Advances in IT and Digital Security, Volume 1, Issue 7, pp. 1-42, June 2026.

Advances in IT and Digital Security
ISSN 3143-4266
DAC Insight Publishers
https://journals.dacinsightpublishers.com/AIDY

An AI-Enabled Aviation Integrity Management Framework for Predictive Maintenance, NDT Data, Cybersecurity, and Operational Risk

Iboro Essien[1], Chinenye C. Udensi[2], Serif Oyindamola Oyesiji[3], Chukwudera Obumneke Anunagba[4] and Kingsley Ndupu[5]*

[1] Trivax Energy Services Ltd, Canada
[2] Elizabeth City State University, Elizabeth City, North Carolina, USA
[3] Harrisburg University of Science and Technology, Harrisburg, Pennsylvania, USA
[4] Barclays, USA
[5] Kennesaw State University, USA
*Corresponding author: kingsleynd34@gmail.com

ABSTRACT

Aviation organizations depend on connected sensors, inspection systems, maintenance platforms, operational databases, and artificial intelligence, yet fragmented governance often separates physical integrity, information integrity, cybersecurity, and operational risk—leaving diagnostic ambiguity unresolved when a single anomaly could plausibly reflect degradation, sensor error, process error, software defect, or manipulation. To address this gap, this paper develops the Aviation Intelligence and Integrity Framework (AIIF) through a disclosed-protocol conceptual synthesis of systems safety, asset management, condition-based maintenance, information systems, cyber-physical security, human-automation interaction, and artificial intelligence risk management. The framework links eight constructs across six layers, differentiates AIIF from EASA, ICAO, NIST, DO-178C, DO-326A, ARP4754A, and asset-management approaches through an explicit comparison, and proposes twelve testable relationships, each with its own causal mechanism and boundary conditions. Furthermore, it presents a diagnostic separability model formalized as a likelihood-based classification problem, a measurement model covering all eight constructs, an eight-step implementation process with step-specific entry and exit criteria, and two fully worked application examples alongside six brief illustrative sketches. Overall, the framework offers a structured, testable basis for designing, governing, and monitoring aviation artificial intelligence; while its claims remain hypotheses pending empirical validation, it effectively places operational efficiency within explicit safety and cybersecurity constraints and specifies the studies needed to test each proposition.

Keywords: *aviation integrity; conceptual framework; artificial intelligence; information systems; predictive maintenance; nondestructive testing; cybersecurity; operational risk; digital twin; assurance*

Received 1 June 2026; Received in revised form 20 June 2026; Accepted 22 June 2026, Available online 30 June 2026

1. INTRODUCTION

1.1 Background

Aviation integrity management seeks continued safe operation across aircraft, engines, avionics, airports, energy systems, and information services. Traditional programs organize responsibility by specialty: engineers manage physical assets, IT teams manage platforms, cybersecurity teams protect networks, safety teams assess hazards, and operational teams deliver flights. Artificial intelligence crosses these boundaries because it learns from shared data and influences decisions across them. A single recommendation may depend on a sensor, network, database, feature pipeline, model, interface, and human response, so no one specialty owns the full chain of evidence behind it.

This paper uses "integrity" in four senses that recur throughout the framework and are kept distinct rather than treated as one property. Physical integrity is the condition of a material asset relative to a defined limit, for example remaining wall thickness against a corrosion allowance. Information integrity is the accuracy, completeness, provenance, and authorization state of the data that represents that condition. Model integrity is the fitness of an analytical or learned model for the domain in which it is applied, including its calibration and its behavior outside that domain. Organizational and operational assurance is the set of governance, authority, and change-control arrangements that keep the first three properties valid over time. AIIF treats aviation decision reliability as a joint function of these four kinds of integrity rather than as a single composite score, because a failure in any one of them can produce the same surface symptom (an anomalous reading, a flagged inspection, a model alert) while requiring a different diagnosis and a different response.

Fragmentation creates diagnostic ambiguity precisely because these four kinds of integrity are usually monitored by different teams using different evidence. A decreasing thickness estimate may indicate corrosion, sensor drift, a unit-conversion error, a database mismatch, or malicious alteration. A photovoltaic power deficit may result from contamination, corrosion, recombination losses, inverter behavior, weather-data error, or false telemetry. A weld-image anomaly may represent a defect, an acquisition artifact, a model shift, or an adversarial perturbation. Separate teams may investigate separate explanations while the operational decision remains unresolved, and the eventual resolution often depends on which team happens to look first rather than on which explanation the evidence actually supports.

The physical-integrity literature provides essential grounding for treating these as distinguishable mechanisms rather than a single undifferentiated "anomaly" class. Risk-based NDT links inspection method and interval to degradation and consequence (Atta *et al.*, 2020). Advanced NDT expands available evidence but retains method-specific limits (Atta *et al.*, 2021). Welding quality assurance links inspection results to process control, standards, and disposition (Atta & Adenuga, 2022). Corrosion-inhibitor research shows that protection depends on material and environment (Atta *et al.*, 2018). These findings indicate that AI outputs require mechanism, context, and controlled action; an anomaly score alone cannot substitute for a mechanism-specific explanation.

The same logic applies to airport energy and sensing assets, which increasingly share the same governance boundary as flight-critical systems because they feed the same operational databases and, in some designs, the same analytics pipelines. Charge transport and recombination influence thin-film performance (Atta *et al.*, 2024). Environmental exposure shapes crystalline-silicon degradation (Atta *et al.*, 2025). Spray-pyrolysis conditions influence film formation (Atta, 2025),

while heterojunction interfaces influence nanowire stability (Atta, 2026). Digital analytics should preserve these causal distinctions rather than treating every deviation as a generic anomaly, and the framework below builds diagnostic separability around exactly this requirement.

This paper proposes the Aviation Intelligence and Integrity Framework. AIIF explains how physical evidence, trusted information, hybrid models, human authority, and governance combine to produce reliable action. The framework answers three questions. Which constructs determine trustworthy AI-supported aviation decisions? How do safety and cybersecurity interact with predictive maintenance and NDT? Which relationships should empirical studies test, and with what design? The contribution is a multilevel model with explicit propositions, a differentiated position relative to existing aviation and AI-assurance frameworks, and an implementation pathway.

2. FRAMEWORK DEVELOPMENT APPROACH

2.1 Manuscript type and rationale

AIIF is presented as a conceptual, theory-building framework rather than a systematic review. The distinction matters for how the procedure below should be read. A systematic review reports the state of evidence on a defined question using a reproducible search and a formal quality appraisal of primary studies. This paper instead uses a disclosed-protocol integrative synthesis: it states what was searched, how sources were screened and coded, and how the resulting constructs, layers, and propositions were derived, but its output is a novel multilevel model rather than a summary of existing findings. Framework papers of this kind are conventionally judged on the transparency of the synthesis procedure and the testability of the resulting propositions rather than on review completeness, and the procedure below is written to be reproducible against those criteria.

2.2 Sources searched

The synthesis drew on five literature pools, searched between March and August 2026. *Aviation regulatory and standards literature.* EASA's AI roadmap and concept papers, ICAO's safety management and cybersecurity guidance, the FAA's research plan, and the RTCA and SAE airworthiness and security standards (DO-178C, DO-326A, ARP4754A) were retrieved directly from the issuing bodies' publication libraries.

Peer-reviewed literature on the framework's seven theoretical foundations (systems safety, asset management, condition-based maintenance, information systems, cyber-physical security, human-automation interaction, and AI risk management) was searched in Scopus, Web of Science, and IEEE Xplore using combinations of the terms "aviation OR aircraft OR airworthiness," "predictive maintenance OR condition-based maintenance OR prognostics," "nondestructive testing OR NDT," "cyber-physical security OR OT security," "explainable AI OR trustworthy AI OR AI assurance," "digital twin," and "human-automation interaction OR situation awareness," restricted to work published from 2000 onward, with a concentration on 2015 onward for AI-specific material given the pace of change in that literature.

Mechanism-level physical-integrity and materials evidence was drawn from a defined author-verified pool (the Atta corrosion, NDT, welding, and photovoltaic-degradation series) selected because it supplies mechanism-specific variables, not because it validates an aviation application; Section 9 states this transfer boundary explicitly for every source in that pool.

Adjacent cross-sector evidence (critical-infrastructure cybersecurity, cloud and data governance, industrial predictive maintenance, and safety-learning research) was searched to identify transferable design patterns for layers and constructs that aviation-specific literature covers only thinly, principally cybersecurity governance for AI pipelines and organizational safety-learning practice. This pool was deliberately bounded: a source was retained only where it supplied a construct, a control, or a mechanism that a later section cites by name, not as general supporting bulk. Section 14 lists what remains from this pool after that filter and states, for each retained source, which construct or layer it supports.

Foundational AI-assurance and explainability literature (uncertainty quantification, adversarial robustness, and post-hoc explanation methods such as LIME and SHAP) was searched to ground the model-validity and diagnostic-separability constructs in established technical practice rather than in aviation-specific claims the literature does not yet support.

