Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 2481–2490
https://doi.org/10.54941/ahfe1008473

Human-Centred Transition Pathways for AI-Enabled Safety-Critical Systems: An Industrial Systems Engineering Perspective

Dimitrios Ziakkas[1] and Debra Henneberry[2]

[1]School of Future Transport Engineering, Coventry University, Coventry, CV1 5FB, UK
[2]School of Aviation and Transportation Technology, Purdue University, West Lafayette, IN 47907, USA

ABSTRACT

Artificial intelligence is transforming safety-critical industries from technology-centred automation towards integrated human-AI work systems. However, many organisations continue to implement AI through isolated technical projects rather than coordinated socio-technical transformation. This paper proposes a Human-Centred Transition Pathway that combines systems engineering, Human Factors Integration, Safety-II, STAMP and Human-AI Teaming into a structured implementation framework. The proposed methodology aligns operational objectives, competency requirements, organisational processes and AI capabilities throughout the system lifecycle. Aviation provides the principal case study through recent developments in EASA AI 3.0 and competency-based operations, while lessons are transferable to rail, maritime and healthcare systems. The framework demonstrates how industrial organisations can move beyond regulatory compliance towards resilient, human-centred AI ecosystems that support sustainable operational perf ormance.

Keywords: Human-centred transition pathway, Human factors integration, Human-AI teaming, STAMP, Safety-II, Systems engineering, EASA artificial intelligence, Competency-based training

INTRODUCTION

Artificial intelligence (AI) is migrating from back-office analytics into the operational core of safety-critical industries, where it assists, advises and increasingly acts alongside professionals who remain accountable for safe outcomes. The European Union has classified many such applications as high-risk under the Artificial Intelligence Act (European Parliament and Council, 2 024), and the European Union Aviation Safety Agency (EASA) has translated this horizontal regulation into sector-specific guidance through its AI Roadmap 2.0 (EASA, 2023), successive AI Concept Papers (EASA, 2024; 2026) and its first rulemaking proposal on AI trustworthiness, Notice of Proposed Amendment (NPA) 2025-07 (EASA, 2025).

Organisational practice lags behind this shift. Many organisations still introduce AI through isolated technical projects, each verified against its specification and handed to operations without a corresponding redesign

Received September 25, 2026; Revised October 31, 2026; Accepted November 6, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

of procedures, competencies, authority relationships or safety oversight. The ironies of automation (Bainbridge, 1983), the misuse and disuse of automated aids (Parasuraman and Riley, 1997) and the difficulty of making automation a team player (Klein et al., 2004) re-emerge, amplified by the opacity and adaptivity of machine learning (National Academies of Sciences, Engineering, and Medicine, 2022). Our earlier work on human performance and accident investigation in transportation indicates that the most consequential failures of AI integration arise not in algorithms but at the interfaces between technology, people and organisations (Ziakkas and Plioutsias, 2024; Ziakkas and Henneberry, 2025).

What is missing is a structured transition pathway: a lifecycle-anchored route from the current work system to a human-AI work system that is demonstrably safe, competent and resilient. Systems engineering provides the lifecycle discipline (INCOSE, 2023), Human Factors Integration (HFI) the human-centred content (Booher, 2003), Safety-II and the System-Theoretic Accident Model and Processes (STAMP) complementary lenses on how safety is created and lost (Hollnagel, 2014; Leveson, 2011), and Human-AI Teaming (HAT) the relational qualities of the joint system (O'Neill et al., 2022). These bodies of knowledge have largely developed in parallel. This paper therefore proposes a Human-Centred Transition Pathway (HCTP) guided by three research questions (RQs): which socio-technical dimensions must be coordinated for safe and sustainable AI implementation (RQ1); how these dimensions can be embedded in systems engineering lifecycle processes (RQ2); and whether a pathway developed in aviation is transferable to other safety-critical industries (RQ3).

