



decision support [21]. Embedding trustworthy AI in ATC requires design choices that foreground traceability of system logic, meaningful communication of uncertainty, and user-centred interfaces that integrate seamlessly into the controller's perception–comprehension–projection cycle. At the systemic level, adaptive function allocation should flexibly shift responsibilities between human and AI depending on real-time traffic complexity and operator workload, while preserving the controller's role as final decision-maker. Critically, regulatory roadmaps such as EASA's AI framework underline that operational deployment of AI in ATM must be accompanied by robust assurance cases, continuous monitoring, and demonstrable alignment with human performance envelopes [7]. In this way, HAIT can move from being a technical aspiration to a practical enabler of safe, transparent, and trustworthy digitalised ATS [22].

III. GAP ANALYSIS

To examine AI integration in RDTs and identify challenges, we performed a gap analysis. This analysis synthesizes key shortcomings in AI implementation and their implications for optimizing its use in ATM. Identified gaps pertain to: (i) trust and user-centric integration; (ii) task allocation and responsibility clarity between ATCOs and AI; and (iii) situational awareness and HMI design limitations.

A. Trust and User-Centric Integration

AI is now integral to RDT operations, supporting tasks such as monitoring, decision-making, and traffic optimisation. Yet several trust-related concerns remain. However, for successful adoption, AI systems must be human-centric, where trust in technology is critical. The integration of AI into RDTs has advanced, but trust-related concerns remain, particularly around AI's explainability, system transparency, and how well AI aligns with ATCO operational needs. Following are the identified trust related gaps:

Lack of explainability: Current AI systems in RDTs provide operational support but often lack sufficient transparency to explain why specific decisions or recommendations are made. This deficiency leads to operator skepticism, especially when AI outputs conflict with ATCO's intuition or established procedures. Without clear explanations, even accurate AI decisions are subject to rejection or underutilization, undermining system effectiveness.

Human-Centric Design (HCD) shortcomings: Existing RDT interfaces and AI workflows frequently overlook user needs, failing to account for cognitive workload, situational stress, and varying experience levels of ATCOs. This lack of Human-Centric design integration results in tools that may technically perform well but are not aligned with operator expectations or operational realities leading to potential inefficiencies and increased workload.

Missing pathways for building trust: Trust in automation is not static; it must be cultivated over time through repeated, positive interactions. Current AI-based RDT systems often lack mechanisms to support gradual trust calibration, such as

feedback loops, adjustable automation levels, or progressive exposure to AI capabilities. Besides, without a clear framework for how AI decisions are made and a way to verify and understand these decisions, ATCOs may hesitate to adopt the technology fully [21].

B. Task Allocation and Responsibility Clarity

The shift towards RDTs introduces complex task-sharing between human controllers and AI systems. AI systems can currently help for assisting with tasks such as runway incursion detection, communication, and traffic sequencing, but the balance between human and machine responsibilities is not clearly defined.

Ambiguity in Human-AI task distribution: It is unclear how tasks are distributed between ATCOs and AI systems, especially during high-traffic periods or in emergency situations. This creates a potential risk of overload or confusion for the human operator. Without well-defined role boundaries, the ATCOs may either over-rely on or underutilize AI systems, increasing the risk of human error. Role clarity is particularly important during transitions between manual and automated modes. The literature on joint cognitive systems emphasizes the need for explicitly defined responsibilities to avoid mode confusion and ensure effective human-AI teaming [23].

Limitations in emergency response support: Current AI systems in RDTs provide valuable support for routine operations but remain limited in their ability to assist ATCOs during abnormal or emergency events. In such situations, controllers must continue to make critical decisions independently, as AI outputs are not yet fully reliable or trustworthy.

A major limitation lies in the training of AI models, which predominantly rely on data from normal operations. As a result, they often lack robustness when confronted with rare or novel scenarios. This restricts their utility in crisis management, where adaptability is essential. The capacity of AI to generalize across scenarios is particularly critical in aviation, where unexpected edge cases can carry severe safety consequences.

