Electronic scientific and practical journal

INTELLECTUALIZATION OF LOGISTICS AND SUPPLY CHAIN MANAGEMENT

#37(2026)
June '26

WWW.SMART-SCM.ORG
ISSN 2708-3195
DOI.ORG/10.46783/SMART-SCM/2026-37

ISSN 2708-3195

Electronic scientific and practical collection
INTELLECTUALIZATION OF LOGISTICS AND SUPPLY CHAIN MANAGEMENT

Electronic scientific and practical publication in economic sciences

Electronic scientifically and practical journal “Intellectualization of logistics and Supply Chain Management” included in the list of scientific publications of Ukraine in the field of economic sciences (category "B"): **Order of the Ministry of Education and Culture of Ukraine dated June 11, 2026 No. 928 (Appendix 13 item 205).**

Cluster: Economic Transformation, Business and Administration

Specialties: C1 – Economics and International Economic Relations (by specializations)
D3 – Management
D5 – Marketing

ISSN 2708-3195
DOI: https://doi.org/10.46783/smart-scm/2026-37

The electronic magazine is included in the international scientometric databases:
Index Copernicus, Google Scholar

Released 6 times a year

№ 37 (2026)
June 2026

Kyiv - 2026

Founder: Viold Limited Liability Company
Editor in Chief: Hryhorak M. Yu. – Doctor of Economics, Ass. Professor.

Technical editor: Harmash O. M. – PhD (Economics), Ass. Professor.
Assistant editor: Davidenko V. V. – PhD (Economics), Ass. Professor.

The electronic scientific and practical journal is registered in international scientometric data bases, repositories and search engines. The main characteristic of the edition is the index of scientometric data bases, which reflects the importance and effectiveness of scientific publications using indicators such as quotation index, h-index and factor impact (the number of quotations within two years after publishing).

In 2020, the International Center for Periodicals (ISSN International Center, Paris) included the Electronic Scientific and Practical Edition "Intellectualization of logistics and Supply Chain Management" in the international register of periodicals and provided it with a numerical code of international identification: ISSN 2708-3195 (Online).

Recommended for dissemination on the Internet by the Academic Council of the Department of Logistics NAU (No. 7 of February 26, 2020). Released 6 times a year. Editions references are required. The view of the editorial board does not always coincide with that of the authors.

Electronic scientifically and practical journal "Intellectualization of logistics and Supply Chain Management" included in the list of scientific publications of Ukraine in the field of economic sciences (category "B"): ***Order of the Ministry of Education and Culture of Ukraine dated June 11, 2026 No. 928 (Appendix 13 item 205).***

Cluster: Economic Transformation, Business and Administration

Specialties: C1 – Economics and International Economic Relations (by specializations); D3 – Management; D5 – Marketing

facebook.com/Smart.SCM.org
DOI: https://doi.org/10.46783/smart-scm/2026-37
тел.: (063) 593-30-41
e-mail: support@smart-scm.org
https://smart-scm.org

Contents

DOI: https://doi.org/10.46783/smart-scm/2026-37-1

UDC 004.8:629.7
JEL Classification: L93, O33, K23, L51, M15, D81.

Received: 2026-04-23
Accepted: 2026-05-28
Published: 2026-06-30

Goncharenko K.V. WELL DIGIT LLC, CEO (Ukraine)

ORCID –
Researcher ID
Scopus author id: –
E-Mail: kseniia.goncharenko@welldigit.com

Bugayko D.O. Doctor of Science (Economics), Professor, Academician of the Academy of Economic Sciences of Ukraine, Corresponding Member of the Transport Academy of Ukraine, Instructor of ICAO Institute, Professor (Full) of the Logistics Department Vice Director for International Cooperation and Education of National University "Kyiv Aviation Institute" (Ukraine)

ORCID – 0000-0002-3240-2501
Researcher ID ABF-5564-2021
Scopus author id: – 57216582348
E-Mail: bugaiko@kai.edu.ua

AI IN AVIATION COMPLIANCE MONITORING: SAFETY BARRIERS, REGULATORY GAPS, AND ARCHITECTURAL CONDITIONS FOR TRUSTWORTHY DEPLOYMENT

Kseniia Goncharenko, Dmytro Bugayko "AI in Aviation Compliance Monitoring: Safety Barriers, Regulatory Gaps and Architectural Conditions for Trustworthy Deployment". *The integration of artificial intelligence (AI) in the aviation industry is not just another step in digitalization; it is a deep system transformation that affects safety, personnel, and organizational processes in multiple ways. According to the ICAO standard (Doc 9859) [1], any aviation system is based on the interaction of people, processes, and technologies. Still, traditional Safety Management Systems (SMS) focus on deterministic failure patterns and predictable behavior, while AI is probabilistic in nature, depends on the representativeness of data, and is prone to unpredictable ways. This review systematizes the key challenges of AI in aviation compliance systems, focusing on organizational maturity, rethinking safety barrier models, and the regulation issues. The final section describes a hybrid Knowledge Graph and LLM retrieval framework (Graph RAG) that is proposed architectural response to the hallucination problem and traceability assurance and derives a minimum structural condition for responsible AI deployment in compliance-critical environments.*