THEORETICAL FOUNDATIONS

Systems Engineering and Human Factors Integration

Systems engineering treats people, processes and technology as elements of one system delivering a capability across a lifecycle, formalised in ISO/IEC/IEEE 15288 through processes such as stakeholder needs definition, architecture definition, integration, verification, validation and operation (INCOSE, 2023; ISO/IEC/IEEE, 2023). HFI requires that manpower, personnel, training, human factors engineering and safety are treated as design variables rather than downstream constraints (Booher, 2003), while ISO 9241-210 establishes that human-centred design is iterative and evaluated with users throughout development (ISO, 2019). Sociotechnical theory adds that technical and social subsystems must be jointly optimised; designing one while treating the other as fixed produces systems that are technically correct but operationally fragile (Clegg, 2000).

Safety-II and STAMP as Complementary Safety Lenses

Safety-II reframes safety from the absence of adverse outcomes to the capacity to succeed under varying conditions, directing attention to the everyday adjustments through which practitioners complete work that procedures cannot fully specify (Hollnagel, 2014). This matters for AI because models

are trained on work-as-imagined or work-as-recorded, whereas safety is produced in work-as-done; an AI function optimised against the former may silently erode the adaptive capacity that sustains the latter. STAMP conceptualises safety as an emergent property of hierarchical control, in which accidents result from inadequate enforcement of safety constraints rather than chains of component failures (Leveson, 2011). Its hazard analysis technique, System-Theoretic Process Analysis (STPA), identifies unsafe control actions that can arise even when every component performs as specified (Leveson and Thomas, 2018), which suits AI, since a machine learning function is effectively a new controller or feedback channel inserted into an existing control structure.

Human-AI Teaming and the Regulatory Trajectory

HAT research moves beyond levels-of-automation taxonomies (Parasuraman, Sheridan and Wickens, 2000) to examine how humans and AI agents coordinate, share situation awareness and recover from each other's errors (O'Neill et al., 2022). Appropriate reliance, in which trust matches actual capability, remains the central design objective and depends on transparency, predictability and directability rather than on accuracy alone (Lee and See, 2004; Shneiderman, 2020). EASA operationalises these ideas through AI levels: Level 1 denotes human augmentation and cognitive assistance, Level 2 human-AI cooperation and collaboration, and Level 3 advanced automation in which the end user may be remote or absent (EASA, 2024). Proposed Issue 03 of the Concept Paper, released for consultation in June 2026 and referred to in this paper as EASA AI 3.0, completes this scope by addressing Level 3 and additional techniques such as reinforcement learning (EASA, 2026), while NPA 2025-07 opens Rulemaking Task RMT.0742 on AI assurance, human factors and ethics (EASA, 2025). The competency-based training and assessment (CBTA) and evidence-based training (EBT) frameworks of the International Civil Aviation Organization (ICAO) define the observable behaviours required of the human team member (ICAO, 2013; 2020). The regulatory architecture thus contains the elements of a human-centred approach, but not guidance on how to sequence and integrate them.

METHODOLOGY

The study adopted a pragmatist philosophy, judging knowledge by its capacity to resolve practical problems (Saunders, Lewis and Thornhill, 2019), and a design science research (DSR) strategy, because the intended contribution is an artefact that must be both theoretically grounded and useful to practitioners (Hevner et al., 2004). The six DSR activities of Peffers et al. (2007) structured the work, and an abductive approach moved iteratively between theory and operational problems. Evidence was drawn from an integrative literature review (Torraco, 2005), a documentary analysis of regulatory and standards texts (EU AI Act, EASA Roadmap, Concept Papers and NPA 2025-07, ICAO training manuals and ISO standards), and an embedded single-case study of European commercial aviation, selected as a

critical case combining mature safety management, a published AI regulatory trajectory and an established CBTA ecosystem (Yin, 2018). Constructs were cross-mapped against ISO/IEC/IEEE 15288 processes and four alignment dimensions derived from the problem definition: operational objectives, competency requirements, organisational processes and AI capabilities. The authors' operational aviation backgrounds informed judgements of plausibility, while the risk of privileging flight-deck interpretations was mitigated through engagement with non-aviation literatures and researcher triangulation. Table 1 summarises the design.

Table 1: Research method, structured by the layers of the research onion (Saunders, Lewis and Thornhill, 2019).

