

The Role of Safety Management Systems (SMS) in the Implementation of Human–Artificial Intelligence (AI) Teaming in the Aviation Ecosystem

Dimitrios Ziakkas¹, Eleftherios Bokas², and Debra Henneberry³

¹Coventry University, Faculty of Engineering, Environment and Computing, Coventry, CV1 5FB, U.K.

²Sky Express, Athens International Airport “El. Venizelos” - Building 23,190 19, Spata, Athens, Greece

³Purdue University, School of Aviation and Transportation Technology, West Lafayette, IN 47907, USA

ABSTRACT

The integration of Human–Artificial Intelligence (AI) teaming in aviation represents a transformative shift in safety-critical operations, requiring robust organisational frameworks to manage emerging risks and sustain human performance. As AI becomes embedded across flight operations, maintenance, training, and safety analytics, Safety Management Systems (SMS) play a central role in ensuring that human–AI collaboration is safely introduced and governed. This paper examines how SMS principles must evolve to address the socio-technical challenges posed by AI, including system opacity, non-determinism, and continuous learning. These characteristics challenge traditional approaches to hazard identification, risk assessment, and safety assurance, requiring more dynamic and data-driven monitoring of human–AI interactions. The study analyses how the four SMS components—Safety Policy, Risk Management, Assurance, and Promotion—can support human–AI teaming. It highlights the need for human-centric policies, new risk identification methods, continuous monitoring of AI performance, and training that fosters both trust and critical evaluation of AI systems. The paper concludes by proposing an enhanced SMS framework incorporating explainable AI, resilience engineering principles, AI-specific safety indicators, and adaptive training aligned with CBTA/EBT methodologies.

Keywords: Safety Management Systems (SMS), Human–AI teaming, Aviation safety, Adaptive automation, Risk management, Safety assurance, Safety culture

INTRODUCTION

Artificial intelligence (AI) is increasingly embedded within the aviation ecosystem, supporting functions such as flight deck decision assistance, predictive maintenance, training analytics, and safety performance monitoring. Unlike legacy automation, contemporary AI systems often operate using probabilistic models, adaptive algorithms, and machine learning techniques that evolve over time. These characteristics fundamentally alter

the nature of human–machine interaction, shifting operational paradigms from human supervision of deterministic automation toward Human–AI teaming (Parasuraman et al., 2000).

Safety Management Systems (SMS) constitute the internationally accepted framework for managing aviation safety. Mandated by the International Civil Aviation Organization (ICAO) through Annex 19 and implemented by authorities such as the Federal Aviation Administration (FAA) and the European Union Aviation Safety Agency (EASA), SMS provides a structured approach to hazard identification, risk management, safety assurance, and organisational learning (ICAO, 2018; FAA, 2023). However, SMS frameworks were largely developed under assumptions of predictable system behaviour, stable functional boundaries, and clearly defined human responsibility. The introduction of Human–AI teaming challenges these assumptions by introducing opacity, dynamic behaviour, and evolving trust relationships.

This paper examines the role of SMS in supporting the safe implementation of Human–AI teaming across the aviation ecosystem. It argues that while the foundational principles of SMS remain valid, their application must evolve to address AI-specific risks, socio-technical complexity, and emergent system behaviour.

CONCEPTUAL FOUNDATIONS OF HUMAN–AI TEAMING AND THE ROLE OF SMS

Human–AI teaming represents a socio-technical transformation rather than a purely technological enhancement. AI systems increasingly participate in perception, diagnosis, prioritisation, and recommendation functions traditionally associated with human cognition. As a result, safety risks increasingly arise from interactional dynamics rather than discrete component failures (Reason, 1997).

Key characteristics of AI-enabled systems include algorithmic opacity, adaptive learning, and probabilistic outputs. These features complicate hazard identification and risk assessment, as system behaviour may vary across contexts and over time (Hollnagel, 2014). Additionally, human cognitive responses to AI—such as automation bias, over-trust, and skill degradation—can introduce latent conditions that erode safety margins (Parasuraman et al., 2000).