Keywords*:* Artificial Intelligence (AI), Aviation Industry, Digitalization, Safety Barriers, Regulatory GAPs, Risk Management

Ксенія Гончаренко, Дмитро Бугайко «Штучний інтелект у моніторингу відповідності авіаційним вимогам: бар'єри безпеки, регуляторні прогалини та архітектурні умови для надійного розгортання». *Інтеграція штучного інтелекту (ШІ) в авіаційну галузь — це не просто*

черговий крок у цифровізації; це глибока системна трансформація, яка впливає на безпеку, персонал та організаційні процеси різними способами. Згідно зі стандартом ICAO (Doc 9859) [1], будь-яка авіаційна система базується на взаємодії людей, процесів та технологій. Тим не менш, традиційні системи управління безпекою (СУБ) зосереджуються на детермінованих моделях відмов та передбачуваній поведінці, тоді як ШІ має ймовірнісний характер, залежить від репрезентативності даних та схильний до непередбачуваних дій. У цьому огляді систематизовано ключові проблеми ШІ в системах авіаційної відповідності, зосереджуючись на організаційній зрілості, переосмисленні моделей бар'єрів безпеки та питаннях регулювання. В заключному розділі описано гібридну структуру пошуку знань на основі графу знань та LLM (Graph RAG), яка є запропонованою архітектурною відповіддю на проблему галюцинацій та забезпечення відстежуваності, а також виводить мінімальну структурну умову для відповідального розгортання ШІ в критично важливих для відповідності середовищах.

Ключові слова: Штучний інтелект (ШІ), авіаційна галузь, цифровізація, бар'єри безпеки, регуляторні прогалини, ризик менеджмент

Introduction. The growing integration of AI changes systems, processes and affects human behavior and decision making, as well as ethical aspects among many organizations, systems and environments.

Traditional safety management systems were developed around deterministic assumptions about failure behavior. AI systems do not always conform to these assumptions. System behavior may change across operational settings, particularly when quality of data or deployment conditions differ from those expected during development. Because of this, existing regulatory and organizational procedures and practices are not always sufficient for assessing AI risks in aviation today.

The purpose of the article. The purpose of this article is to conduct a systematic analysis of the factors and challenges of AI implementation in aviation compliance monitoring system, with particular attention to regulatory issues, human factors, safety implications, and the organizational conditions required for reliable deployment. And also, to consider the proposed architectural solution of the AI system, which is the most relevant in application in the compliance monitoring system.

The main tasks. This article considers the following issues which form a picture from a different angle regarding the challenges and possibilities of implementing AI into the compliance management system:

- How the level of organizational maturity (Safety Maturity Framework) defines the readiness of organizations to implement AI systems and manage AI failures.
- Examining how AI integration changes established safety models such as SHELL and the Swiss Cheese model, creating conditions in which the same error can propagate across both human decision-making and AI-supported operational processes.
- Overviewing the regulatory framework in terms of AI risk management, namely the EU AI Act and EASA amendments to aviation requirements.
- Considering the human factors consequences of AI integration: the double-bind accountability effect, bias, de-skilling, and the decrease of safety reporting culture.
- The potential use of Graph RAG architectures to improve traceability and limit hallucinated outputs in aviation compliance environments. Also, considering the minimum conditions necessary to ensure safe AI introduction in compliance management context.

Methodology. This article is an analytical review. Sources include ICAO and EASA regulatory documents, peer-reviewed human

factors literature, EU legislation, and a pre-print technical paper [24]. They were chosen to represent key regulatory, human factors, and AI architecture sources directly relevant to the analytical arguments.

Presentation of the main results.

1. Organizations' readiness to implement AI: the role of organizational maturity

The effectiveness of implementing advanced technologies such as AI directly depends on the level of the Safety Maturity Framework (SMF) within the organization.

EASA Part-IS contains general guidance, for defining or adopting a maturity model for information systems. The following existing models may be considered: Cybersecurity Capability Maturity Model (C2M2), version 1.1; Systems Security Engineering – Capability Maturity Model (SSE-CMM); NIST Cybersecurity Framework (NIST CSF), version 1.1; and ATM Cybersecurity Maturity Model, edition 1. [2]

The document defines a hypothetical five-level maturity model, ranging from «Non-existent» to «Adaptive», inspired by the «Tier» terminology from the NIST CSF. In fact, the model is based on NIST CSF, together with some elements of ISO/IEC 27001: Initial, Defined, Implemented, Managed, and Improved.

Also, Aviation organizations describe their status of Safety Management system with respect to ICAO SMS framework as following:

- SMS planning performed. Identification of SMS accountable and implementation team, SMS gap analysis, SMS implementation plan, establishment of SMS training plan.

- In addition to the above: reactive processes implemented. Safety policy, safety management responsibilities across the organisation, SMS coordination mechanism, departmental action groups where applicable, SMS documentation and emergency response plan (ERP).

- In addition to the two above: predictive processes implemented. Voluntary reporting procedure, safety risk management procedures, occurrence reporting and investigation procedures, high-consequence safety performance indicators, management of change procedure, internal and external audit programme.