Layer	Design Decision	Justification
Philosophy	Pragmatism	Knowledge judged by practical utility for industrial AI implementation
Approach to theory	Abduction	Iteration between theory and operational problems generates a new framework
Methodological choice	Multi-method qualitative	Review, documentary analysis and case study provide complementary evidence
Strategy	Design science research (Hevner et al., 2004; Peffers et al., 2007)	The contribution is a usable, theoretically grounded artefact
Time horizon	Cross-sectional; regulatory state to June 2026	Captures NPA 2025-07 and the Concept Paper Issue 03 consultation
Data sources	Peer-reviewed literature; EU, EASA, ICAO, IMO and ISO documents; aviation critical case	Theoretical, regulatory and operational triangulation
Analysis	Cross-mapping to ISO/IEC/IEEE 15288 processes and four alignment dimensions	Links human-centred constructs to auditable lifecycle processes
Evaluation	Demonstration in aviation; analytical transferability to rail, maritime and healthcare	Establishes utility and scope of generalisation
Quality criteria	Theoretical, data and researcher triangulation; reflexive positionality	Enhances credibility and confirmability

FINDINGS: THE HUMAN-CENTRED TRANSITION PATHWAY

The synthesis produced a five-phase pathway (Figure 1; Table 2). Each phase maps to ISO/IEC/IEEE 15288 processes, integrates specific theoretical contributions and produces evidence that aligns the four dimensions. The phases form a closed loop, because operational learning in the final phase continuously revises the operational intent that initiates the first.

Phases of the Pathway and the Aviation Case

Phase 1, operational intent and work-as-done baseline, begins not with the algorithm but with the operational problem and the human work that currently solves it. Stakeholder needs definition captures work-as-done through observation and practitioner narratives (Hollnagel, 2014), and concludes with an explicit declaration of the intended AI level, so that the human role is fixed as a requirement rather than discovered after deployment. In the aviation case, an operator considering AI support for in-flight diversion planning would first document how crews integrate weather, fuel, technical and commercial information, then declare the function as Level 1 cognitive assistance with crew authority retained (EASA, 2024).

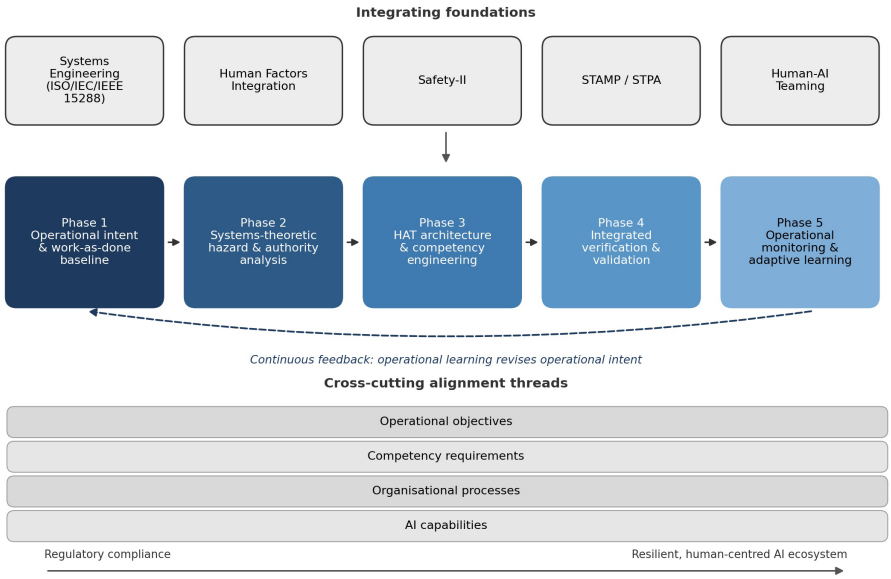


Figure 1: The Human-Centred Transition Pathway (HCTP) for AI-enabled safety-critical systems.

Phase 2, systems-theoretic hazard and authority analysis, inserts the AI function into a STAMP control structure spanning flight crew, operations control, maintenance, training and the regulator. STPA treats the AI both as a controller issuing recommendations and as a feedback channel shaping human process models (Leveson and Thomas, 2018). Unsafe control actions arising from human-AI interaction, such as automation bias, confusion over which agent holds authority, and erosion of the crew's independent process model (Bainbridge, 1983), generate derived requirements spanning technology, procedures and training.