2.3 Screening and inclusion

A source was included if it met at least one of three tests: it defined a construct or mechanism later formalized in Sections 5 through 8; it supported a specific, named proposition in Section 7; or it documented an aviation-specific hazard, regulatory requirement, or operational constraint that the framework had to accommodate. A source was excluded if its relevance to AIIF was topical rather than mechanistic, for example a general AI-governance survey with no construct, control, or mechanism that a later section actually uses. This test was applied retrospectively to the full draft reference pool assembled during search; sources that survived only as background color were removed in the revision reported in this manuscript, which is why the reference list here is materially shorter than in the version this synthesis first produced. Standards and regulatory documents (EASA, ICAO, FAA, RTCA, SAE, ISO, NIST) were treated as primary evidence of institutional requirements and governance practice, not as peer-reviewed findings, and are cited accordingly.

2.4 Coding and synthesis procedure

Two of the authors independently read the retained sources and extracted, for each one, the mechanism or claim it supported, the aviation or cross-sector context it came from, and the layer, construct, or proposition it was judged to inform. The two extraction sets were compared; disagreements were resolved by a third author reading the source directly and by tightening the construct definition until the source's fit (or exclusion) was unambiguous. Constructs were retained in the model in Section 5 only where at least two independent sources supported their inclusion as analytically distinct from an existing construct, which is the test that led to collapsing several early candidate constructs (for example, separate "sensor quality" and "acquisition procedure" constructs) into physical-state observability, and to keeping diagnostic separability and integration quality as separate constructs from model validity despite their partial overlap, on the grounds that each has its own failure mode: a model can be well calibrated (model validity) yet unable to distinguish competing causes (diagnostic separability), and evidence can be individually valid yet poorly linked across systems (integration quality). The twelve propositions in Section 7 were derived by tracing, for each pair of adjacent constructs in the causal chain shown in Figure 2, whether the synthesized literature supported a directional claim, a moderation claim, or a mediation claim, and by keeping only relationships that a plausible empirical design could falsify with fielded or simulated evidence. Table 1 summarizes the resulting evidence base by pool and disposition.

Table 1: Evidence base by source pool, screening disposition, and role in AIIF. Source: authors' synthesis.

Source pool	Sources screened	Sources retained	Primary role in AIIF
Aviation regulatory and standards	14	11	Governance boundary, Layer 6, comparison table (Section 3)
Theoretical-foundation peer-reviewed literature	46	24	Table 2 foundations, construct definitions (Section 5)
Physical-integrity and materials evidence (Atta series)	9	9	Mechanism-specific variables, diagnostic separability
Adjacent cross-sector evidence	61	19	Layer 3 and Layer 6 controls, implementation process
Foundational AI-assurance and explainability literature	28	16	Model-validity and diagnostic-separability formalization

Generative artificial intelligence assisted language editing and structural organization during manuscript preparation and during this revision. The authors performed the search, screening, coding, and synthesis described above, reviewed and verified the evidence, citations, interpretations, and propositions, and accept full responsibility for the manuscript.

3. POSITIONING AIIF RELATIVE TO EXISTING AVIATION AND AI-ASSURANCE FRAMEWORKS

Aviation already has extensive guidance for airworthiness, safety management, and AI assurance individually. EASA's human-centric AI roadmap, ICAO's safety-management and cybersecurity guidance, DO-178C and DO-326A, ARP4754A, ISO 55000 asset management, and NIST's AI and cybersecurity frameworks each govern a piece of the problem this paper addresses. None of them, individually or in combination as currently practiced, gives an aviation organization a single construct set for deciding whether a specific AI-supported recommendation is trustworthy enough to act on when the evidence behind it could plausibly reflect several different underlying causes. Table 2 maps eight AIIF dimensions against these existing approaches to make that gap, and what AIIF adds to it, explicit rather than asserted.

Table 2: AIIF dimensions compared with established aviation and AI-assurance practice. Source: authors' synthesis.

AIIF dimension	Established practice it draws on	What AIIF adds
Physical-state observability	Risk-based NDT and inspection planning (Atta *et al.*, 2020); ISO 55000 asset data requirements (ISO, 2014)	Defines observability as a construct with formative indicators shared across maintenance, energy, and structural domains, rather than a method-specific inspection coverage metric
Information integrity	DO-326A airworthiness security process (RTCA, 2014); NIST cybersecurity governance (NIST, 2024)	Ties information integrity explicitly to downstream model validity (Proposition 2) rather than treating data protection and model performance as separately governed

AIIF dimension	Established practice it draws on	What AIIF adds
Model validity	EASA's AI roadmap and machine-learning approval research (EASA, 2023, 2024); NIST AI RMF (NIST, 2023)	Specifies calibration, abstention, and external-validity indicators as a measurable construct usable outside EASA's approval process, for organizations not yet seeking certification credit
Diagnostic separability	DO-178C software assurance (RTCA, 2011); ARP4754A system development assurance (SAE International, 2010)	Not established practice: no cited framework formalizes distinguishing physical, sensor, process, software, and malicious causes as a single measurable construct; Section 8 develops this as AIIF's central contribution
Human decision readiness	ICAO safety management (ICAO, 2018); human-automation interaction research (Endsley, 1995; Parasuraman *et al.*, 2000)	Combines competence, workload, and authority into one construct with defined boundary conditions, separated from explanation quality (Section 5), which existing guidance treats as a design checklist item rather than a moderating construct
Integration quality	ISO 55000 lifecycle alignment (ISO, 2014); asset-management maturity practice	Established practice governs integration within one specialty (asset management, IT, or security); AIIF treats cross-specialty traceability itself as a measurable construct
Assurance continuity	DO-178C change control (RTCA, 2011); NIST AI RMF monitoring function (NIST, 2023)	Extends change control from software configuration to model drift, supplier change, and organizational change as one continuity construct with shared indicators
Decision reliability	ICAO safety-management outcome metrics (ICAO, 2018)	Reframes the outcome variable from safety-event counts alone to a risk-weighted, closure-inclusive measure that also accommodates non-safety-critical AI use

Three conclusions follow from Table 2. First, most of AIIF's individual constructs have a direct antecedent in existing practice; the framework does not claim to invent physical-state observability or information integrity from nothing. Second, several constructs (information integrity linked to model validity, human decision readiness separated from explanation quality, integration quality as a measurable cross-specialty property) integrate existing practice in a way current guidance does not, because that guidance is organized by specialty rather than by decision. Third, diagnostic separability has no direct antecedent among the compared frameworks; it is the element of AIIF that existing aviation and AI-assurance guidance does not yet formalize, and Section 8 develops it accordingly as the framework's central and most novel contribution. This differentiation is a claim about structure, not about maturity: every AIIF construct still requires the aviation-specific hazard analysis, certification evidence, and operational trial that Table 2's established-practice column represents, and Section 13 states that requirement explicitly.

4. THEORETICAL FOUNDATIONS

Table 3: Theoretical foundations of AIIF. Source: authors' synthesis.

Foundation	Role in AIIF
Systems safety	Safety emerges from controlled interactions among components, organizations, and decisions. Hazard analysis should examine unsafe control actions and feedback, not only component failure (Leveson, 2011).
Asset management	Value, risk, performance, and lifecycle decisions should align across assets and organizational objectives (ISO, 2014).

Foundation	Role in AIIF
Condition-based maintenance	Diagnostics and prognostics convert observed condition into maintenance decisions, subject to data quality, failure modes, and uncertainty (Jardine *et al.*, 2006; Lei *et al.*, 2018); deployed models require ongoing surveillance because operating conditions drift (Gama *et al.*, 2014).
Information systems	Data architecture, interoperability, configuration control, identity, and service continuity shape whether analytics enters work reliably; least privilege, complete mediation, and economy of mechanism remain the relevant design principles for that architecture (Saltzer & Schroeder, 1975).
Cyber-physical security	Physical processes and computation constrain each other, allowing independent physical relationships to expose manipulated data (Humayed *et al.*, 2017).
Human-automation interaction	Automation changes information acquisition, analysis, choice, and action. Appropriate allocation preserves situation awareness and responsibility (Endsley, 1995; Parasuraman *et al.*, 2000).
AI risk management	Trustworthy AI requires governance, mapping, measurement, and management across the lifecycle (NIST, 2023; EASA, 2023), grounded in established practice for uncertainty quantification (Goodfellow *et al.*, 2015) and post-hoc explanation (Ribeiro *et al.*, 2016; Lundberg & Lee, 2017).

No single theory in Table 3 covers material degradation, digital integrity, organizational authority, and adaptive computation together, which is why AIIF combines seven rather than adopting one wholesale. The integration uses the operational decision, not the asset or the model, as the common unit of analysis: evidence about an asset travels upward through the layers defined in Section 6, a person or an approved automated function acts on it within the authority that decision carries, and feedback from that action updates the integrity record that later decisions will draw on. Each foundation contributes a different part of that path rather than a redundant restatement of the same integration logic. Systems safety and human-automation interaction jointly explain why the framework treats the human decision step as a control action with its own failure modes, not as a neutral endpoint that simply receives a model output. Asset management and condition-based maintenance jointly explain why physical-state observability and assurance continuity are defined as lifecycle properties rather than point-in-time measurements. Information systems and cyber-physical security jointly explain why information integrity is defined broadly enough to include configuration and provenance, not only data accuracy, and why an independent physical relationship (for example, a redundant sensor obeying different physics than the primary one) is treated as security-relevant evidence, not only as a reliability check. AI risk management supplies the lifecycle governance vocabulary, mapping, measuring, and managing, that Sections 10 and 11 use to operationalize model validity and diagnostic separability into testable indicators.