- In addition to all of the above: operational safety assurance implemented. Enhancement of disciplinary procedures, hazard analysis integrates occurrence investigation and voluntary reporting, integration of hazard analysis and risk management with subcontractor's SMS, low-consequence events included in data collection, processing and measuring, SMS audit programmes implemented, training for all relevant personnel completed, safety promotion in place.

Hamidreza Golabchi et al. also, noted that: "To evaluate and enhance safety culture maturity, researchers have developed maturity models, which classify organizations into progressive stages of safety development. Prominent examples include Fleming's safety culture maturity model and Hudson's safety maturity ladder, both of which provide structured pathways for organizations to transition from reactive to proactive safety cultures" [3]. He describes the maturity of organizations as having progressed also through five distinct levels: Initial, Managed Standardized, Advanced, Optimized.

All these maturity definitions have different names but are very similar to each other. Their main common feature is that the top level of maturity ensures proactive actions, forecasting, data governance, constant monitoring and staff involvement, which in turn contributes to the sustainability of the system and organization processes. At this stage, organization achieves a high level of safety maturity, where the focus is on continuous improvement and innovation and there AI solutions can be integrated with maximum value. At the same time

organizational maturity determines not only whether AI can be deployed but also whether the organization can detect and recover from AI-induced failures.

2. Rethinking the safety barriers: SHELL and Swiss Cheese Models in the Age of AI

The traditional SHELL model, as described in ICAO Doc 9859 [1], outlines interactions between Liveware and four main system elements: Software, Hardware, Environment, and other Liveware. Yet, emergence of AI introduces a fundamentally new dimension (layer) to this structure.

Within the SHELL framework, this AI layer cannot be fully classified as either Software or Hardware. Unlike conventional system components, AI systems can interpret, adapt, and make context-dependent decisions, participating directly in operational interactions rather than only supporting them. It does not have to replace existing systems; instead, AI layers are integrated into ongoing processes and software.

As a result, the interaction model within SHELL is changing. It is no longer limited to human–system or human–human interfaces: new ones are emerging, such as Human-AI Teaming, and variants including Human-AI-AI-Human and Human-AI-Human-AI, potentially allowing problems to propagate unchecked across traditional «defense-in-depth» boundaries. These interactions may occur with limited or no direct human oversight [4].

This shift directly affects safety risk management. The addition of an AI layer brings new kinds of dependencies, uncertainties, and emergent behaviors the original SHELL framework does not directly address. Because of this, human factors assessment and risk evaluation methods may need to be reconsidered to reflect these new patterns of interaction.

The Swiss Cheese model which is one of the most widely used frameworks in aviation safety management, developed by Professor James Reason [5], and is specified in ICAO Doc 9859 [1] SMS methodology, illustrates that accidents involve successive violations of multiple defenses. The model proposes that accidents occur when vulnerabilities across multiple independent protective barriers, namely organizational factors, preconditions, unsafe acts, and physical defenses, align simultaneously. The independence of those barriers is the core safety mechanism: a failure penetrating one layer is unlikely to penetrate the next if each layer operates from different information sources, different users, and different institutional standpoints.

AI disrupts this architecture in ways that existing safety models were not designed to anticipate. As Kirwan notes, AI can affect all layers of the Swiss Cheese model simultaneously, either increasing or decreasing the size and number of holes and, critically, reducing the independence between barriers so that the system effectively has fewer protective layers than it appears to have [4].

Two distinct mechanisms drive this risk, and they operate differently enough to warrant separate treatment.

While Kirwan [4] notes that AI can affect all layers of the Swiss Cheese model at the same time, there is a specific issue, namely the loss of independence caused by a shared knowledge-base architecture during system design, which does not appear to have been clearly addressed in previous research. This has an important implication: once independence is removed at the architectural level, operational controls alone cannot restore it.

The first is correlated failure through shared organizational processes. In the compliance monitoring and quality assurance context, traditional safety management relies on planning, execution, types and objects of audits, compliance and safety performance indicators, corrective actions management, staff competency, organizational maturity, etc., each constituting an independent check. If two or more of these layers draw from the same AI knowledge base or consult the same AI-assistant, a systematic error in that base,

such as a misinterpretation of a regulatory requirement or an incorrectly ingested amendment, propagates through every layer simultaneously. The holes align not by chance but by design. This vulnerability is architectural: it is introduced at system design time, before any query has been made, and cannot be fully addressed by operational controls applied after deployment. Restoring genuine independence requires either maintaining separate knowledge sources across critical oversight layers or building verification mechanisms capable of detecting cross-layer convergence on a shared incorrect conclusion before it becomes consequential.

The second mechanism is cascading error propagation across chained human-AI systems. Kirwan identifies an emerging interaction pattern: Human-AI-AI-Human and Human-AI-Human-AI configurations, where multiple AI agents process and pass outputs sequentially, and where a false conclusion generated by one agent can propagate unchecked through subsequent human and AI decision points [4]. This is qualitatively different from the first mechanism. Shared infrastructure creates a single point of failure affecting multiple layers simultaneously. Chained agents create a propagation pathway in which an error moves through the system progressively, potentially amplifying at each stage, until it manifests as an active failure. Classical human-machine interaction models, which assume defined handoff points between a human and a single automated system, have no adequate representation of this failure mode.