Insufficient real-time workload management: Current RDT implementations show limited integration between AI systems and human operators in managing workload distribution. The absence of an effective real-time support and dynamic workload adjustments, responsive to both traffic density and operator status, can result in cognitive overload or understimulation, each of which negatively impacts performance and safety.

C. Situational Awareness and HMI Design Limitations

RDTs offer remote monitoring of airports through high-definition video feeds, sensors, and other surveillance tools. AI systems can assist ATCOs in maintaining situational awareness but need to be more interactive and adaptive to changing conditions.

1) Inadequate dynamic situational awareness: The inability to effectively integrate diverse data sources, such as visual, radar and audio inputs, into a unified operational picture

significantly impairs situational awareness. Current AI systems often lack the advanced capabilities necessary to dynamically enhance this awareness, particularly in contexts requiring the synthesis of multiple, heterogeneous information streams. As a result, ATCOs may not receive comprehensive or timely updates, especially when dealing with rapidly evolving conditions.

2) *Poor HMI design:* Human-machine interfaces in RDTs often fall short in supporting the seamless integration of AI assistance into ATCOs workflows. In many cases, existing interfaces do not provide the level of real-time feedback required for effective monitoring and management of air traffic, limiting the operational value of AI support. Besides, deficiencies in information management further constrain their effectiveness. The HMI may overload operators with non-essential data or, conversely, fail to provide critical information when needed, leading to inefficient decision-making which can negatively affect operational safety.

IV. ADDRESSING THE IDENTIFIED GAPS

A. Enhancing trust and explainability through XAI

1) *Implementation of clear explanations:* To enhance trust and usability, AI-based decision support systems for ATC should incorporate several key features. First, explainability frameworks are essential to visually clarify the AI's decision-making process, illustrating not only the outcomes but also the underlying reasoning behind specific recommendations, such as rerouting an aircraft or delaying take-off due to adverse weather conditions. Complementing this, confidence metrics should be implemented to provide quantifiable reliability indicators, such as percentage scores, associated with each recommendation, which could be intuitively conveyed through color-coded visual elements on operational dashboards. Furthermore, robust data integration is required, enabling the system to process and cross-validate information from diverse sources, including radar, weather sensors, and visual feeds, thereby ensuring transparent and well-founded recommendations with direct access to the supporting data. Finally, interpretability must be prioritized for non-technical users, ensuring that explanations are clear and visually intuitive, allowing ATCOs to readily understand and act upon the AI's outputs.

2) *Iterative, user-centric feedback loops:* To promote a co-adaptive relationship between human controllers and AI systems, iterative feedback mechanisms should be embedded into the HMI. These allow operators to assess, confirm, or contest AI outputs, enabling the system to learn from operational context and controller input over time. Drawing from interactive machine learning paradigms, the feedback loop enhances system personalization while reinforcing user confidence through consistent responsiveness. Thus, leading to smoother technology adoption.

3) *Gradual trust-building frameworks:* Given the safety-critical nature of ATM, a phased approach to trust development should be implemented. Initially, AI modules act in an advisory capacity, with decision authority fully retained by the operator. Over time, and contingent upon proven

system reliability and controller trust ratings, selected tasks (e.g., routine traffic monitoring) may be semi-automated. This graduated framework should follow established automation trust calibration models, ensuring that trust is neither prematurely inflated nor unjustly suppressed. Alongside, an AI trust-building process through training programs and live operational scenarios can be created to allow ATCOs to learn and develop confidence in AI systems progressively.

B. Clarifying Task Allocation and Responsibility

1) *Development of clear task allocation guidelines:* Clear guidelines should be defined for task allocation between the AI and ATCOs, ensuring that ATCOs understand the role of AI systems in both routine and emergency scenarios.

2) *Improved algorithms for emergency situations:* AI should be capable of handling increased complexity during peak times or emergencies. Thus, we should develop algorithms that provide ATCOs with real-time suggestions backed by transparent reasoning. AI behavior under non-nominal conditions must prioritize controller awareness and timely handover. Emergency-specific reasoning algorithms trained on historical incident data and validated through synthetic emergency scenarios should be developed to detect anomalies and trigger prioritized alerts. These models should incorporate uncertainty estimation to human-in-the-loop control when confidence levels drop.