Regulatory and standards guidance completes the foundation set by defining the boundary within which any of this can be deployed. EASA links aviation AI to human-centered assurance and approval evidence (EASA, 2023, 2024). The FAA's research plan sets research priorities across safety, operations, and emerging technology (FAA, 2023). ICAO connects cybersecurity and safety management to accountable risk controls (ICAO, 2018, 2019). DO-178C governs airborne software evidence and DO-326A addresses airworthiness security processes (RTCA, 2011, 2014); ARP4754A connects aircraft and system development to structured assurance (SAE International, 2010); ISO 55000 and ISO 31000 provide the asset-management and risk-management vocabulary Table 3 draws on (ISO, 2014, 2018). These sources are treated as evidence of institutional requirements and governance practice, not as peer-reviewed empirical findings, and every AIIF

construct that touches certification, approval, or airworthiness remains subordinate to them; Section 3 states this relationship as an explicit comparison rather than a general disclaimer.

5. CONSTRUCT DEFINITIONS

AIIF is built from eight core constructs, defined operationally in Table 4 and developed in the paragraphs that follow.

Table 4: Core AIIF constructs. Source: authors' synthesis.

Construct	Operational definition
Physical-state observability	Extent to which sensors and inspections reveal the asset and environmental conditions relevant to a decision.
Information integrity	Accuracy, completeness, timeliness, provenance, semantic consistency, and protection of data and configurations.
Model validity	Demonstrated fitness of an AI or hybrid model within a defined operating domain, including calibration and uncertainty.
Diagnostic separability	Ability to distinguish physical degradation, sensor failure, process error, software defect, and malicious manipulation.
Human decision readiness	Competence, situation awareness, authority, workload capacity, and access to safe fallback available to the responsible user.
Integration quality	Strength of traceable links among assets, evidence, models, users, actions, and governance records, including the explanation quality of the interface each layer presents to the next.
Assurance continuity	Degree to which evidence remains valid and controlled through drift, updates, incidents, and organizational change.
Decision reliability	Consistency with which actions manage operational risk within approved limits and achieve the intended outcome.

Physical-state observability is the extent to which sensors and inspections reveal the asset and environmental conditions relevant to a decision. The construct is measured through multiple indicators because a single technical metric hides organizational and operational variation. Following standard practice for multi-indicator constructs, the framework distinguishes formative indicators, which jointly define the construct and need not correlate with one another, from reflective indicators, which are interchangeable effects of it and should correlate. For physical-state observability, sensor and inspection-method coverage of the relevant failure modes, probability of detection, and independent-modality count are formative: each captures a different, non-substitutable aspect of what "observable" means for a given asset. Missingness by operating condition and the rate of unexplained discrepancies between independent readings are reflective: both should fall together as observability genuinely improves, and either alone is an adequate, if noisy, indicator of the underlying construct. This distinction, and the justification for it, is repeated for each construct below because the choice of formative versus reflective specification changes how a future study should compute the construct and interpret its indicators, a point Section 11 returns to for the full measurement model.

Information integrity is the accuracy, completeness, provenance, timing, authorization state, and configuration context of data moving through the decision chain. Provenance coverage and timestamp accuracy are formative, since a record can have full provenance but stale timestamps

or the reverse. Schema-error rate, unauthorized-change rate, and recovery time are reflective consequences of the same underlying integrity state.

Model validity is the degree to which a model performs reliably within its approved operating domain. Discrimination, calibration, and robustness are formative, since a model can discriminate well while being poorly calibrated. Abstention quality, external validity, and risk-weighted error are reflective indicators of overall validity once the formative components are in place.

Diagnostic separability is the ability to distinguish physical degradation from sensor failure, process error, software defect, and malicious manipulation, defined and formalized as a decision problem in Section 8. Class-specific sensitivity and evidence independence (whether the discriminating signals come from genuinely different measurement channels) are formative; causal consistency across repeated cases and correct-escalation rate are reflective.

Human decision readiness combines competence, situation awareness, workload capacity, authority, and access to a rehearsed fallback. This definition deliberately excludes explanation quality, which the original construct set conflated with readiness; explanation quality is instead a property of the human-machine interface itself and is carried under integration quality, so that Proposition 5 can treat explanation quality as a predictor of decision reliability and human decision readiness as a separate, person-level moderator of that relationship without the two constructs overlapping. Competence and authority are formative; appropriate reliance, escalation accuracy, and recovery time under degraded conditions are reflective.

Integration quality is the strength of traceable links among assets, evidence, models, users, actions, and governance records, including the explanation quality of the interface each layer presents to the next. Identity consistency and interface error rate are formative; trace completeness and closed-loop action verification are reflective.

Assurance continuity is the preservation of valid evidence through drift, updates, incidents, supplier changes, and organizational change. Drift-detection delay and audit completeness are formative; regression-escape rate, rollback performance, and incident-learning time are reflective.

Decision reliability is the consistency with which authorized actions keep risk within approved limits and achieve the intended operational outcome. Correct-action rate and risk-weighted error are formative; repeat-inspection rate, unscheduled-removal rate, disruption, and closure quality are reflective. Decision reliability is the framework's outcome construct; Section 7 develops how the other seven constructs are proposed to relate to it.

6. AIIF ARCHITECTURE

AIIF organizes these eight constructs into six architectural layers, summarized in Table 5 and shown with their evidence-flow and governance-feedback paths in Figure 1.

Table 5: Six-layer AIIF architecture. Source: authors' synthesis.

Layer	Contents
Layer 1: Physical assets and environment	Aircraft structures, engines, avionics, facilities, photovoltaic systems, climate, loads, and operating conditions.
Layer 2: Trusted sensing and inspection	Sensors, NDT, calibration, acquisition procedures, inspectors, probability of detection, and independent confirmation.

Layer	Contents
Layer 3: Governed information infrastructure	Asset identity, data models, integration, provenance, access control, configuration, storage, and service continuity.
Layer 4: Hybrid intelligence and digital twins	Rules, physics, statistical learning, simulation, uncertainty, anomaly detection, diagnosis, and prognosis.
Layer 5: Human decision and controlled action	Role-specific explanation, authority, work orders, operational restrictions, escalation, override, and feedback.
Layer 6: Safety-security governance	Hazards, threats, risk acceptance, assurance cases, supplier controls, monitoring, incident response, and change approval.

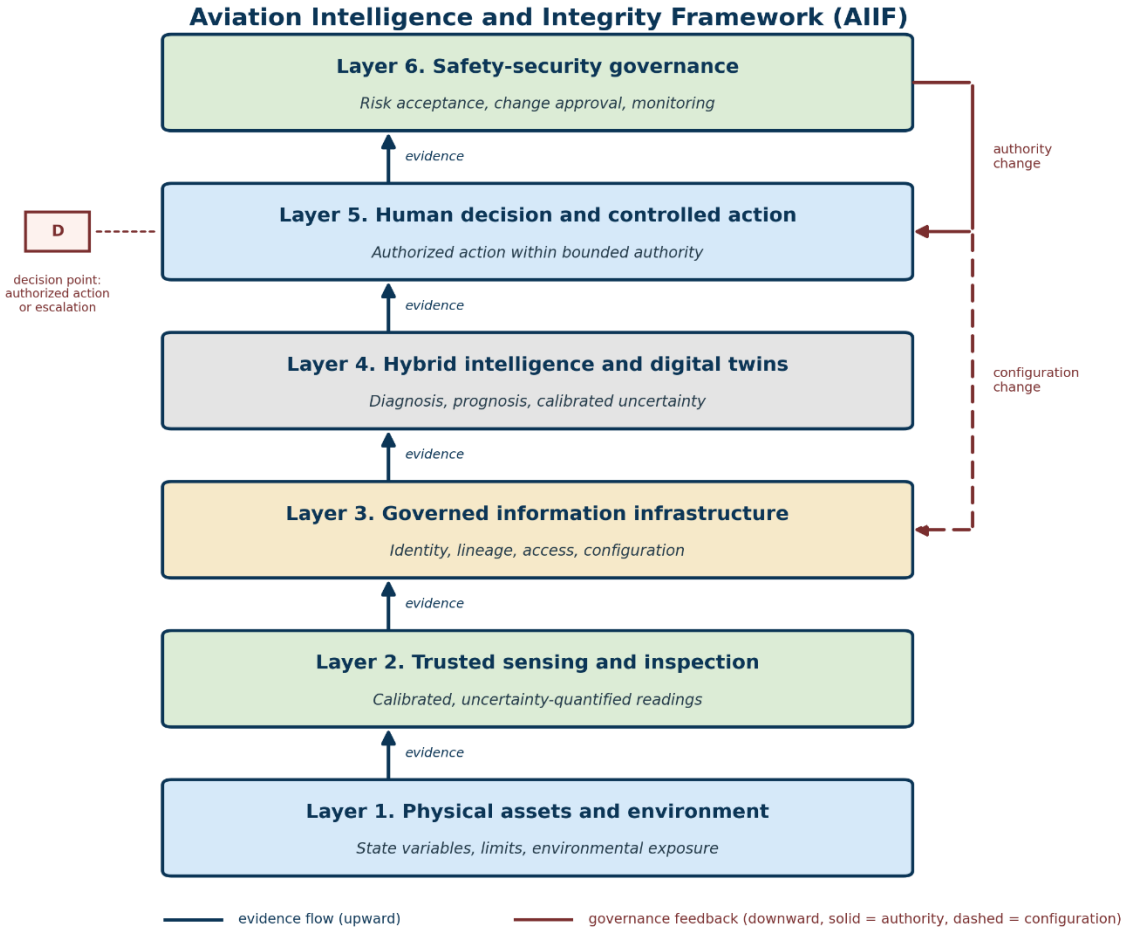


Figure 1: Six-layer Aviation Intelligence and Integrity Framework, with evidence flow, decision points, and the assurance feedback path labeled. Source: authors' synthesis.