Both mechanisms point toward the same implication for investigation and classification frameworks. Human Factors Analysis and Classification System (HFACS) [6] and equivalent taxonomies were developed on the foundation of the Reason model and share its assumptions about human-machine interaction structure. As Kirwan argues, these frameworks will need new categories to capture algorithmic bias, complacency toward AI output, situational awareness loss due to AI opacity, and the propagation dynamics specific to human-AI chains [4].

It creates a safety investigation challenge: to understand the root causes of AI safety impacts or incidents, to review the data that was used, and to understand how AI responded to the user. It is very important to ensure capability of data logging and audit. An organization that integrates a shared AI compliance assistant across compliance and safety management functions, without right architecture approach, has structurally reduced the independence of its safety layer before any operations begin. The combination of training, procedural guidance, or oversight requirements cannot restore independence that was never built into the system architecture.

3. Legal issues: regulation of the EU AI Act and risk assessment

There is a significant contradiction between the legal definition of high-risk AI and its actual hazard to the organization.

One of the most challenging aspects of integrating AI into aviation is the discrepancy between European legal regulation and engineering risk assessment.

According to Article 6 of the European Artificial Intelligence Act [7], in the aviation context, a system is classified as high-risk AI if it is a «safety component» of an aviation product, or if its functions fall under Annex III. At the same time, Article 6(3) contains exceptions (derogations): the system loses its high-risk status if it performs only a narrow procedural task, detects patterns in already made decisions (passive monitoring), or performs a preparatory action without affecting the final decision.

EASA is drafting document DS.AI in order to meet the EU AI Act in terms of «safety component» and interaction AI with human.

It is important point that, AI-based systems that incorporate logic- and knowledge-based AI or hybrid AI, for which failure contribution is more stringent than «no safety effect» are excluded from DS.AI (DS.AI.010 item (d). At the same time AI

system, designed in this approach, is the most reliable for regulatory framework.

To determine if the AI system is high-risk or is excluded, organization should follow EU AI Act provisions, but this do not guarantee the safety itself if Functional Hazard Assessment (FHA) was not performed.

According to the EASA DS.AI, the FHA score classifies hazards on a five-point scale (from H1: unacceptable - potential for fatalities to H5: no risk). If, after aggregating all scenarios, the system obtains the H5 (No risk) level, it is assigned the lowest guaranteed level AL6.

An organization can develop an AI assistant (for example, for audit or compliance management, based on hybrid AI solutions) that successfully defined status under Article 6(3) exceptions and is legally exempt from the strict EU AI Act regulation of high-risk systems. However, an internal risk assessment may show that this supposedly safe tool poses serious threats to business processes (fines, staff discrimination and omission of critical documentary inconsistencies).

In such a situation, if it is general organization, they should voluntarily apply, for example, cross-industry standards such as NIST AI RMF [9] (for enterprise risk management) or HUDERIA (for assessing the impact on people's rights) or other risk management framework to determine its risk tolerance or even abandon the implementation of AI. But if it is aviation organization, there is misunderstanding how to apply the EASA regulation.

The regulatory gap identified above has a direct human consequence. When the legal framework does not clearly allocate responsibility between the AI system and its human operators, individual professionals as auditors, engineers, safety managers are left in an ambiguous accountability position.

4. Human-AI Teaming, Safety Culture, and De-skilling

The implementation of decision-making AI algorithms is classified by EASA as follows. As specified in the EASA Concept Paper [10]: Level 2A corresponds to the implementation of an AI-based system capable of teaming with an end user. The operation is expected to change by moving from human-human teams to human-AI-based system teams. More specifically, Level 2A introduces the notion of cooperation as a process in which the AI-based system helps the end user accomplish their own objective and goal. The operation evolves by considering the work from the AI-based system based on a predefined task allocation pattern e.g., AI advanced assistant supporting landing phases (automatic approach configuration), or conflict detection and resolution in ATM.

This transition creates an acute conflict with the basic principle of safety management - Safety Culture and the safety reporting culture. There is a «Double-Bind» syndrome occurs:

- If an employee does not agree with the recommendation of AI, acts at their own discretion, and an incident occurs, the organization may accuse them of ignoring the technology (algorithm aversion).
- If an employee trusts an algorithm that gives false advice, and this leads to an incident or penalties, they will be accused of overtrust (automation bias) and loss of vigilance.

This double-bind is not merely a theoretical concern. Kirwan documents the legal and professional exposure it creates for individual aviation personnel [4]. Franchina similarly examines its implications for Just Culture frameworks in European aviation [11]. Due to fear of legal responsibility, staff may stop openly reporting AI errors or their doubts, which will destroy the fundamental principle of safety reporting in aviation (reporting culture). To mitigate this, EASA and other methodologies require the introduction of operational explainability mechanisms, where the system can provide information about its level of statistical certainty and protect a person from manipulation.

To overcome this threat, the modern science of Human Factors puts forward two critical requirements:

– *Operational Explainability (OpXAI):* AI should not be a «black box». The algorithm should provide a person with understandable information about the reasons for its decision (rationale), data sources, as well as the metric of its own uncertainty bounds. If the level of AI confidence drops, the system should transfer control to a person, forming the so-called «Calibrated Trust» [4]. Operational Explainability (OpXAI) is a fundamental requirement described in detail in the EASA Concept Paper [10]. In particular, the AI operational explainability section states that the system must provide clear information about how the algorithm obtained its results.