3) *Intelligent workload distribution algorithms:* To avoid cognitive overload and underload in RDTs, intelligent workload balancing algorithms should be introduced. These leverage real-time data on task complexity, airspace density, and controller workload (including biometric indicators) to adapt task allocation and interface prioritization. AI offers significant potential to address this challenge by monitoring physiological and behavioral indicators of workload and enabling adaptive redistribution of tasks and prioritization of alerts.

C. Improving situational awareness and user interfaces

1) *Integrated multi-source situational awareness system:* AI systems must be designed to intelligently fuse data and present it in a coherent and actionable manner. The incorporation of high-fidelity visualizations and predictive overlays is essential for helping operators construct and maintain accurate mental models of both current and anticipated airspace states, thereby supporting informed decision-making in real time.

2) *Adaptive and customizable HMI:* HMI design should be improved by offering customizable user interfaces that allow ATCOs to adjust the amount of information displayed based on task complexity or urgency. AI systems should provide just-in-time information and visual data that assist with decision-making.

3) *Smart alert systems for relevant information:* An intelligent alerting system is needed to reduce information overload and ensure that ATCOs can focus on critical tasks without unnecessary distraction.

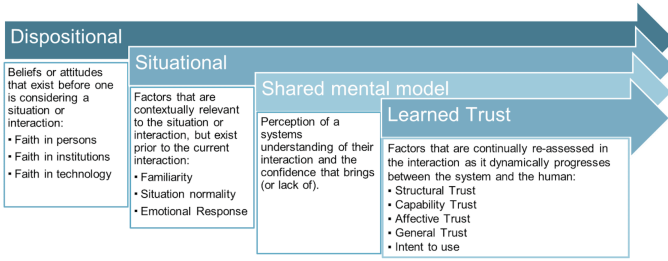


Figure 1. Factors of trust. The model highlights four components: (i) Dispositional trust—attitudes toward people, institutions, and technology; (ii) Situational trust—familiarity, context, and emotions; (iii) Shared mental models—alignment between human and system understanding; and (iv) Learned trust—adaptation through experience. These dimensions explain how controllers build and calibrate trust in AI for Remote Digital Towers.

V. MEASURING TRUST AND PERFORMANCE

Building on the identified gaps, the TRUSTY project consortium adopted a double funnel methodology to identify the main trust-related Key Performance Indicators (KPIs) for AI integration and Human-AI Teaming (HAIT). This methodology comprised a top-down and a bottom-up approach to define the KPIs from two different perspectives. Firstly, “a-priori” (i.e., scientifically based) ideas about the components of trust, originating from a literature analysis, were combined with “a-posteriori” (empirically based) findings grounded in a project workshop with stakeholders and end-users.

A. Top-down Approach

The top-down approach adopted a multi-factorial trust model encompassing four key levels (Figure 1):

- **Dispositional Trust:** Pre-existing beliefs or attitudes towards people, institutions, and technologies.
- **Situational Trust:** Context-specific factors influencing trust, such as familiarity with the AI system and emotional responses.
- **Shared Mental Model:** The human’s perception of the AI’s understanding of their interaction [24].
- **Learned Trust:** The dynamic nature of trust, continuously reassessed throughout human-AI interaction, considering structural trust, capability trust, affective trust, and willingness to use AI.

Building upon this foundation, we explored the potential for integrating trust into the Theory of Planned Behaviour [25] (Figure 2), through the lens of the Norm Activation Model [26] leveraging the work performed by [27]. The Theory of Planned Behaviour is a psychological theory that connects a person’s beliefs to their behaviour. The theory suggests that there are three core components which together shape an individual’s intentions in terms of behaviour. These are i) attitude; ii) subjective norms; and iii) perceived behavioural control, as represented in (Figure 2). Trust can be considered a factor influencing the components of this model.

The four levels of trust, namely dispositional, situational, shared mental model and learned trust can be considered part of the Theory of Planned Behaviour in the following way.