Each layer below is described through the same seven elements, purpose, inputs, outputs, responsible roles, controls, characteristic failure modes, and assurance evidence, so that the layers can be compared, but the content under each element is specific to that layer rather than a restatement of the layer above or below it. One illustrative aviation domain is developed per layer rather than the same three domains repeated across all six, so that the architecture reads as six analytically distinct stages rather than one template applied six times.

6.1 Layer 1: Physical assets and environment

Purpose. Layer 1 represents the asset and its operating environment as the ground truth that every other layer ultimately references. It does not sense or interpret; it defines what exists to be sensed and interpreted, including the limits (corrosion allowance, fatigue life, thermal envelope) against which later layers judge deviation.

Inputs. Type-design and maintenance-program limits, material specification, build and modification records, and environmental-exposure history for the specific asset and location.

Outputs. A defined set of state variables and degradation mechanisms, each with an approved limit and an owner, that Layer 2 is built to observe.

Responsible roles. Design and structures engineering define the limits; the asset owner maintains the configuration record against which Layer 2 evidence is later compared.

Controls. Configuration baseline management, life-limited-part tracking, and environmental-exposure monitoring tied to the asset's actual operating history rather than a fleet average.

Characteristic failure modes. Undocumented modification or repair, environmental exposure outside the design envelope, and configuration records that fall out of step with the physical asset, each of which invalidates a later layer's evidence without that layer necessarily detecting it.

Assurance evidence. Type-design and modification records, maintenance history, and environmental logs tied to the specific tail number, engine, or facility.

For airframe corrosion and structural health, Layer 1 must represent humidity, salts, dissimilar-metal couples, coatings, joints, and hidden moisture paths as explicit state variables with defined limits, because the mechanism, not just the symptom, determines what Layer 2 needs to observe and what Layer 4 can validly infer from it (Atta *et al.*, 2018, 2020). A thickness limit defined without the associated environmental-exposure variable cannot later distinguish ordinary corrosion progression from an anomalous rate that should trigger escalation.

6.2 Layer 2: Trusted sensing and inspection

Purpose. Layer 2 acquires evidence about Layer 1 state with known, bounded uncertainty, and attaches that uncertainty to every reading rather than passing forward a bare value.

Inputs. The Layer 1 state-variable and limit definitions, plus the sensor, NDT method, and inspector assigned to observe them.

Outputs. Time-stamped, uncertainty-quantified readings tied unambiguously to asset identity, method, and operator, forwarded to Layer 3 with acquisition metadata attached rather than as a bare measurement.

Responsible roles. NDT inspectors and instrumentation engineers acquire the evidence; quality assurance verifies method qualification and calibration currency.

Controls. Calibration traceability to a reference standard, probability-of-detection studies for each method-defect combination in use, and an independent-confirmation protocol for high-consequence findings.

Characteristic failure modes. Drift and miscalibration, coupling or contact loss, inspector variability, and image or signal artifacts that a downstream model could mistake for a genuine finding.

Assurance evidence. Calibration certificates, probability-of-detection study results, and inspector qualification and currency records.

For weld and bonded-joint inspection, Layer 2 must represent manufacturing discontinuities, fatigue cracks, porosity, lack of fusion, debonding, and repair quality through the specific method (ultrasonic, radiographic, thermographic, acoustic, visual, or digital) used to detect each, because method-specific detection limits, not a single inspection score, determine what Layer 4 can later separate from a competing explanation (Atta *et al.*, 2021; Atta & Adenuga, 2022).

6.3 Layer 3: Governed information infrastructure

Purpose. Layer 3 preserves the identity, meaning, and authorization of evidence as it moves between the systems that acquire it and the systems that interpret it, which is where most aviation AI failures originate even though each individual component works locally.

Inputs. Readings and acquisition metadata from Layer 2, asset identity from Layer 1, and the access and retention policy set by Layer 6.

Outputs. Governed data products, schema-conformant, lineage-tracked, access-controlled, and quality-flagged, that Layer 4 can consume without independently re-verifying provenance.

Responsible roles. Data engineering and IT maintain the infrastructure; information security enforces access control and change management.

Controls. Master data management for asset identity, schema governance, configuration management for the data pipeline itself, and a defined retention and disposal policy.

Characteristic failure modes. Schema drift between system versions, identity mismatch across systems that name the same asset differently, stale cache serving outdated readings as current, and unauthorized or unlogged data change.

Assurance evidence. Data-lineage records, access logs, and a schema and configuration change history for the pipeline.

For engine and rotating-component maintenance, Layer 3 must reconcile engine health monitoring, oil-debris, vibration-spectrum, and borescope-image records that typically originate in separate vendor systems under separate identifiers for the same physical engine, because a governed link between those identifiers is what allows Layer 4 to treat them as corroborating or conflicting evidence about the same asset rather than as unrelated data streams (Jardine *et al.*, 2006; Lei *et al.*, 2018).

6.4 Layer 4: Hybrid intelligence and digital twins

Purpose. Layer 4 converts governed evidence into diagnosis, prognosis, and a recommended action, with an explicit account of which competing cause the evidence best supports and how confident that judgment is.

Inputs. Governed data products from Layer 3, physics-based or engineering models of the relevant degradation mechanism, and historical labels where available.

Outputs. A diagnosis that states competing-cause likelihoods rather than a single anomaly score, a prognosis where applicable, a recommended action, and a calibrated uncertainty estimate.

Responsible roles. Data science and reliability engineering build and maintain the model; the domain engineer who owns the physical mechanism validates its physical plausibility.

Controls. A validation protocol using temporal and external holdouts, defined abstention rules for low-confidence cases, and champion-challenger testing before a model update is promoted.

Characteristic failure modes. Distribution shift between training conditions and current fleet or environmental conditions, mislabeled or single-cause-only training data that cannot represent competing causes, and overconfident extrapolation outside the validated domain.

Assurance evidence. Validation reports, calibration curves, and a version record linking each deployed model to the evidence that approved it.

For airport renewable-energy and sensing infrastructure, Layer 4 must combine a physics-based model of charge transport and recombination with a data-driven anomaly model, because photovoltaic power deficits share surface symptoms (a falling output curve) across mechanistically distinct causes, contamination, recombination loss, inverter behavior, and manipulated telemetry, that only a hybrid model with mechanism-aware features can begin to separate (Atta *et al.*, 2024, 2025).

6.5 Layer 5: Human decision and controlled action

Purpose. Layer 5 converts a Layer 4 recommendation into an authorized action within the competence and authority of a specific, accountable person, or into an approved automated action bounded by an explicit authority limit.

Inputs. The Layer 4 diagnosis with its explanation and uncertainty, and the role-specific authority limits set by Layer 6.

Outputs. A work order, operational restriction, escalation, or documented override, each with its rationale and the evidence it relied on recorded against the decision.

Responsible roles. The maintenance controller, dispatcher, duty engineer, or flight crew member accountable for the specific decision domain.

Controls. An authority matrix defining what each role may decide unilaterally versus escalate, a rehearsed fallback procedure for when Layer 4 output is unavailable or not trusted, and mandatory logging of any override with its stated reason.

Characteristic failure modes. Automation surprise, where the human does not understand why the system recommended an action; excessive delegated authority relative to the uncertainty actually present; and poorly communicated uncertainty that leaves the human unable to tell a routine recommendation from a marginal one.

Assurance evidence. Decision and override logs, and training and competence records for each authorized role.

For avionics and software assurance, Layer 5 must present flight-relevant uncertainty (a sensor-fault flag, a reversion to a degraded control law) in a form the flight crew can act on within seconds, which is a materially different design problem from a maintenance controller's minutes-to-hours decision window, and the framework treats this time-criticality as a boundary condition on how much explanation Layer 5 can present before it becomes a hazard in its own right (RTCA, 2011, 2014).

6.6 Layer 6: Safety-security governance

Purpose. Layer 6 sets risk acceptance criteria, approves changes to any layer below it, and monitors the system as a whole, joining safety and security governance rather than running them as parallel, uncoordinated processes.

Inputs. Monitoring data, incident reports, audit findings, and supplier or configuration change notices from every layer.

Outputs. Risk-acceptance decisions, approved or rejected changes, an updated assurance case, and corrective actions that flow back down to Layer 5 (authority changes) or Layer 3 (configuration changes).