From an SMS perspective, these characteristics necessitate a shift from compliance-focused safety management toward performance-based and resilience-oriented approaches capable of addressing uncertainty and emergence (Hollnagel, 2014). Table 1 summarises key Human–AI teaming characteristics and their implications for SMS.

Table 1: Research methodology overview - alignment of analysis, design, development, implementation, evaluation (addie) phases with human–ai teaming competencies.

Human–AI Characteristic	Safety Implication	SMS Impact
Algorithmic opacity	Reduced transparency	Enhanced assurance and explainability
Adaptive learning	Performance drift	Continuous monitoring
Probabilistic outputs	Increased cognitive load	Trust calibration training
High automation	Skill degradation	Competency management

METHODOLOGY

This study adopts a qualitative, theory-driven analytical methodology grounded in socio-technical systems theory and established aviation safety management principles. Given the conceptual and regulatory nature of Human–AI teaming in contemporary aviation, an interpretive research design was selected to examine how existing SMS frameworks can accommodate and govern emerging AI-enabled operational practices.

The methodology consists of three complementary components. First, a structured literature review was conducted across aviation safety, human factors, automation, and artificial intelligence domains. Peer-reviewed journal articles, authoritative books, and regulatory publications were analysed to identify recurring themes related to human–automation interaction, trust calibration, resilience engineering, and organisational accident theory. Foundational works by Reason (1997), Parasuraman et al. (2000), and Hollnagel (2014) were used as theoretical anchors to interpret the safety implications of adaptive and intelligent systems.

Second, a regulatory and standards analysis was performed focusing on ICAO, FAA, and EASA safety management documentation. Key sources included ICAO Annex 19, the FAA Safety Management System Manual, and the EASA Artificial Intelligence Roadmap. These documents were examined to identify explicit and implicit assumptions regarding system predictability, human accountability, performance-based oversight, and continuous assurance. The analysis assessed the extent to which current SMS requirements can be mapped onto Human–AI teaming characteristics such as opacity, adaptive learning, and probabilistic behaviour.

Third, an SMS component-based analytical framework was applied. The four core SMS elements—Safety Policy, Safety Risk Management, Safety Assurance, and Safety Promotion—were used as deductive categories to structure the analysis. For each component, AI-related hazards, governance challenges, and potential adaptations were identified and synthesised. Industry practices reported by major operators, including Turkish Airlines, the Lufthansa Group, and Sky Express were reviewed as illustrative examples of SMS-aligned AI implementation, without claiming empirical generalisability.

Data synthesis followed an iterative thematic approach, integrating insights from literature, regulatory texts, and industry practices. The objective was not to validate a specific AI system or operational concept, but to evaluate SMS as a governance mechanism capable of managing emergent human–AI

interaction risks. This methodological approach aligns with prior safety research addressing complex, non-deterministic systems where experimental or quantitative methods are not yet feasible.

Table 2: Research methodology overview.

Methodological Component	Data Sources	Analytical Focus	Output
Literature review	Peer-reviewed journals, books	Human–AI interaction, automation, safety theory	Theoretical grounding
Regulatory analysis	ICAO, FAA, EASA documents	SMS assumptions and governance	Regulatory alignment
SMS framework analysis	SMS components (Policy, SRM, Assurance, Promotion)	AI-related risks and adaptations	Structured synthesis
Industry illustration	Public operator practices	Practical applicability	Contextual validation

FINDINGS

Safety Policy establishes organisational commitment, accountability, and governance. In the context of Human–AI teaming, Safety Policy must explicitly articulate a human-centric philosophy that preserves human authority over safety-critical decisions (ICAO, 2018). Policies should define acceptable AI roles, clarify responsibility boundaries, and address ethical considerations related to algorithmic decision support.

Furthermore, Safety Policy should address AI lifecycle governance, including data management, model updates, validation constraints, and operational limitations. Clear policy articulation reduces the risk of implicit authority transfer to AI systems and supports regulatory accountability under existing SMS requirements (FAA, 2023; EASA, 2023).