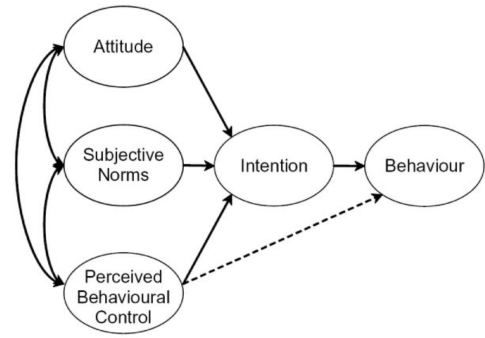


Figure 2. Theory of Planned Behaviour (TPB). Attitudes, subjective norms, and perceived behavioural control shape intention, which drives behaviour. In Remote Digital Towers, these constructs map to trust in AI systems: attitudes (dispositional trust), subjective norms (shared mental models), and perceived control (explainability and transparency), influencing controllers’ intention and behaviour with AI.

Dispositional Trust and Attitude: Our research conceptualised attitudes as a form of dispositional trust, reflecting general faith in people, institutions (including government, technology certifiers, and internal testing bodies), and technology itself.

Shared Mental Model and Subjective Norms: The Norm Activation Model suggests that a subjective norm arises from the combined awareness of both the necessity of a particular behaviour and the consequences associated with it [26]. Based on this model, we hypothesised that the shared mental model between humans and AI assistants could function as the subjective norm [24].

Shared Mental Model and Perceived Behavioural Control / Intention: Furthermore, we hypothesised that the explainability and transparency of an AI solution might enhance or reduce the perceived behavioural control of the human-AI teaming behaviour. This perceived control is expected to mediate the relationship between trust factors (particularly dispositional and subjective norm) and the intention to engage in teaming behaviour. We also considered the situational dimension of trust, acknowledging that the specific air traffic scenario where human ATCOs interact with the AI system will influence their trust perception.

Learned Trust and Behaviour: Finally, we proposed that the learned dimension of trust can be mapped onto the outcome of the teaming behaviour itself. Positive interactions would presumably lead to a higher willingness to continue teaming with the AI, while negative interactions would have the opposite effect.

Building upon these psychological models, it’s clear that trust is not instantaneous; therefore, understanding the path to trust is critical when designing and integrating novel technologies, particularly AI.

B. Bottom-Up Approach

To optimise human-system integration and understand ATCOs’ needs for trust in RDTs, we adopted a bottom-up approach, combining a workshop with structured qualitative interviews to capture insights from ATCOs, regulatory

practitioners, and equipment manufacturers. These activities identified key trust drivers and informed the extraction of trust-related KPIs, aligned with the EASA AI trustworthiness framework and EU guidelines [7]–[10].

1) *Methodology*: 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. Participants came from seven different countries, including Sweden, France, Italy, Spain, Switzerland, Serbia, and Finland, ensuring broad geographical representation across Europe. 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), and 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.

2) *Qualitative results*: The workshop analysis revealed the four interrelated key levels of trust discussed previously that characterize how ATCOs conceptualize and build trust in technological systems.

Dispositional trust was described as both organizational and personal. At the organizational level, participants emphasized confidence in technologies that had undergone structured processes of development, integration, and certification. At the personal level, trust was shaped by individual attitudes toward technology, as well as by detailed knowledge of how a system functions, including its features and potential sources of error. Participants also highlighted a strong sense of trust in colleagues, grounded in shared training, certification, and operational experience.

Situational trust was more context-dependent, reflecting the role and criticality of the technology. Controllers distinguished between safety net systems, which provide monitoring and alerts, and safety-critical systems, where reliability was considered non-negotiable. Trust was supported when systems served informative rather than intrusive roles, allowed cognitive engagement with their functioning, and were accompanied by clear procedures in the event of failure. Other important factors included knowledge of system capabilities and limitations, systematic recording of performance, and organizational clarity regarding liability. Participants also described trust as being strengthened when outputs aligned with operational sense-making, when technologies offered simple decision support in high workload conditions, and when redundancy or cross-checking mechanisms were available.