Responsible roles. A joint safety-security governance function, quality assurance, and an executive risk owner accountable for residual risk.

Controls. A maintained assurance case linking claims to evidence, a change-approval board covering both safety and security changes, an incident-response process, and supplier oversight covering data, model, and platform dependencies.

Characteristic failure modes. Safety and security governed as separate processes that each declare an incident outside their scope when the evidence is genuinely ambiguous between a physical and a malicious cause, and incident learning that is too slow to prevent recurrence.

Assurance evidence. The assurance case itself, audit reports, change-approval records, and incident-investigation reports.

For aviation cybersecurity and data integrity, Layer 6 is where a false-data-injection or spoofing event and a genuine sensor or physical fault must be adjudicated jointly, because containment actions appropriate to a cyber event (isolating a data feed) can themselves create an availability risk that a purely physical fault would not justify, which is the specific joint-governance problem Proposition 11 formalizes (ICAO, 2019; NIST, 2024).

6.7 Inter-layer interfaces

Table 6 specifies what crosses each adjacent layer boundary, in what form, and what verification applies before the receiving layer may rely on it, replacing the repeated general statement that "inputs and outputs must be controlled" with the actual control at each boundary.

Table 6: Inter-layer interfaces in the AIIF architecture. Source: authors' synthesis.

Boundary	What crosses it	Form	Verification before reliance
Layer 1 to Layer 2	Asset identity, state-variable definitions, limits	Configuration record reference	Configuration record matches physical asset (walk-down or equivalent)

Boundary	What crosses it	Form	Verification before reliance
Layer 2 to Layer 3	Readings with acquisition metadata	Time-stamped record with method, operator, and uncertainty attached	Calibration currency and method qualification confirmed
Layer 3 to Layer 4	Governed data products	Schema-conformant, lineage-tracked record	Lineage complete to Layer 2 source; quality flags absent or explicitly handled
Layer 4 to Layer 5	Diagnosis, prognosis, recommendation, uncertainty	Role-appropriate explanation with competing-cause likelihoods	Model version within its approved and current validation domain
Layer 5 to Layer 6	Decision and override record, escalation	Logged decision with rationale and evidence reference	Authority matrix confirms the decision was within the deciding role's authority
Layer 6 to Layer 3/5 (feedback)	Approved change, authority adjustment, restriction	Change record referencing the assurance case claim it affects	Change approved through the change-approval board before it takes effect

7. CAUSAL LOGIC AND PROPOSITIONS

The causal chain linking the eight constructs is summarized in Figure 2 and stated as twelve testable propositions in Table 7, each developed individually in Sections 7.1 through 7.12.

Table 7: Testable AIIF propositions. Source: authors' synthesis.

ID	Proposition
P1	Greater physical-state observability is positively associated with model validity because representative, mechanism-relevant evidence reduces label ambiguity.
P2	Information integrity is positively associated with model validity, and this relationship strengthens as system integration complexity rises.
P3	Hybrid physics-data modeling improves diagnostic separability when physical failure, sensor error, and malicious manipulation obey different constraints.
P4	Diagnostic separability mediates the relationship between model validity and decision reliability.
P5	Human decision readiness strengthens the positive relationship between explanation quality and decision reliability.
P6	Excessive automation authority weakens decision reliability when uncertainty or operating-domain boundaries are poorly communicated.
P7	Cybersecurity control maturity strengthens the relationship between information integrity and decision reliability.
P8	Integration quality reduces the time between anomaly detection, diagnosis, authorized action, and verified closure.
P9	Assurance continuity strengthens the effect of initial model validity on long-term decision reliability under distribution shift.
P10	Independent multimodal evidence reduces false diagnoses more strongly in high-ambiguity cases than in routine cases.

ID	Proposition
P11	Joint safety-security governance reduces residual operational risk more than separate governance when cyber and physical causes share observable symptoms.
P12	Decision reliability improves operational efficiency only when fallback design prevents risk transfer during data, model, or platform failure.

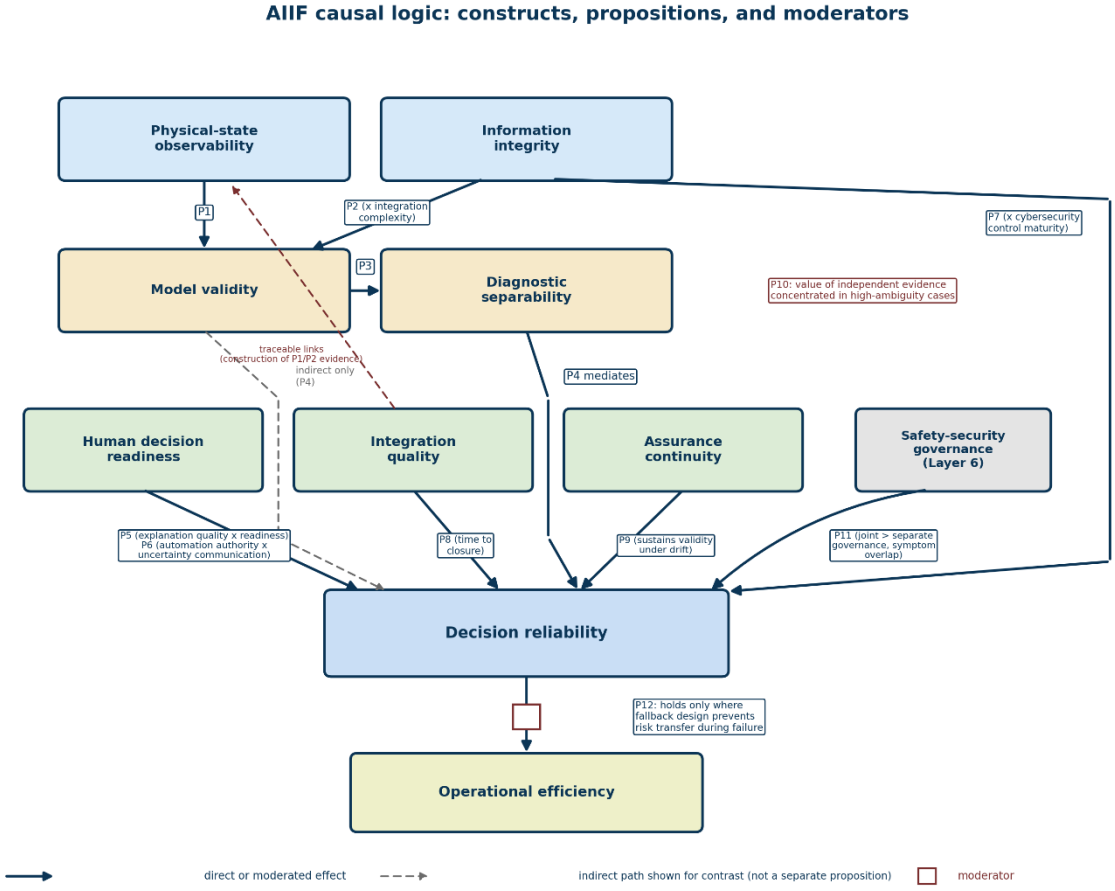


Figure 2: AIIF causal logic and assurance feedback, with each labeled arrow naming the proposition it represents. Source: authors' synthesis.

7.1 P1: Physical-state observability and model validity

Mechanism. Representative, mechanism-relevant evidence reduces label ambiguity at the point a model is trained or validated: when the sensing layer captures the actual degradation mechanism rather than a correlate of it, the labels a model learns from are less likely to encode the wrong cause, so calibration and discrimination both improve.

Boundary conditions. The effect weakens when hazards are inherently poorly observable (subsurface flaws below method detection limits), when ground-truth labels are delayed relative to the decision they would inform, or when field workarounds bypass the sensing the model was trained on. It should strengthen where independent sensing modalities, stable configuration, and explicit uncertainty reporting are already in place.

Competing explanations. Fleet age, maintenance-program maturity, and workforce experience could each independently improve apparent model validity without any change in observability itself, so a test of P1 must control for these rather than attribute all improvement to observability.

Empirical test. A prospective multilevel study comparing model calibration and discrimination across assets or sites with naturally varying observability (number of independent sensing modalities, inspection coverage), controlling for fleet age, maintenance maturity, and workforce experience as covariates, with observability measured prior to and independent of the model outcomes being explained.

7.2 P2: Information integrity, model validity, and integration complexity

Mechanism. Integrity failures, mislabeled assets, stale timestamps, unauthorized changes, inject noise into the data a model consumes. That noise compounds as more systems intermediate the same evidence, because each additional handoff is another point where identity or meaning can be lost, so the same underlying integrity failure does more damage to model validity in a highly integrated pipeline than in a simple, single-source one.

Boundary conditions. The moderation should be strongest where integration spans more than two independently governed systems and weakest where a single team controls acquisition through model input with no intermediate handoff.

Competing explanations. General IT operational maturity could explain both higher information integrity and higher model validity without a causal link between them; a test of P2 needs an integrity measure that varies independently of overall IT maturity, for example a targeted master-data-quality intervention.