– *Knowledge-Based Behaviour (KBB):* According to the Rasmussen model [12], the highest level of human cognitive activity is knowledge-based behavior (KBB). This classic model of cognitive activity is referenced both in the draft EASA specifications NPA 2025-07(B) Proposed DS.AI [8] (where solutions are divided into Skill-based, Rule-based, and Knowledge-based) and in Kirwan research this was also mentioned [4].

AI assistants take on analytical work, which inevitably leads to the degradation of professional skills of staff (de-skilling). If AI is faced with an unpredictable situation ("corner case"), a person may be unable to take control because they have lost a deep understanding of the physics and logic of processes. Kirwan notes that some skill sets may be lost as automation increases, and that knowledge-based behavior, incorporating both factual or declarative knowledge and experience accumulated over years of operating complex systems represents the most critical and hardest-to-recover cognitive capability [4].

The risk of staff dequalification due to AI is also addressed in the EASA Concept Paper (in the adaptation of the ALTAI checklist, item G6.g) [10]. It states: «Can an AI-based system create a risk of workforce dequalification?... When new working methods are introduced, there is a risk of dequalification, which means that staff will no longer use their competence...». The risk of a person's inability to take control over the limits of the system's knowledge (edge and corner cases) is actively considered in the literature on Human Factors and is mentioned in the requirements for testing HAT systems [4]. On the other hand, it also can be noted that a person's hard skills are just transformed including the cognitive context, so it can be called competence challenges or transformation to the new ones.

5. AI in Compliance Monitoring: advantages and limitations

The implementation of artificial intelligence systems, particularly AI assistants and agents, in compliance monitoring, quality assurance, and knowledge base management is one of the most promising areas of digital transformation. According to research, knowledge base management and IT are the areas where the use of AI solutions, including AI agents is scaling the fastest across different industries [12,13].

5.1. Benefits of AI Assistants in compliance and quality assurance

Effective management of knowledge bases and deep research: AI solutions can perform the function of «deep research», quickly analyzing large volumes of unstructured data, texts, and reports to synthesize complex conclusions from them. AI systems applying natural language processing (NLP) can retrieve relevant content and surface cross-document patterns that manual search would miss or make prohibitively time-consuming. The break-down of data silos between, for example, safety reporting and audit findings, is a frequently cited operational benefit [12,14].

Improving the quality of inspections and audits: During inspections or audits, AI assistants can retrieve applicable regulatory text, highlight the compliance gaps, permissible limits, and historical records in response to natural language queries. This reduces the time an auditor or inspector spends on manual cross-referencing and allows more attention to be given to observation and judgment. The practical value is not in replacing inspector expertise

but in reducing friction in accessing the knowledge that expertise depends on.

Automation of routine processes and reporting: AI solutions can speed up operations that are too time-consuming, automate repetitive tasks, reporting on compliance status, corrective action deadline monitoring, highlight trends and patterns and free the aviation specialist to devote more time to critical thinking and new analyses. For compliance teams already operating under resource pressure, the ability to reduce time spent on structured but low-cognitive-demand tasks creates capacity for more substantive analytical work [16].

Shift from reactive to predictive safety management: Perhaps the most significant potential contribution is in safety intelligence: identifying correlations and trends across incident records, safety reports and audit data that would not be apparent from reviewing cases individually. This capability supports the transition from capturing incidents after they occur to detecting precursors before they result in harm a goal explicitly reflected in ICAO Doc 9859 [1], ICAO Safety Intelligence Manual (Doc 10159) [15], and EASA SMS frameworks.

5.2. Technical limitations and their compliance implications

The limitations of common AI technologies are well-documented in the machine learning publications. Their implications in compliance contexts are serious because the consequences of acting on incorrect information can affect regulatory outcomes and, in safety-critical industries, operational safety.

Hallucinations and reliability: LLMs generate outputs by predicting statistically likely continuations of their input. They do not differentiate between what they know and what they are generating plausibly. The result is a list of failures in which the model produces output that is grammatically coherent and contextually appropriate in tone but factually wrong, citing regulations that do not exist, attributing requirements to incorrect articles, or constructing plausible sounding but erroneous procedural conclusions. In a regulatory retrieval context, this is not just a trouble; it is a failure mode with potential safety implications. [17,4,16].

Black Box problem (Inscrutability): The complexity of machine learning algorithms often renders the AI decision-making process difficult for humans to understand. If the system cannot explain on what data or regulations it detected noncompliance or violations (lack of explainability), it is difficult for auditors to trust such conclusions. EASA NPA 2025-07(B) explicitly addresses this through the requirement for operational explainability (OpXAI) - the ability of a system to provide, in terms an auditor can verify, the basis for each advisory output [8].

Data drift dependency: AI systems reflect the data on which they were designed. Regulatory environments are subject to constant changes: regulatory requirements and standards are amended, new AMC&GM materials are issued, and organizational procedures are often revised. If the knowledge base contains outdated standards, historical bias, or errors, the AI will consistently issue false recommendations. In addition, the lack of AI's ability to understand context (common sense) makes it vulnerable in non-standard situations.