Shared mental models were viewed as critical for sustaining trust over time. Participants noted that technologies could be introduced gradually, beginning as supportive safety nets and progressively taking on greater responsibility as reliability and user confidence increased. Trust was also reinforced through assumptions of common mental models among ATCO colleagues, stemming from shared training and certification. Conversely, low-reliability systems were perceived as detrimental, increasing workload and distracting from core responsibilities. Continuous, often subconscious, analysis of both the

situation and system performance was described as part of routine practice, serving as a constant reinforcement of trust calibration.

Finally, learned trust developed iteratively through experience with technology. Participants reported that trust increased when system outputs consistently matched operational reality, when systems exhibited predictability over time, and when familiarity with the technology grew. Interaction with the system allowed trust to be adjusted upward or downward depending on performance. Emotional and social dimensions also played a role: shared experiences during high-stress events fostered interpersonal trust and provided emotional support. At the same time, poorly integrated alerting features or frequent false alarms were seen as disruptive, undermining both trust and operator focus.

To operationalise the findings of the gap analysis and the trust framework, a set of KPIs was derived for providing measurable evidence for the development and validation of trustworthy AI in RDTs. These KPIs span four dimensions: trust, situational awareness, workload and HMI, and system performance. They combine subjective assessments (e.g., perceived explanation utility, workload ratings) with objective behavioural and system-level measures (e.g., reliance appropriateness, detection times, false alarm rates, recovery from disruptions). As summarised in Table I, the indicators offer a concrete basis for linking psychological trust constructs with operational performance, thereby enabling systematic evaluation of AI-driven RDT solutions in both simulation and live settings. These KPIs are intended not only as evaluation tools but also as design targets, guiding system developers toward solutions that are transparent, adaptive, and aligned with ATCO cognitive needs.

VI. DISCUSSION

This study demonstrates both the potential and challenges of integrating AI into RDTs. Trust emerges as a central determinant of successful adoption, as AI's uncertainty and opacity can limit operator acceptance. Transparent, interpretable outputs aligned with ATCO cognitive models, combined with phased automation and adaptive feedback, are critical for gradual trust calibration [6], [22].

AI systems perform well in routine monitoring, anomaly detection, and predictive analytics but are less reliable in non-nominal or emergency scenarios, necessitating that controllers retain ultimate decision-making authority. To balance human and AI roles, clear role definitions and adaptive task allocation are essential to prevent over- or under-reliance while optimizing workload. Furthermore, effective AI support for situational awareness depends on HMI design that integrates multi-source data, adjusts information density to traffic complexity, and employs smart alerting to highlight critical events without causing distraction.

The introduction of a multi-factorial trust model [21], [22], [24] within this research represents a step toward systematically measuring how trust evolves in operational contexts. By linking dispositional, situational, shared mental models, and



TABLE I. KEY PERFORMANCE INDICATORS (KPIs) FOR TRUSTWORTHY AI INTEGRATION IN REMOTE DIGITAL TOWERS.

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.
	Unified Operational Picture Consistency	Percentage of scenarios where fused AI outputs align with operator's mental model of the traffic situation.
Workload & HMI	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.

learned trust to established behavioral theories, our framework provides a pathway to quantify trust as both a psychological construct and an operational performance indicator. This approach could inform certification, training, and continuous monitoring practices, ensuring that trust remains appropriately calibrated throughout the system lifecycle.

Despite the significant potential of AI in RDTs, several challenges must be addressed to ensure safe and trustworthy integration. A key concern is system robustness and security, as AI systems must be resilient against cyber threats in highly digitalized environments. Equally important are regulatory and standardization frameworks: the adoption of AI-assisted remote operations requires compliance with stringent safety standards, extensive testing, and close collaboration with regulatory bodies, as well as adequate training for personnel.

Another challenge lies in balancing model accuracy and interpretability. Highly interpretable models may sacrifice predictive accuracy, whereas more accurate models may be less explainable, making the trade-off a central issue. Scalability in real-time operations further complicates XAI integration, as explainability tools must process large volumes of data without introducing delays.