Empirical test. A quasi-experimental before-after comparison of model validity following a defined information-integrity intervention (master-data cleanup, provenance-tagging rollout) at sites that differ in integration complexity, testing whether the pre-post validity gain is larger at higher-complexity sites.

7.3 P3: Hybrid physics-data modeling and diagnostic separability

Mechanism. Physics constraints rule out explanations that are mechanically implausible given the observed evidence, narrowing the hypothesis space before any learned component acts; the data-driven component then captures patterns physics does not encode, such as acquisition artifacts or adversarial perturbations. Separability improves specifically when the competing causes obey genuinely different constraints, because that is what gives the physics component something to rule out.

Boundary conditions. The advantage over a purely data-driven model should be largest when labeled examples of every competing cause are scarce, since physics constraints substitute for missing labels; it should shrink as labeled multi-cause data becomes abundant, since a sufficiently large dataset can let a purely data-driven model learn the same distinctions empirically.

Competing explanations. A purely data-driven model trained on enough labeled multi-cause data might match hybrid performance without any physics component, which is the alternative P3 must be tested against rather than against a null model.

Empirical test. A controlled fault-injection study comparing pure-physics, pure-data, and hybrid models' class-specific sensitivity and specificity across physical, sensor, and injected-manipulation cause classes, with labeled training data held constant across the three model types.

7.4 P4: Diagnostic separability as a mediator between model validity and decision reliability
Mechanism. A model can be well calibrated in the aggregate (high validity) while still recommending the wrong type of action, because calibration measures whether a probability is trustworthy, not whether the model has identified the correct cause. Diagnostic separability is the missing link: model validity improves decision reliability specifically by improving the correctness of the causal attribution the human or automated system acts on.

Boundary conditions. The mediated path should be weaker where the appropriate action does not depend on the specific cause, for example when any anomaly above a threshold triggers the same conservative response regardless of whether it is physical or a sensor fault; in that setting validity can improve reliability directly without separability doing any mediating work.

Competing explanations. A direct validity-to-reliability path that does not route through separability would appear as a significant direct effect alongside a null indirect effect in a mediation model; P4 predicts a significant indirect effect specifically.

Empirical test. A field mediation analysis with bootstrapped confidence intervals on the indirect effect, using paired records of model validity, diagnostic-separability performance, and downstream decision-reliability outcomes from the same case population.

7.5 P5: Human decision readiness as a moderator of explanation quality
Mechanism. An explanation only improves a decision if the person receiving it has the competence, authority, and available attention to use it; absent that readiness, additional explanation is inert at best and an added workload cost at worst. Positioning readiness as a moderator of the explanation-reliability relationship, rather than as a component of readiness itself, is what P5 depends on, and Section 5 states why explanation quality is defined outside the human decision readiness construct for exactly this reason.

Boundary conditions. The moderation may be curvilinear: at very high readiness, additional explanation quality could show diminishing returns because the operator already has enough context; at very low readiness, explanation quality may show little effect because the operator cannot act on it regardless of how it is presented.

Competing explanations. Explanation quality might have a direct, readiness-independent effect on decision reliability if a sufficiently well-designed explanation can compensate for low readiness rather than requiring it; this is the null-interaction alternative a test of P5 must rule out.

Empirical test. A factorial simulation experiment manipulating explanation quality and operator readiness (via role, training level, or induced workload) independently, measuring decision reliability as the outcome and testing for the interaction term specifically, not only the two main effects.

7.6 P6: Automation authority, uncertainty communication, and decision reliability
Mechanism. Authority delegated to automation without matched communication of its uncertainty or the boundaries of its validated domain produces overreliance: the human stops independently

verifying because the system's confidence is not visibly bounded, and errors that a well-informed human would have caught pass through uncorrected.

Boundary conditions. The proposition predicts a conditional, not a universal, effect: authority itself is not proposed to weaken reliability where uncertainty and domain boundaries are clearly communicated, and could plausibly improve reliability in that condition by reducing unnecessary human intervention.

Competing explanations. If the harm is fully explained by the interaction and automation authority shows no effect at good communication quality, the finding supports the interaction-only account; if authority shows a negative main effect even under good communication, that would instead point to a distinct problem with delegated authority itself, which P6 as stated does not predict.

Empirical test. A field comparison crossing sites with high versus low delegated automation authority against independently rated uncertainty-communication quality, measuring override-appropriateness and error rate, with the interaction term as the primary estimate of interest.

7.7 P7: Cybersecurity control maturity as a moderator of information integrity

Mechanism. Information-integrity indicators such as schema-error rate or timestamp accuracy describe data quality under benign operating conditions; they do not by themselves indicate whether that integrity would hold under deliberate manipulation. Cybersecurity control maturity is what extends integrity's benefit to decision reliability into the adversarial case, so the relationship between integrity and reliability should be stronger where cybersecurity controls are more mature.

Boundary conditions. The moderation should be strongest where a plausible manipulation threat exists and weakest in closed or air-gapped systems where that threat is not credible, in which case information integrity could show a strong direct effect on reliability with little role for cybersecurity maturity.

Competing explanations. Cybersecurity control maturity could affect decision reliability directly, for instance through faster incident response, independent of any interaction with information integrity; a test of P7 needs to estimate the interaction term net of this main effect.

Empirical test. A cross-sectional audit study across multiple sites measuring information-integrity indicators, an independently scored cybersecurity-maturity index, and decision-reliability outcomes, testing the integrity-by-maturity interaction term.

7.8 P8: Integration quality and time to resolution

Mechanism. Integration quality reduces resolution time because the evidence and authority paths a response needs already exist before an event occurs; without them, the same evidence and authority must be assembled ad hoc during the event, which is the dominant source of delay in a poorly integrated system.

Boundary conditions. The effect should weaken for event types the pre-arranged paths did not anticipate, since novel event categories require the same ad hoc assembly regardless of integration quality for previously characterized cases; it should be strongest for recurring, well-characterized cause classes.

Competing explanations. Staffing levels and organizational size could independently explain faster resolution without any change in integration quality; a test of P8 must control for these.

Empirical test. A process-mining study measuring time from detection to verified closure across a case population, using integration-quality indicators (identity consistency, trace completeness) as the predictor of interest and staffing level as a covariate.

7.9 P9: Assurance continuity under distribution shift

Mechanism. An initially valid model's performance decays as operating conditions shift away from its training and validation domain; assurance continuity practices, drift detection, scheduled revalidation, and rollback capability, catch and correct that decay before it reaches decision reliability, so the long-term benefit of initial validity depends on whether continuity practices are in place, not on initial validity alone.

Boundary conditions. The moderation matters most in non-stationary operating environments (seasonal variation, changing fleet composition, evolving threat patterns) and matters comparatively little in genuinely stable operating envelopes, where initial validity alone may predict long-term reliability adequately.

Competing explanations. An inherently robust model architecture could substitute for active continuity practice, showing stable long-term reliability without drift detection or revalidation; a test of P9 should include model architecture as a covariate rather than attributing all stability to continuity practice.

Empirical test. A longitudinal study tracking model validity and decision reliability over time at sites that do and do not actively practice assurance continuity, testing whether the slope of reliability over time differs between the two groups under comparable distribution shift.

7.10 P10: Independent multimodal evidence and ambiguity

Mechanism. In a low-ambiguity case, a single modality already resolves the diagnosis, so an additional independent modality is redundant rather than resolving; in a high-ambiguity case the competing hypotheses would otherwise be indistinguishable, and an independent modality shifts the likelihood ratio in a way that measurably changes the diagnosis. The benefit of independent evidence is therefore proposed to be concentrated in high-ambiguity cases rather than uniform across all cases.

Boundary conditions. The benefit requires genuine independence between modalities, meaning uncorrelated failure modes; two sensors that tend to fail together for the same underlying reason provide little separability benefit even in high-ambiguity cases, which is a boundary condition distinct from the ambiguity level itself.

Competing explanations. A general "more evidence is always better" account predicts a uniform benefit across ambiguity levels rather than the concentrated benefit P10 predicts; the two accounts are distinguishable only by stratifying on pre-registered ambiguity level before observing outcomes.

Empirical test. A stratified field or simulation study that pre-registers case ambiguity level using evidence available before diagnosis, then compares false-diagnosis rate with and without an added independent modality separately within each ambiguity stratum.

7.11 P11: Joint versus separate safety-security governance

Mechanism. When physical and cyber causes share observable symptoms, separate safety and security governance structures can each reasonably conclude the event lies outside their own scope

and defer to the other, creating an adjudication gap during which risk persists unmanaged. A joint governance structure removes that gap by requiring a single adjudication regardless of eventual cause.

Boundary conditions. The advantage of joint governance should be largest exactly where symptom overlap between cyber and physical causes is highest and should shrink toward zero where the two are easily distinguished from the outset, in which case separate governance can adjudicate each correctly without coordination cost.

Competing explanations. If incident-response speed is already fast under separate governance because of strong informal coordination between teams, the formal governance structure may matter less than P11 predicts; a test should measure actual adjudication delay rather than assume it from the governance chart alone.

Empirical test. A comparative case study or natural experiment across organizations with joint versus separate safety-security governance, measuring adjudication delay and residual-risk outcomes specifically for incidents that were symptom-ambiguous at the time of detection.