Safety Risk Management

Safety Risk Management (SRM) must expand beyond traditional hazard identification methods focused on equipment failure or procedural non-compliance. AI-related hazards often emerge from human–AI interaction, including mismatches between AI intent and operator mental models, degraded vigilance due to perceived reliability, and loss of redundancy in highly automated environments (Reason, 1997).

Effective SRM requires multidisciplinary hazard identification incorporating operational expertise, human factors analysis, and data science. Risk assessments must explicitly consider degraded data quality, algorithmic bias, and uncertainty in AI confidence levels. Table 3 illustrates representative AI-related hazards and corresponding SMS risk controls.

Table 3: AI-Related hazards and sms risk controls.

Hazard Category	Example	SMS Mitigation
Automation bias	Over-reliance on AI advice	Training, SOP reinforcement
Algorithmic drift	Gradual loss of accuracy	Continuous validation
Data bias	Skewed operational inputs	Data governance controls
Role ambiguity	Unclear decision authority	Policy clarification

Safety Assurance

Safety Assurance ensures that risk controls remain effective and that safety performance meets organisational expectations. In Human–AI teaming contexts, assurance activities must incorporate continuous monitoring of AI system performance and human–AI interaction patterns (Hollnagel, 2014). This includes tracking AI outputs, operator responses, and deviations from expected behaviour.

Explainability audits and periodic validation against operational assumptions are essential to maintaining confidence in AI-supported systems (EASA, 2023). Safety Assurance must also monitor shifts in trust and reliance, recognising that gradual changes in human behaviour may signal emerging safety risks.

Safety Promotion

Safety Promotion supports organisational readiness through training, communication, and reporting. Effective Human–AI teaming requires AI literacy, awareness of system limitations, and the ability to critically evaluate AI-generated outputs. Training programmes should integrate scenario-based exercises addressing ambiguous or conflicting AI recommendations, aligned with Competency-Based Training and Assessment (CBTA) and Evidence-Based Training (EBT) frameworks.

A robust reporting culture is essential for capturing AI-related anomalies and near-misses. Encouraging reporting of unexpected AI behaviour supports organisational learning and continuous improvement within SMS (ICAO, 2018).

Industry and Regulatory Perspectives

Industry operators such as Turkish Airlines, Lufthansa Group and Skyexpress have planned and begun integrating AI-enabled analytics within SMS-driven oversight structures, reinforcing SMS as the primary governance mechanism for intelligent systems. These initiatives demonstrate the feasibility of embedding AI within existing safety frameworks while maintaining regulatory compliance.

Regulatory authorities have similarly acknowledged the implications of AI for safety oversight. The FAA emphasises data-driven assurance and continuous monitoring within SMS, while EASA highlights human–machine

interaction and explainability as critical considerations for AI deployment (FAA, 2023; EASA, 2023). ICAO continues to promote performance-based safety management as a foundation for managing emerging technologies (ICAO, 2018). Table 4 summarises regulatory perspectives relevant to Human–AI teaming.

Table 4: Regulatory perspectives on AI and SMS.

Authority	Key Focus	SMS Relevance
ICAO	Performance-based safety	SMS adaptability
FAA	AI oversight, data analytics	Continuous assurance
EASA	Human–AI interaction	Human-centric governance

DISCUSSION

The findings underscore the need for a systemic rethinking of aviation training. Human–AI teaming cannot be addressed through incremental updates; it requires a transformation of instructional design, assessment philosophy, and organisational learning mechanisms.