Data quality and availability also remain critical, since incomplete or inconsistent datasets undermine AI performance. Moreover, integration with legacy ATM systems presents technical and operational difficulties, requiring interoperable interfaces and comprehensive training. Finally, trust in AI systems adds another layer of complexity: trust evolves gradually and must therefore be measured and managed over time to ensure sustained operator confidence [6], [22], [28].

VII. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

This paper addressed the urgent need to strengthen trust and reliability in AI applications within RDTs to ensure safer and more efficient ATM. A comprehensive gap analysis identified three critical challenges: (i) trust and human-centric integration, requiring transparent, explainable outputs aligned with

ATCO workflows and gradual trust-building; (ii) task sharing and responsibility, demanding clear allocation guidelines and AI support for abnormal or high-traffic situations; and (iii) situational awareness and interface design, necessitating intuitive, customizable HMIs and intelligent alerting to reduce cognitive overload. To close these gaps, the project proposed actionable strategies including the development of interpretable XAI features, iterative user-centred design processes, calibrated trust frameworks, algorithms for workload balancing and emergency response, and real-time multimodal data integration. Initial KPIs were identified for trust, situational awareness, and workload management, guiding future system validation and operational deployment.

Looking ahead, future research must push the boundaries of HAIT in ATM. Advancing autonomous yet human-supervised decision-making systems will be key to leveraging AI's predictive and optimisation power without compromising oversight. Research should refine trust calibration models that dynamically capture how trust evolves with system performance, experience, and operator state, and should pursue adaptive AI systems capable of tailoring task allocation and interface prioritisation to real-time workload and behavioural indicators. Progress in generalised XAI techniques is essential to deliver explanations that remain intuitive and accurate across models and operational contexts while preserving efficiency. Crucially, longitudinal studies of HAIT are needed to evaluate the sustained effects of XAI on controller performance, well-being, and trust in live operations.

In conclusion, this paper highlights that the success of RDTs depends not only on advanced sensing and AI, but above all on calibrated trust, clear human-AI role allocation, and user-centred interfaces. By linking psychological trust models with operational KPIs and translating them into actionable design requirements, a concrete pathway is provided for building RDT systems that are transparent, resilient, and accepted by controllers.

ACKNOWLEDGMENT

This research was conducted within the framework of the TRUSTY project, which has received funding from the SESAR 3 Joint Undertaking (JU) under grant agreement No 101114838.