7.12 P12: Fallback design as a condition for efficiency gains

Mechanism. Efficiency gains from AI-driven optimization, fewer inspections, less unscheduled downtime, are real during normal operation, but if the system lacks a validated fallback for the periods when data, model, or platform fail, that efficiency is purchased by an unmanaged risk spike during exactly those periods. This is risk transfer rather than risk reduction, and P12 predicts the efficiency-reliability relationship holds only where fallback design prevents that transfer.

Boundary conditions. The proposition is strict where failure is rare but not negligible; if a system's failure probability were genuinely zero over the relevant horizon, fallback design would matter less, but that probability is rarely knowable in advance, which is itself part of the argument for building fallback capability regardless of the estimated failure rate.

Competing explanations. Efficiency and reliability might simply trade off along a Pareto frontier with no fallback-contingent interaction, in which case fallback design would show no effect on the relationship between decision reliability and operational efficiency once baseline efficiency is held constant.

Empirical test. A scenario-based simulation comparing operational-efficiency and risk-transfer indicators (for example, value-at-risk during simulated outage windows) across system designs with strong versus weak fallback capability, holding baseline efficiency fixed across scenarios.

8. DIAGNOSTIC SEPARABILITY MODEL

Table 8: Diagnostic separability model: cause classes and discriminating evidence. Source: authors' synthesis.

Cause class	Examples	Discriminating evidence
Physical degradation	Corrosion, crack growth, wear, thermal damage, photovoltaic aging	Persistent mechanism-consistent change across independent measurements
Sensor or acquisition fault	Drift, miscalibration, coupling loss, image artifact	Local inconsistency, calibration failure, modality disagreement

Cause class	Examples	Discriminating evidence
Process or human error	Wrong asset, unit, procedure, label, or disposition	Workflow anomaly, metadata conflict, sequence violation
Software or integration defect	Mapping error, stale cache, version mismatch, timing fault	Interface-specific pattern, reproducibility under configuration
Malicious manipulation	Spoofing, poisoning, privilege abuse, record alteration	Provenance break, security telemetry, coordinated inconsistency

The taxonomy in Table 8 is necessary but not sufficient for diagnostic separability; a list of cause classes and typical evidence does not by itself specify how competing hypotheses should be weighed or when a decision should be made. This section formalizes that step as a likelihood-based classification problem, consistent with how Section 5 defines diagnostic separability as a measurable construct rather than a taxonomy.

8.1 Formal structure

Let C = {c1, …, c5} denote the five cause classes in Table 8, treated as mutually exclusive single-cause hypotheses in the base case (Section 8.4 relaxes this for genuinely multi-cause events). Let E = (e1, …, ek) denote evidence drawn from k available channels, for example an independent sensor reading, a provenance record, a security-telemetry feed, and a workflow log for the same finding. Each channel carries a likelihood P(ei | c), the probability of observing that specific evidence if cause c were true; Table 8's "discriminating evidence" column is, in this formulation, a qualitative statement of which likelihoods are high or low for each cause. Given a prior P(c) reflecting the base rate of each cause class for the specific asset type and context, the posterior probability of each cause given the full evidence set is

P(c | E) is proportional to P(c) multiplied by the product over i of P(ei | c),

under the assumption that the evidence channels are conditionally independent given the cause. That assumption is explicit and testable rather than implicit: it fails when two evidence channels share a common failure mode, for example two sensors on the same power bus that can both be spoofed by the same intrusion, and Proposition 10's boundary condition on genuine modality independence is exactly this assumption stated as an empirical requirement.

8.2 Decision rule

AIIF does not recommend acting on the single most probable cause without regard to consequence. Instead, an action a is chosen to minimize the consequence-weighted expected cost, summing over every cause the posterior probability of that cause multiplied by the cost of taking action a if that cause were true. This is what Table 8's action thresholds "tied to consequence" mean formally: the same posterior distribution over causes can justify different actions depending on how asymmetric the costs are. A posterior that is only moderately concentrated on physical degradation still justifies conservative operational restriction if the cost of missing true degradation is high, while the same degree of concentration on a sensor fault may justify continued monitoring if the cost of an unnecessary inspection is the only downside.

8.3 Measurable outcomes

Four outcomes operationalize diagnostic separability for the measurement model in Section 11. Class-specific sensitivity and specificity are the probability of correctly deciding cause c given

that c is the true cause, and the probability of correctly rejecting c given that c is not the true cause, estimated separately for each of the five classes rather than pooled into one accuracy figure, since the classes carry different consequences. Misclassification cost is the consequence-weighted sum across all cause pairs where the decided cause differs from the true cause, which is the quantity Section 8.2's decision rule is built to minimize prospectively. Time to correct causal classification is the elapsed time from first evidence to the posterior concentrating above a defined confidence threshold on the true cause, which is separability's contribution to Proposition 8's resolution-time outcome. The value of independent evidence is the expected reduction in posterior uncertainty, operationalized as the expected drop in posterior entropy, from adding one more genuinely independent evidence channel, which gives Proposition 10 a specific quantity to estimate rather than a qualitative claim that more evidence helps.

8.4 Assumptions and ambiguous or multi-cause events

Three assumptions make Sections 8.1 and 8.2 usable and each has a stated failure mode. Conditional independence of evidence channels given the cause can fail under correlated sensor or system failures, as noted above; where it is suspected to fail, the framework requires treating the correlated channels as one combined channel rather than as two independent ones, which reduces the effective value of that evidence rather than double-counting it. The likelihood functions P(ei | c) are assumed to be known or estimable from historical or deliberately injected multi-cause data; where a cause class has no labeled history at a given site, the framework requires falling back to the qualitative discriminating-evidence pattern in Table 8 and explicitly flagging the resulting decision as evidence-limited rather than presenting a numerical posterior with unwarranted precision. Cause classes are assumed mutually exclusive in the base case, which does not hold for every real event: a corroded fitting can also carry a genuine sensor drift, and an insider can exploit a genuine process gap. For these cases, AIIF extends C to include compound-cause hypotheses (for example, concurrent physical degradation and sensor drift) and compares single-cause against compound-cause hypotheses using their relative posterior support rather than forcing a single-class decision. Where no hypothesis, single or compound, achieves posterior concentration above the confidence threshold defined for the applicable consequence tier, the framework treats that state itself as the decision, not as a failure of the model: high-consequence ambiguity triggers conservative operation and parallel engineering-security investigation to acquire additional independent evidence, and low-consequence ambiguity triggers additional low-cost sensing or monitored continuation, exactly the two dispositions Table 8's cause classes were originally described against, now derived from the decision rule rather than asserted separately for each class.

9. CROSS-SECTOR MECHANISM INTEGRATION

Table 9: Cross-sector mechanisms and their aviation relevance. Source: authors' synthesis.

Evidence source	AIIF contribution
Atta *et al.* (2018)	Mechanism and context dependence of corrosion control; supports physical-state variables and treatment-boundary logic.
Atta *et al.* (2020)	Risk-based NDT and integrity management; supports inspection planning, criticality, and evidence fusion.
Atta *et al.* (2021)	Advanced NDT for weld integrity; supports multimodal sensing, method limits, and digital defect analysis.

Evidence source	AIIF contribution
Atta and Adenuga (2022)	Inspection standards and quality assurance; supports traceable disposition and controlled records.
Atta *et al.* (2024)	Charge transport and recombination; supports physics-informed diagnosis for thin-film airport assets and sensors.
Atta *et al.* (2025)	Environmental photovoltaic degradation; supports climate-aware monitoring and remaining-life analysis.
Atta (2025)	Spray-pyrolysis mechanisms; supports manufacturing provenance and process-to-performance models.
Atta (2026)	GeS heterojunction instability; supports interface-aware uncertainty and nanoscale sensor reliability.

Every source in Table 9 is used the same way, and that use is stated once here rather than repeated after each row. None of these sources by itself validates an aircraft application; each supplies constructs, variables, or failure mechanisms from adjacent materials and energy-system research for empirical testing under aviation-specific materials, standards, operating envelopes, and certification requirements. What differs across the eight sources is which part of the framework each one grounds. The corrosion and NDT series (Atta *et al.*, 2018, 2020, 2021; Atta & Adenuga, 2022) grounds Layer 1 and Layer 2 in Section 6 and the physical-degradation and sensor-fault cause classes in Section 8. The photovoltaic and thin-film series (Atta *et al.*, 2024, 2025; Atta, 2025, 2026) grounds Layer 4's hybrid-modeling illustration and the framework's claim, made in Section 1, that airport energy and sensing assets sit inside the same governance boundary as flight-critical systems rather than outside it. Treating these sources as analogical evidence rather than as aviation findings is a deliberate limit on what Table 9 claims, and Section 15 restates that limit for the framework as a whole.

10. IMPLEMENTATION PROCESS

Table 10 summarizes the eight-step implementation process at a glance; Sections 10.1 through 10.8 develop each step's entry criteria, exit criteria, accountable role, required evidence, approval authority, and rollback condition, and Section 10.9 crosswalks the steps against the architecture and constructs.

Table 10: AIIF implementation process. Source: authors' synthesis.