De-skilling: If an AI assistant takes over all the analytical work, there is a high risk of degradation of the professional skills of auditors and inspectors (loss of knowledge-based behavior). In the event of a system failure or a non-standard situation (corner case), the specialist may not be able to independently detect violations or noncompliance. The approaches to auditor competence assessment and their training should be reviewed to mitigate these risks.

5.3. Systemic Risks: when individual limitations become organizational hazards

There are systemic risks that emerge from how these systems are integrated into organizational safety structures. These risks are probably more serious than the technical

limitations themselves because they are less visible and less likely to be identified before causing harm.

1) Bias in AI-assisted compliance monitoring. The most widely discussed is automation bias: when humans work alongside automated systems, they systematically over-rely on system output, even when that output is wrong [18,19,4]. In compliance monitoring it reflects as a specific failure pattern: findings not flagged by an AI assistant are at elevated risk of being overlooked, not because the inspector or auditor lacked competence, but because the system's silence implied acceptability. This is documented behavior in automated decision-support systems across aviation and other high-reliability industries, not a theoretical concern.

The second vector runs in the opposite direction. Historical audit records, finding classifications, and incident reports reflect the judgment of the people who created them. If certain violation types or specific nonconformities were systematically under-reported due to organizational culture, just culture failures, or established inspector blind spots an AI system grounded on that data will reproduce those gaps, consistently and invisibly. NIST AI RMF explicitly addresses this under its MAP and MEASURE functions: bias is not only what a model does with data, but also what the data carried in [9]. In regulated environments where historical records form the primary knowledge base, this is a material risk that pre-deployment data quality assessment cannot fully eliminate.

The third is the most structurally difficult to control. When an auditor reviews an AI recommendation, they are not starting from a neutral position the AI's framing of a finding shapes what they look for and how they interpret the evidence in front of them, even when they believe they are exercising independent judgment. This confirmation bias effect [20] means that human-in-the-loop oversight, the standard mitigation for automation bias, does not straightforwardly restore independence. If the review step is itself shaped by the output it is meant to verify, its value as an independent check is compromised in proportion to how strongly the AI framing anchors the auditor's attention.

2) Double-Bind syndrome in matters of responsibility: AI deployment in compliance contexts creates a structural accountability problem for individual professionals. If an auditor overrides an AI recommendation and an adverse outcome follows, they risk being accused of ignoring available system guidance. If they follow a flawed AI recommendation and an outcome follows from that, they risk being accused of uncritical reliance on an automated tool. This can seriously harm the culture of open safety reporting [4,11].

3) Cybersecurity and sensitive data protection: Compliance monitoring as a safety systems work with sensitive corporate and personal information. There are serious risks of data breaches, privacy breaches (such as GDPR [21]), and cyberattacks such as data poisoning, where an attacker can discreetly change training data or knowledge base content to hide certain types of breaches from AI algorithms. Unlike conventional data breaches, data poisoning may be undetectable until after outcomes have occurred because the system continues to function normally for all other query types.

5.4. The hallucination problem as a structural issue

Automation bias and de-skilling can be partially addressed through training, workflow design, and human oversight requirements. The accountability trap requires institutional and regulatory attention. But hallucination or algorithms failures, the generation of factually incorrect information presented with apparent confidence is a structural property of current LLM architectures, and procedural mitigations (human review, citation checking) are, at best, unreliable controls applied after the fact.

The core issue is that a standard LLM generates responses by sampling from a probability distribution over possible next token. It has no internal representation of what it knows versus what it is estimating. A human expert who does not know something will typically express uncertainty; an LLM will typically produce a confident-sounding response regardless of whether the relevant information was present in its training data. In a regulatory retrieval context, where the specific wording of a standard or the specific article number of a requirement may determine compliance outcomes, this is not an acceptable failure mode.

Requiring auditors to verify every AI citation is reasonable, but it imposes a verification burden that scales with query volume, is susceptible to automation bias (auditors who consistently find citations to be correct will become less rigorous over time), and does not prevent incorrect citations from being acted upon when verification steps are skipped or rushed.

An architectural solution is required one that constrains what the system can generate, not one that attempts to detect errors after generation.

5.5. A proposed architectural response: Graph RAG for compliance environments

Graph Retrieval-Augmented Generation (Graph RAG) is a hybrid architecture that combines the natural language capability of LLMs with the deterministic reasoning of structured Knowledge Graphs [22]. The combination addresses the hallucination problem at its architectural root rather than at the point of output review.

The fundamental distinction from conventional RAG (retrieval-augmented generation) is the structured representation of knowledge [24]. Standard RAG systems retrieve text chunks from a vector database based on semantic similarity to a query; the LLM then generates a response conditioned on the retrieved text. The model still generates, it can still hallucinate because it synthesizes from retrieved fragments, not reading from a verified structured record. Graph RAG replaces the unstructured retrieval layer with a knowledge graph in which entities, relationships, and provenance are explicitly encoded.