REFERENCES

- [1] K. Yamazaki, S. Inoue, H. Aoyama, and T. Kimiaki, "Remote digital tower to support air traffic control systems," *Advances in Human Factors of Transportation*, vol. 148, no. 148, 2024.
- [2] C. Westin, S. Boonsong, B. Josefsson, and J. Lundberg, "Building trust in autonomous system competence—the dita digital tower assistant for multiple remote towers, an early concept evaluation," in *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*. IEEE, 2020, pp. 1–9.
- [3] P. Kearney, W.-C. Li, J. Zhang, G. Braithwaite, and L. Wang, "Human performance assessment of a single air traffic controller conducting multiple remote tower operations," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 30, no. 2, pp. 114–123, 2020.
- [4] C. Liu, Y. Lu, and T. Mou, "Research on ergonomic evaluation of remote air traffic control tower of hangzhou jiangde airport," in *International Conference on Human-Computer Interaction*. Springer, 2023, pp. 306–319.
- [5] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.
- [6] J. Y. Zhong, "Heading toward trusted atco-ai systems: A literature review," *Engineering Archive*, 2021.
- [7] EASA, "Artificial intelligence roadmap. a human-centric approach to ai in aviation," Tech. Rep., 2020.
- [8] —, "Artificial intelligence concept paper issue 2: Guidance for level 1 2 machine learning applications," European Aviation Safety Agency, Tech. Rep., 2024.
- [9] —, "Technical and operational requirements for remote tower operations," Tech. Rep., 2017.
- [10] SESAR Joint Undertaking, "European atm master plan: 2025 edition," SESAR Joint Undertaking, Brussels, Belgium, Tech. Rep., Dec. 2024.
- [11] EUROCONTROL, "European atm master plan implementation report," Tech. Rep., 2021.
- [12] A. Papenfuss and M. Friedrich, "Head up only—a design concept to enable multiple remote tower operations," in *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*. IEEE, 2016, pp. 1–10.
- [13] M. Peukert, L. Meyer, and B. Josefsson, "Multi remote tower-challenge or chance? an empirical study of air traffic controllers performance," in *International Conference on Human-Computer Interaction*. Springer, 2020, pp. 611–618.
- [14] R. M. Jacobsen, J. Wester, H. B. Djernæs, and N. van Berkel, "Distributed cognition for ai-supported remote operations: Challenges and research directions," *arXiv preprint arXiv:2504.14996*, 2025.
- [15] M. T. Ribeiro, S. Singh, and C. Guestrin, "“why should i trust you?”: Explaining the predictions of any classifier," 2016.
- [16] A. M. Salih, Z. Raisi-Estabragh, I. B. Galazzo, P. Radeva, S. E. Petersen, K. Lekadir, and G. Menegaz, "A perspective on explainable artificial intelligence methods: Shap and lime," *Advanced Intelligent Systems*, vol. 7, no. 1, Jun. 2024. [Online]. Available: <http://dx.doi.org/10.1002/aisy.202400304>
- [17] A. Degas, M. Islam, C. Hurter, S. Barua, H. Rahman, M. Poudel, D. Ruscio, M. Ahmed, S. Begum, M. Rahman, S. Bonelli, G. Cartocci, G. Di Flumeri, G. Borghini, F. Babiloni, and P. Aricò, "A survey on artificial intelligence (ai) and explainable ai in air traffic management: Current trends and development with future research trajectory," *Applied Sciences*, vol. 12, no. 3, p. 1295, 2022.
- [18] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," in *Situational awareness*. Routledge, 2017, pp. 9–42.
- [19] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, vol. 30, no. 3, pp. 286–297, 2000.
- [20] NASEM, "Human-ai teaming: State-of-the-art and research needs," *National Academies of Sciences Engineering and Medicine*. Washington, DC: The National Academies Press, vol. 10, p. 26355, 2022.
- [21] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human factors*, vol. 46, no. 1, pp. 50–80, 2004.
- [22] M. R. Endsley, "Supporting human-ai teams: Transparency, explainability, and situation awareness," *Computers in Human Behavior*, vol. 140, p. 107574, 2023.
- [23] M. R. Endsley and D. B. Kaber, "Level of automation effects on performance, situation awareness and workload in a dynamic control task," *Ergonomics*, vol. 42, no. 3, pp. 462–492, 1999.
- [24] R. W. Andrews, J. M. Lilly, D. Srivastava, and K. M. Feigh, "The role of shared mental models in human-ai teams: a theoretical review," *Theoretical Issues in Ergonomics Science*, vol. 24, no. 2, pp. 129–175, 2023.
- [25] I. Ajzen, "The theory of planned behavior," *Organizational Behavior and Human Decision Processes*, vol. 50, no. 2, pp. 179–211, 1991.
- [26] S. H. Schwartz and J. A. Howard, "A normative decision-making model of altruism," in *Altruism and Helping Behavior: Social, Personality, and Developmental Perspectives*, J. P. Rushton and R. M. Sorrentino, Eds. Lawrence Erlbaum, 1981, pp. 189–211.
- [27] F. Brambati, D. Ruscio, F. Biassoni, R. Hueting, and A. Tedeschi, "Predicting acceptance and adoption of renewable energy community solutions: the prosumer psychology," *Open Research Europe*, vol. 2, no. 115, pp. –, 2023.
- [28] A. Silva, M. Schrum, E. Hedlund-Botti, N. Gopalan, and M. Gombolay, "Explainable artificial intelligence: Evaluating the objective and subjective impacts of xai on human-agent interaction," *International Journal of Human-Computer Interaction*, vol. 39, no. 7, pp. 1390–1404, 2023.