Phase 3, HAT architecture and competency engineering, co-designs the AI function, the interface and the competency framework. Interaction design addresses transparency, uncertainty communication and directability, consistent with EASA human factors objectives (EASA, 2024; 2025). Training needs analysis maps new behaviours to ICAO competencies such

as situation awareness and information management, problem solving and decision-making, and workload management (ICAO, 2020). AI-related behaviours, including interrogating a recommendation, detecting degraded model performance and reverting to independent reasoning, are specified observably so that they can be trained and assessed within CBTA and EBT (ICAO, 2013).

Phase 4, integrated verification and operational validation, recognises that model verification against learning assurance objectives is necessary but insufficient. Human-in-the-loop validation uses EBT-style simulator scenarios containing model edge cases and degraded data, with criteria covering team performance, trust calibration (Lee and See, 2004), workload and recovery when the AI is wrong or unavailable. Procedures, safety management system (SMS) interfaces and data governance are validated in the same activity, so that acceptance addresses the whole work system.

Phase 5, operational monitoring and adaptive learning, treats the human-AI work system as continuously evolving. Conventional safety performance indicators are combined with Safety-II indicators of how practitioners adapt to, work around or over-rely on the AI. Model drift and practice drift are monitored together, because their interaction can produce boundary migration even when each appears acceptable in isolation (Rasmussen, 1997; Dekker, 2011). AI-supported occurrence analysis can accelerate learning, provided investigative judgement remains human-led (Ziakkas and Henneberry, 2025; Ziakkas, Henneberry and Plioutsias, 2025). Findings return to Phase 1, closing the loop.

Table 2: Findings: phases of the human-centred transition pathway with integrating foundations, alignment outputs and aviation illustration.

Phase (ISO/IEC/IEEE 15288 Processes)	Integrating Foundations	Outputs Across the Four Alignment Threads	Aviation Illustration
1. Operational intent and work-as-done baseline (mission analysis; stakeholder needs)	Safety-II; human-centred design	Operational need statement; work-as-done description; declared AI level and human role	Diversion-planning support declared Level 1; crew authority retained
2. Systems-theoretic hazard and authority analysis (system requirements; risk management)	STAMP/STPA; HAT	Human-AI control structure; unsafe control actions; authority allocation; derived requirements	Unsafe action: crew accepts a diversion recommendation built on stale weather data
3. HAT architecture and competency engineering (architecture and design definition)	HFI; HAT; CBTA	Transparency and uncertainty cues; AI-related observable behaviours; instructor guidance	Interrogation and reversion behaviours mapped to ICAO competencies within EBT

(Continued)

Table 2: Continued.

Phase (ISO/IEC/IEEE 15288 Processes)	Integrating Foundations	Outputs Across the Four Alignment Threads	Aviation Illustration
4. Integrated verification and validation (integration; verification; validation; transition)	Learning assurance; simulation; EBT	Assurance evidence; team performance and trust-calibration results; organisational readiness	Simulator scenarios with degraded AI output; SMS and procedure changes accepted jointly
5. Operational monitoring and adaptive learning (operation; maintenance)	Safety-II; drift models; AI-supported investigation	Paired model-drift and practice-drift indicators; lessons returned to Phase 1	Flight data monitoring and crew reports reviewed for over-reliance and workarounds

Cross-Cutting Alignment Threads

Four alignment threads maintain coherence between dimensions typically owned by different organisational functions. Operational objectives ensure that AI is pulled by operational need rather than pushed by technical availability; competency requirements ensure that each change in function allocation produces a matching change in trained and assessed behaviour; organisational processes ensure that procedures, the SMS and change management evolve with the technology; and AI capabilities ensure that model performance and limitations are communicated in terms the other threads can use. The central finding is that transition risk concentrates where these threads fall out of alignment, for example when an AI capability is fielded before the corresponding competency is defined.

Transferability to Other Safety-Critical Industries

The pathway's structure appears transferable, while its content must be localised. Rail shares aviation's procedural, signalling-based authority structures and already treats the railway as an integrated socio-technical system (Wilson et al., 2005). Maritime autonomy is moving from trials towards regulation, and the International Maritime Organization (IMO) interim guidelines already require risk assessment, qualified personnel and attention to the human element (IMO, 2019). Healthcare is least standardised, yet the Systems Engineering Initiative for Patient Safety (SEIPS) work system model (Carayon et al., 2006) and analyses of human factors challenges in clinical AI (Sujan et al., 2019) map directly onto the pathway's emphasis on work-as-done, competency and organisational readiness. The principal adaptation lies in Phase 3, because competency frameworks outside aviation are less mature and may need to be developed alongside the AI capability.