The analysis presented in this paper indicates that existing SMS frameworks provide a robust foundation for governing Human–AI teaming, but require targeted enhancement to remain effective in environments characterised by adaptive, opaque, and probabilistic technologies. An enhanced SMS framework must preserve the core principles of hazard identification, risk control, and continuous improvement, while explicitly addressing the socio-technical dynamics introduced by intelligent systems. A central requirement of this enhanced framework is the systematic integration of explainable artificial intelligence (XAI) within SRM and Safety Assurance processes. Explainability enables human operators, safety managers, and regulators to understand the basis of AI-generated recommendations and predictions, supporting informed decision-making and effective oversight (EASA, 2023). Within SRM, explainability facilitates more accurate hazard identification by revealing underlying assumptions, data dependencies, and confidence levels associated with AI outputs. Within Safety Assurance, explainability audits provide a mechanism for verifying that AI behaviour remains aligned with operational expectations and safety objectives over time.

Resilience engineering principles further strengthen SMS capability by addressing the inherent uncertainty and variability of Human–AI teaming. Traditional safety management has focused primarily on preventing predefined failures; however, AI-enabled systems may fail in novel and unanticipated ways (Hollnagel, 2014). By adopting a resilience-oriented perspective, SMS shifts emphasis toward monitoring system performance under varying conditions, supporting adaptive responses, and maintaining functional safety margins even when AI behaviour deviates from nominal patterns. This perspective is particularly relevant in high-automation or reduced-crew operations, where recovery capacity may be constrained. The development of AI-specific safety performance indicators represents another critical enhancement. Conventional safety indicators are often retrospective and

event-based, whereas Human–AI interaction risks may emerge gradually through changes in trust, reliance, or cognitive workload. Leading indicators tailored to AI operations—such as trends in AI override rates, frequency of conflicting human–AI decisions, or anomalies in confidence signalling—can provide early warning of degraded system performance. Integrating such indicators into Safety Assurance supports proactive intervention before adverse outcomes occur, consistent with performance-based oversight principles promoted by ICAO and national regulators (ICAO, 2018; FAA, 2023).

Competency development constitutes a further pillar of the enhanced SMS framework. Human–AI teaming alters the competencies required of pilots, engineers, instructors, and safety professionals. Technical proficiency must be complemented by skills in monitoring intelligent systems, interpreting probabilistic information, and maintaining appropriate trust calibration (Parasuraman et al., 2000). Aligning training with CBTA and EBT frameworks ensures that competency development remains adaptive and data-informed. SMS plays a coordinating role by linking training outcomes to observed safety performance and operational data. Organisational culture remains a decisive factor in the effectiveness of any enhanced SMS framework. The introduction of AI into safety-critical operations can create reluctance to challenge automated systems or report anomalies perceived as technical rather than operational. A mature safety culture must explicitly recognise Human–AI interaction as a legitimate source of safety concern and encourage transparent reporting of unexpected AI behaviour. Safety Promotion activities should therefore emphasise shared responsibility between humans and AI systems, reinforcing that accountability ultimately resides with the organisation and its personnel (Reason, 1997).

Finally, regulatory alignment is essential to the viability of an enhanced SMS framework. While certification and approval processes for AI systems continue to evolve, SMS provides an immediately applicable mechanism for demonstrating control of AI-related risks within existing regulatory structures. By embedding AI governance within SMS, organisations can support regulatory confidence, facilitate oversight, and enable incremental adoption of intelligent technologies without compromising safety margins. In this sense, SMS functions as both a safety management and regulatory bridging mechanism during a period of rapid technological change.

In summary, an enhanced SMS framework for Human–AI teaming integrates explainability, resilience engineering, AI-specific safety indicators, adaptive competency development, and cultural reinforcement within the established SMS structure. A further strengthening element within this enhanced framework is the integration of the FAIR³ System (Functionality, Authority, Information, Responsibility, and Resilience), which provides a structured lens for governing Human–AI teaming in safety-critical environments. The FAIR³ System extends traditional human factors and safety governance concepts by explicitly addressing how authority, responsibility, and information flows are allocated and maintained when intelligent systems participate in operational decision-making. Within an SMS context, FAIR³

supports systematic alignment between technological capability, human accountability, and organisational oversight.