The main points:

- *Knowledge graph construction.* A Graph RAG approach is used to extract entities and relationships from source documents: regulatory texts, incident reports, audit records and in which each structured graph node and edge is traceable to a specific source document, article, and version. This step converts unstructured regulatory text into a deterministically queryable knowledge structure. In the aviation safety domain, Iyengar et al. demonstrate this approach at scale, constructing an Aviation Safety Knowledge Graph from NTSB accident reports and FAA databases, yielding over 205,000 entities and 137,000 verified relationships [24].
- *Fact-bound response assembly.* The response is assembled from graph nodes returned by the database query. The system does not generate from its parametric memory; it retrieves from a verified, versioned structured source. A regulatory citation that does not exist in the graph cannot be returned. Iyengar et al. evaluate this grounding property across three dimensions schema accuracy, query precision, and contextual grounding, confirming that responses are strictly derived from retrieved graph nodes rather than hallucinated from base model training data [24].

The practical consequences of this architecture are significant for compliance and safety platforms. First, hallucination is constrained by design: the system cannot return information that is not present in the verified graphs, which eliminates the primary failure mode of factual confabulation. Second, every response is structurally traceable: each advisory output can be accompanied by the specific graph nodes that supported it: source document ID, article reference, amendment

version, and ingestion date, without additional engineering effort. This satisfies the operational explainability requirement (OpXAI) that EASA NPA 2025-07(B) establishes for AI-based systems in aviation [8].

Third, the architecture aligns with EASA's characterization of hybrid AI systems as combining machine learning constituents with logic- and knowledge-based (LKB) constituents. The Knowledge Graph is the LKB element: it provides deterministic reasoning and constrains the probabilistic LLM. EASA's AI Roadmap 2.0 [25] identifies hybrid AI as a priority area for regulatory development, acknowledging that current guidance under DS.AI NPA 2025-07(B) does not fully address hybrid architectures. This represents both a regulatory gap and a research opportunity.

It is important to be precise about what this architecture does and does not solve. It eliminates hallucination in the sense that the system cannot invent facts that are absent from the graph. It does not eliminate errors that originate in the graph itself: an incomplete graph, an incorrectly ingested document, or an outdated knowledge base will produce incorrect responses - confidently, consistently, and without any indication that something is wrong. The quality of the knowledge graph construction process and the currency of the underlying sources are therefore not secondary engineering concerns. It is one of the primary safety controls in the system. Knowledge base change management, amendment tracking, and outdated information detection are as safety critical as the query architecture itself.

5.6. Implications for deployment in safety-critical environments

The analysis above specifies to the following requirements that appear necessary for any AI assistant deployment, based on logic- and knowledge-based approach, in compliance monitoring contexts. These are not novel recommendations, each appears in existing AI governance frameworks [8,9,7].

The knowledge base integrity is probably one of the most consequential and the least visible in current deployment practice. The validity, completeness, and source of the knowledge base on which an AI compliance assistant operates should be actively managed as a safety-critical configuration item. This includes amendment tracking, identification, version control, and a defined change management process for knowledge base updates. A system with excellent architecture but with a poorly maintained knowledge base is not safer than one without those attributes; it may be more dangerous because the architecture creates confidence that the underlying quality does not justify.

Also, human oversight should be architecturally ensured, not procedurally recommended. Workflow should ensure that human review or approval occurs before AI advisory outputs will be implemented into any regulated process, and it is not as a guideline that users are expected to follow, but as a system constraint they cannot avoid. This distinction matters because automation bias operates exactly in the space where human oversight is available but optional. And at the same time there are biases related also to users who provide this oversight. For example, HUDEIRA AI risk methodology specifies «positionality reflection»: developers must examine their own privileges, background, and blind spots to recognize the limits of their own perspective and take into account the viewpoints that are missing for an objective assessment of the impact of AI. So, it is important to provide independent competence assessment of auditors and safety staff, which should include some ethical and positionality issues.

The third one is operational explainability. The ability to explain, in verifiable terms, the basis for an AI recommendation must be a part of the system's architecture, not a post-hoc capability added at the reporting layer. For compliance and safety platforms, it means references and citations to specific source documents, articles, and amendment

versions, not confidence scores or free AI interpretation, which are not interpretable by auditors in regulatory terms.

Besides, important point is effective AI risk management system. This system should cover the full AI system lifecycle: design, development, deployment, and operation, since some failure modes or mitigation actions can only be addressed at the architectural stage.

These conditions map directly onto the organizational maturity framework. A company's readiness for AI is determined not only by its technological base but also by management's ability to decentralize safety management and ensure a continuous improving process. Also, the adoption of AI algorithms requires strong data governance. Thus, AI should be implemented not as an isolated IT solution but as an element of a socio-technical system. Organizations that deploy AI compliance assistants before reaching the appropriate maturity level for each condition are not merely accepting higher risk, they are deploying a system whose safety controls cannot function as designed.

These requirements are understandable to state and are significantly harder to implement in practice. The research challenge is not identifying them but developing verification methods, test procedures, and monitoring options that can confirm they are met and continue to be met over the operational life of a deployed system.

Conclusions. When organizations introduce AI systems before achieving an adequate level of operational readiness, existing safety barriers may no longer function reliably. In the context of safety, the maintaining a high level of organization maturity should be a central corporate goal.