DISCUSSION

The principal contribution of this paper is a pathway that places the transition, rather than the technology, at the centre of the engineering effort. For RQ1, four dimensions, namely operational objectives, competency requirements, organisational processes and AI capabilities, were identified whose misalignment constitutes the dominant source of transition risk. For RQ2, these dimensions were embedded within existing ISO/IEC/IEEE 15288 processes without creating a parallel governance structure. For RQ3, the architecture proved analytically transferable to rail, maritime and healthcare, with competency engineering requiring the greatest adaptation.

Theoretically, the pathway integrates traditions often presented as competing. STAMP defines the constraints the human-AI system must enforce, Safety-II describes the adaptive performance through which practitioners enforce them when the world departs from the model, and HAT specifies the transparency, directability and calibrated trust that make such adaptation possible (Klein et al., 2004; O'Neill et al., 2022). This extends calls for human-centred AI (Shneiderman, 2020) by locating human-centredness in auditable lifecycle processes and evidence. Practically, the pathway supports a movement from compliance towards resilience, because NPA 2025-07 defines what must be demonstrated for an AI system but not how an organisation should reshape itself around it. Industrial systems engineers should therefore involve training, safety and operations specialists from the first lifecycle stage, budget for human-in-the-loop validation and post-deployment monitoring, and treat competency frameworks as configuration items under the same change control as software.

Several limitations apply. The pathway has been demonstrated analytically but not yet evaluated within an operating organisation; the aviation case reflects a regulatory environment still in flux, as Concept Paper Issue 03 is under consultation and the second step of RMT.0742 is in preparation; and the authors' backgrounds may have privileged flight-deck perspectives over maintenance, air traffic management and ground operations. Future research should evaluate the pathway through a cross-industry Delphi study, apply it prospectively within an operator's AI implementation, and test the Phase 4 criteria in controlled simulation studies.

CONCLUSION

AI will not make safety-critical systems safer simply by being accurate; it will do so only if organisations redesign the work systems into which it is introduced, so that people, procedures, competencies and technology evolve together. By combining systems engineering, Human Factors Integration, Safety-II, STAMP and Human-AI Teaming, and by aligning operational objectives, competency requirements, organisational processes and AI capabilities from concept to operation, the Human-Centred Transition Pathway offers industrial organisations a route beyond regulatory compliance towards resilient, human-centred AI ecosystems. Aviation provides the first proving ground; rail, maritime and healthcare can benefit from the same attention to the people who remain at the centre of safe operation.

ACKNOWLEDGMENT

The authors thank colleagues at Coventry University and Purdue University for discussions that informed this framework.

REFERENCES

Bainbridge, L. (1983). Ironies of automation. Automatica, 19(6), pp. 775–779.

Booher, H.R. (ed.) (2003). Handbook of Human Systems Integration. Hoboken, NJ: Wiley.

Carayon, P., Schoofs Hundt, A., Karsh, B.-T., Gurses, A.P., Alvarado, C.J., Smith, M. and Flatley Brennan, P. (2006). Work system design for patient safety: The SEIPS model. Quality and Safety in Health Care, 15(Suppl. 1), pp. i50–i58.

Clegg, C.W. (2000). Sociotechnical principles for system design. Applied Ergonomics, 31(5), pp. 463–477.

Dekker, S. (2011). Drift into Failure: From Hunting Broken Components to Understanding Complex Systems. Farnham: Ashgate.

EASA (2023). EASA Artificial Intelligence Roadmap 2.0: A Human-Centric Approach to AI in Aviation. Cologne: European Union Aviation Safety Agency.

EASA (2024). EASA Concept Paper: Guidance for Level 1 & 2 Machine Learning Applications, Issue 02. Cologne: European Union Aviation Safety Agency.

EASA (2025). Notice of Proposed Amendment 2025-07: Artificial Intelligence Trustworthiness (RMT.0742). Cologne: European Union Aviation Safety Agency.