From a Safety Policy perspective, the FAIR³ System reinforces the principle that authority and responsibility must remain clearly assigned to human operators and accountable management, regardless of AI system sophistication. AI functionality is formally bounded by policy-defined operational limits, ensuring that decision authority is not implicitly transferred to automated agents. This aligns with ICAO's emphasis on clear accountability structures within SMS and mitigates ethical and legal ambiguity associated with autonomous or semi-autonomous systems (ICAO, 2018). Within Safety Risk Management, FAIR³ provides a structured approach to identifying interactional hazards arising from misaligned functionality, inadequate information transparency, or ambiguous responsibility allocation. By explicitly examining whether AI systems provide sufficient, accurate, and context-relevant information to support human decision-making, organisations can better assess risks related to automation bias, over-reliance, and mode confusion. The inclusion of resilience as a core FAIR³ element further supports the identification of recovery pathways when AI behaviour deviates from expected performance, consistent with resilience engineering principles (Hollnagel, 2014).

Safety Assurance processes benefit from FAIR³ by enabling targeted monitoring of human–AI interaction quality rather than solely system outputs. Assurance activities can evaluate whether AI functionality remains within defined authority boundaries, whether information presentation supports effective human oversight, and whether responsibility attribution remains clear in operational practice. These assessments complement AI-specific safety indicators by focusing on governance integrity in addition to technical performance.

Finally, within Safety Promotion, the FAIR³ System supports training and cultural initiatives that reinforce shared understanding of roles, limitations, and accountability in Human–AI teaming. Training aligned with CBTA and EBT can incorporate FAIR³ principles to strengthen operators' ability to manage authority gradients, interrogate AI outputs, and maintain adaptive resilience under uncertainty. Embedding FAIR³ concepts within Safety Promotion also reinforces reporting behaviours by legitimising Human–AI interaction issues as safety-relevant events.

CONCLUSION

The rapid integration of artificial intelligence across the aviation ecosystem marks a decisive shift in how safety-critical operations are designed, managed, and governed. Human–AI teaming introduces forms of variability, opacity, and adaptive behaviour that extend beyond the assumptions underpinning traditional automation and challenge long-standing safety management practices. This paper has demonstrated that, while these challenges are significant, SMS remains the most appropriate and resilient governance mechanism for managing the risks associated with Human–AI collaboration—provided they evolve in a targeted and human-centric manner. Through

structured analysis of the four SMS components, the study has shown that Human–AI teaming does not invalidate existing safety management principles but instead exposes the need for their recalibration. Safety Policy must explicitly preserve human authority and accountability in environments increasingly influenced by algorithmic decision support.

In conclusion, Human–AI teaming represents not a departure from established aviation safety principles but an evolutionary challenge that tests their adaptability. By evolving SMS to incorporate explainability, resilience, competency adaptation, and structured governance frameworks such as FAIR³, the aviation industry can ensure that Human–AI collaboration enhances—rather than erodes—system safety. The continued relevance of SMS in the age of intelligent systems ultimately reaffirms a central tenet of aviation safety: technology may change, but effective safety governance remains fundamentally human-centred.

ACKNOWLEDGMENT

The authors thank Coventry University, Purdue University faculty members, Sky Express and HF Horizons analysts for their invaluable feedback, which contributed to this work.

REFERENCES

- European Union Aviation Safety Agency (EASA). (2023). *eMCO-SIPO - Minimum crew operations-single pilot operations*. European Union Aviation Safety Agency.
- Federal Aviation Administration (FAA). (2023). *Single pilot operations research roadmap*. Federal Aviation Administration.
- Hollnagel, E. (2014). *Safety-I and Safety-II: The past and future of safety management*. Ashgate.
- International Civil Aviation Organization (ICAO). (2018). *Safety management manual (SMM)* (Doc 9859, AN/474, 4th ed.). International Civil Aviation Organization.
- Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000, May). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 30(3), 286-297. <https://doi.org/10.1109/3468.844354>.
- Reason, J. (1997). *Managing the risks of organizational accidents*. Ashgate.