AI can affect the independence of safety barriers. The SHELL and Swiss Cheese models demonstrate this issues in two main ways: correlational failures within shared knowledge architectures and error propagation within aviation compliance workflows. When people or systems use the same AI models and data, the same erroneous output may affect several operational functions at once. Once such dependencies are embedded into operational workflows, procedural controls and additional training may have only limited corrective impact. For this reason, such dependencies should already be considered during early stages of system design.

Aviation regulators are still working toward a harmonized approach to AI governance and certification. EASA in its approaches is guided by the requirements of the EU AI Act, but the regulation is still pending.

The AI architecture considered in this article helps with governance of compliance-related knowledge databases in aviation. The proposed architecture combines logic-based and knowledge-based AI methods within a hybrid framework. However, hybrid AI category is currently excluded from DS.AI, which creates a regulatory gap.

AI-supported decision-making may place personnel in a double-bind situation, contribute to the gradual loss of professional skills and reduce the willingness to submit voluntary safety reports. To mitigate these risks, an AI system should provide operational explainability, and the organization should maintain personnel competence through revised training and competency assessment practices.

Graph-based RAG architectures can improve response reliability where LLM-generated outputs based on structured knowledge representations. An additional advantage of this architecture is improved traceability and explainability of generated responses, which is particularly relevant in aviation compliance context. However, the reliability of such systems still depends on how the underlying knowledge graph is built, maintained, updated, and validated over time.

The field of ongoing work in AI trustworthiness and safety risk management

requires closer collaboration between the AI research community, domain experts, and regulators than has characterized most AI deployment and safety challenges before.

AI will be particularly important in the post-war recovery of Ukraine's air transport industry [26]. The dynamic emergence of new hazards and the increasing intensity of traditional ones, coupled with the objective challenges of ensuring the fourth critical element CE 4 "Technical Personnel Qualification," underscores the relevance and practical value of developing an innovative safety data management tool such as AI.

References

1. International Civil Aviation Organization. Safety management manual (Doc 9859, 4th ed.). ICAO; 2018.
2. EASA. Part-IS: Information security for aviation organisations. European Union Aviation Safety Agency; 2022.
3. Golabchi H, Pereira E, Lefsrud L, Mohamed Y. Proposal of a safety maturity framework in construction: Implementing leading indicators for proactive safety management. Journal of Safety and Sustainability. 2025.
4. Kirwan B. Human factors requirements for human-AI teaming in aviation. Future Transportation. 2025;5:42. https://doi.org/10.3390/futuretransp5020042
5. Reason JT. Human error. Cambridge University Press; 1990.
6. Shappell SA, Wiegmann DA. The human factors analysis and classification system — HFACS. DOT/FAA/AM-00/7. U.S. Department of Transportation / FAA; 2000.
7. European Parliament. Regulation (EU) 2024/1689 of the European Parliament and of the Council (EU AI Act). Official Journal of the European Union; 2024.
8. EASA. NPA 2025-07(B): Proposed DS.AI specifications. European Union Aviation Safety Agency; 2025.
9. National Institute of Standards and Technology. Artificial intelligence risk management framework (AI RMF 1.0). NIST AI 100-1; 2023. https://doi.org/10.6028/NIST.AI.100-1
10. EASA. Concept paper: Guidance for level 1 & 2 machine learning applications, issue 02. European Union Aviation Safety Agency; 2024.
11. Franchina F. Artificial intelligence and the just culture principle. Hindsight. 2023;35:39–42. EUROCONTROL.
12. Rasmussen J. Skills, rules, knowledge; signals, signs, and symbols; and other distinctions in human performance models. IEEE Transactions on Systems, Man, and Cybernetics. 1983;3:257–266.
13. McKinsey & Company. The state of AI in 2025. McKinsey & Company; 2025.
14. McKinsey & Company. Technology trends outlook 2025, fifth edition. McKinsey & Company; 2025.
15. International Civil Aviation Organization. Safety intelligence manual (Doc 10159). ICAO; 2025.
16. Madanchian M, Taherdoost H. The impact of artificial intelligence on research efficiency. Results in Engineering. 2025.

17. Miller JK, Zhao Y. Hallucination in domain-specific large language models: An empirical study in aviation safety. Safety Science. 2024;172:106382.
18. Parasuraman R. Humans and automation: Use, misuse, disuse, and abuse. Human Factors. 1997;39:230–253.
19. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: An attentional integration. Human Factors. 2012;52(3).
20. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. Science. 1974;185:1124–1131.
21. European Parliament. Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union; 2016.
22. Edge D, Trivedi R, Mozafari M. Graph RAG: Unleashing the power of knowledge graphs with large language models. arXiv:2401.15841; 2024.
23. Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems. 2020;33:9459–9474.
24. Iyengar A, Tiselska A, Samaraweera D, Liu H. Building trust in the skies: A knowledge-grounded LLM-based framework for aviation safety. arXiv:2604.13101v1; 2025.
25. EASA. EASA AI roadmap 2.0. European Union Aviation Safety Agency; 2023.
26. Kharazishvili, Y., Kwilinski, A., Bugayko, D., Hryhorak, M., Butorina, V., & Yashchyshyna, I. (2022). Strategic scenarios of the post-war recovery of the aviation transport sustainable development: The case of Ukraine. *Virtual Economics*, 5(3), 7–30. https://doi.org/10.34021/ve.2022.05.03(1)