EASA (2026). EASA Artificial Intelligence Concept Paper, Proposed Issue 03. Cologne: European Union Aviation Safety Agency.

European Parliament and Council (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L series, 12 July 2024.

Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), pp. 75–105.

Hollnagel, E. (2014). Safety-I and Safety-II: The Past and Future of Safety Management. Farnham: Ashgate.

ICAO (2013). Manual of Evidence-based Training (Doc 9995). Montréal: International Civil Aviation Organization.

ICAO (2020). Procedures for Air Navigation Services – Training (PANS-TRG, Doc 9868), 3rd edn. Montréal: International Civil Aviation Organization.

IMO (2019). Interim Guidelines for MASS Trials (MSC.1/Circ.1604). London: International Maritime Organization.

INCOSE (2023). INCOSE Systems Engineering Handbook, 5th edn. Hoboken, NJ: Wiley.

ISO (2019). ISO 9241-210:2019 Ergonomics of Human-System Interaction – Part 210: Human-Centred Design for Interactive Systems. Geneva: International Organization for Standardization.

ISO/IEC/IEEE (2023). ISO/IEC/IEEE 15288:2023 Systems and Software Engineering – System Life Cycle Processes. Geneva: International Organization for Standardization.

Klein, G., Woods, D.D., Bradshaw, J.M., Hoffman, R.R. and Feltovich, P.J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. IEEE Intelligent Systems, 19(6), pp. 91–95.

Lee, J.D. and See, K.A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), pp. 50–80.

Leveson, N.G. (2011). Engineering a Safer World: Systems Thinking Applied to Safety. Cambridge, MA: MIT Press.

Leveson, N.G. and Thomas, J.P. (2018). STPA Handbook. Cambridge, MA: Massachusetts Institute of Technology.

National Academies of Sciences, Engineering, and Medicine (2022). Human-AI Teaming: State-of-the-Art and Research Needs. Washington, DC: The National Academies Press.

O'Neill, T., McNeese, N., Barron, A. and Schelble, B. (2022). Human–autonomy teaming: A review and analysis of the empirical literature. Human Factors, 64(5), pp. 904–938.

Parasuraman, R. and Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), pp. 230–253.

Parasuraman, R., Sheridan, T.B. and Wickens, C.D. (2000). 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), pp. 286–297.

Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), pp. 45–77.

Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. Safety Science, 27(2–3), pp. 183–213.

Saunders, M., Lewis, P. and Thornhill, A. (2019). Research Methods for Business Students, 8th edn. Harlow: Pearson.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. International Journal of Human–Computer Interaction, 36(6), pp. 495–504.

Sujan, M., Furniss, D., Grundy, K., Grundy, H., Nelson, D., Elliott, M., White, S., Habli, I. and Reynolds, N. (2019). Human factors challenges for the safe use of artificial intelligence in patient care. BMJ Health & Care Informatics, 26(1), e100081.

Torraco, R.J. (2005). Writing integrative literature reviews: Guidelines and examples. Human Resource Development Review, 4(3), pp. 356–367.

Wilson, J.R., Norris, B., Clarke, T. and Mills, A. (eds.) (2005). Rail Human Factors: Supporting the Integrated Railway. Aldershot: Ashgate.

Yin, R.K. (2018). Case Study Research and Applications: Design and Methods, 6th edn. Thousand Oaks, CA: Sage.

Ziakkas, D. and Henneberry, D. (2025). The challenges of integrating AI in aviation incident-accident investigations: A human-centric approach. In: Intelligent Human Systems Integration (IHSI 2025): Integrating People and Intelligent Systems. AHFE International, 160, pp. 246–255.

Ziakkas, D., Henneberry, D. and Plioutsias, A. (2025). Designing human-centric intelligent systems in aviation: Applications of artificial cognitive systems, AI-enhanced investigations, and immersive eVTOL simulation training. Human-Intelligent Systems Integration. https://doi.org/10.1007/s42454-025-00083-x

Ziakkas, D. and Plioutsias, A. (2024). Artificial Intelligence and Human Performance in Transportation. Boca Raton, FL: CRC Press. https://doi.org/10.1201/9781003480891