Step	Required output
Define the decision	State the user, asset, timing, authority, consequence, baseline, and fallback.
Map hazards and threats	Develop joint physical, safety, cybersecurity, and human-factor scenarios.
Establish evidence requirements	Select sensors, NDT, records, provenance, labels, and independent confirmation.
Build governed data products	Control identity, semantics, quality, access, configuration, retention, and lineage.
Develop hybrid models	Combine rules, physics, statistics, uncertainty, and abstention according to the decision.

Step	Required output
Validate technically and operationally	Use temporal and external testing, simulations, challenge cases, and human-in-the-loop trials.
Deploy with bounded authority	Stage release, restrict actions, monitor interfaces, and retain validated fallback.
Maintain continuous assurance	Monitor drift, incidents, overrides, outcomes, supplier changes, and update impacts.

10.1 Define the decision

Entry criteria are a proposed decision scope from the asset owner. The deliverable is a decision charter stating the user, asset, timing, authority, consequence tier, baseline performance, and fallback action. Exit criteria are the accountable decision owner's signature on that charter. The asset owner and the decision owner are accountable; a high-consequence decision additionally requires Layer 6 sign-off on the consequence-tier classification itself, since that classification sets the evidence and validation bar every later step must clear. The evidence produced is the signed charter. If the consequence tier is later found to have been misclassified, the charter reopens and every downstream step that depended on the original tier is re-reviewed against the corrected one, rather than being grandfathered in.

10.2 Map hazards and threats

Entry criteria are an approved decision charter. The deliverable is a joint hazard and threat register covering physical, safety, cybersecurity, and human-factor scenarios for that specific decision, not a generic asset-level register. Exit criteria are joint sign-off from safety and cybersecurity leads on the same register, which operationalizes Proposition 11's joint-governance requirement at the point in the process where it first matters. The evidence produced is the signed register. A hazard or threat discovered after this step reopens the register and, because Section 10.3's evidence plan is built directly from it, triggers a review of whether the evidence plan already covers the new entry.

10.3 Establish evidence requirements

Entry criteria are an approved hazard and threat register. The deliverable is an evidence plan mapping each register entry to the sensors, NDT methods, records, provenance requirements, and independent-confirmation protocol that would provide evidence about it, consistent with the Layer 1 to Layer 2 interface defined in Table 6. Exit criteria are a one-to-one mapping with no register entry left without an evidence source. Engineering and instrumentation are accountable; the evidence produced is the plan itself together with the probability-of-detection study for each proposed method. A gap discovered later returns the plan to this step rather than being absorbed informally into a later one.

10.4 Build governed data products

Entry criteria are an approved evidence plan with acquisition underway or already available. The deliverable is a set of governed data products meeting the Layer 3 controls in Section 6.3: identity resolution, schema conformance, access control, configuration management, retention policy, and lineage tracking, consistent with the Layer 2 to Layer 3 and Layer 3 to Layer 4 interface checks in Table 6. Exit criteria are passing those interface checks, not merely producing data. Data engineering, IT, and information security are accountable, and information security's approval is required before a data product may be consumed by Step 10.5, since an ungoverned data product entering model development at this stage is the specific failure mode Layer 3 exists to prevent.

10.5 Develop hybrid models

Entry criteria are governed data products in sufficient volume and quality to support the modeling approach chosen. The deliverable is a validated model or model ensemble combining rules, physics, and statistical learning as appropriate to the decision, with defined abstention behavior. Exit criteria are a validation report meeting pre-specified thresholds on the calibration, discrimination, and class-specific separability measures defined in Sections 5 and 8, not a single aggregate accuracy figure. Data science is accountable for the model; the engineer who owns the underlying physical mechanism must independently confirm physical plausibility before the model proceeds. A model that fails validation returns to this step; it does not proceed to Step 10.6 on a provisional basis.

10.6 Validate technically and operationally

Entry criteria are a technically validated model from Step 10.5. The deliverable is operational validation evidence: temporal and external holdout testing, simulation against challenge cases including injected competing-cause scenarios, and human-in-the-loop trials conducted under realistic workload rather than idealized conditions. Exit criteria are trial results showing decision reliability at or above the threshold set for the decision's consequence tier, under both normal and degraded conditions. Operations and human-factors specialists are accountable jointly, since a model that performs well in isolation but fails under realistic operator workload has not met this step's exit criteria. A failed trial returns either to Step 10.5, if the failure traces to model behavior, or to Step 10.3, if the failure traces to an evidence gap the trial exposed.

10.7 Deploy with bounded authority

Entry criteria are passed operational validation. The deliverable is a staged deployment with an explicit authority matrix (Section 6.5), monitored interfaces at each Table 6 boundary, and a rehearsed, validated fallback that has itself been exercised, not merely documented. Exit criteria for each stage are pilot-site performance within bounds before authority or scope expands to the next stage. Operations is accountable for execution; Layer 6's change-approval board must approve each authority expansion individually rather than approving the full rollout in advance. Monitored performance that degrades below threshold triggers rollback to the previous authority level or to full fallback, whichever the degradation severity requires.

10.8 Maintain continuous assurance

Entry criteria are a live deployment from Step 10.7. The deliverable is ongoing monitoring of drift, incidents, overrides, outcomes, supplier changes, and update impacts, feeding a continuously updated assurance case. This step has no single exit criterion because it does not end while the system remains deployed; instead it carries periodic review gates, at an interval set by the decision's consequence tier, at which Layer 6 reconfirms the current authority level against accumulated evidence. Layer 6 is accountable overall, drawing reports from every layer owner. A material change, a model update, a supplier change, a configuration change, does not self-approve within this step; it re-enters Step 10.5 or 10.6 as appropriate to its scope, and sustained drift or a repeated override pattern triggers an automatic authority reduction pending re-validation rather than waiting for the next scheduled review gate.

10.9 Steps, layers, constructs, and assurance artifacts

Table 11 links each implementation step to the AIIF layers it primarily operates on, the constructs it is intended to build or protect, and the assurance artifact it produces, so that the process in Sections 10.1 through 10.8 can be read against the architecture in Section 6 and the constructs in Section 5 rather than as a free-standing checklist.

Table 11: Implementation steps mapped to AIIF layers, constructs, and assurance artifacts. Source: authors' synthesis.

Step	Primary layer(s)	Primary construct(s)	Assurance artifact produced
10.1 Define the decision	Layer 1, Layer 6	Decision reliability (target definition)	Signed decision charter
10.2 Map hazards and threats	Layer 6	Assurance continuity (scope)	Joint hazard and threat register
10.3 Establish evidence requirements	Layer 1, Layer 2	Physical-state observability	Evidence plan and probability-of-detection study
10.4 Build governed data products	Layer 3	Information integrity	Governed data product with lineage record
10.5 Develop hybrid models	Layer 4	Model validity, diagnostic separability	Model validation report
10.6 Validate technically and operationally	Layer 4, Layer 5	Decision reliability, human decision readiness	Operational validation and human-in-the-loop trial report
10.7 Deploy with bounded authority	Layer 5, Layer 6	Integration quality	Authority matrix and fallback rehearsal record
10.8 Maintain continuous assurance	All layers, coordinated by Layer 6	Assurance continuity	Updated assurance case

11. MEASUREMENT AND EMPIRICAL TESTING

Table 12: Measurement model for all eight AIIF constructs. Source: authors' synthesis.

Construct	Illustrative measures
Physical-state observability	Coverage of relevant failure modes; probability of detection; independent modality count; missingness by condition.
Information integrity	Provenance completeness; schema error rate; timestamp accuracy; unauthorized change rate; recovery time.
Model validity	Calibration error; sensitivity; specificity; false-negative cost; robustness; abstention quality; external validity.
Diagnostic separability	Class-specific sensitivity and specificity by cause class; misclassification cost; time to correct causal classification; expected value of independent evidence (Section 8.3).
Human decision readiness	Competence scores; situation awareness; workload; appropriate reliance; escalation accuracy; recovery time.
Integration quality	Identity consistency across systems; interface error rate; trace completeness; closed-loop action verification; explanation-quality rating of the layer-to-layer interface.

Construct	Illustrative measures
Assurance continuity	Drift detection delay; update regression escape rate; audit completeness; rollback success rate; incident learning cycle time.
Decision reliability	Correct action rate; risk-weighted error; repeat inspection; unscheduled removal; disruption; closure quality.

Physical-state observability should be measured through failure-mode coverage, probability of detection, independent-modality count, and missingness by condition, with denominators and confidence intervals stated explicitly and errors weighted by safety consequence rather than treated uniformly. Information integrity should be measured through provenance completeness, schema-error rate, timestamp accuracy, unauthorized-change rate, and recovery time, tested against corrupted, delayed, incomplete, and misidentified records rather than against clean data alone. Model validity should combine calibration, sensitivity, specificity, false-negative cost, robustness, abstention quality, and external validation, using time-based and fleet-level holdouts rather than random splits where leakage across related records is plausible. Diagnostic separability should be measured using the four outcomes formalized in Section 8.3, estimated separately by cause class rather than pooled, since pooling would hide the classes a given deployment handles poorly. Human decision readiness should combine competence, situation awareness, workload, appropriate reliance, escalation accuracy, and recovery time, tested under normal, time-pressured, and degraded conditions in simulation. Integration quality should combine identity consistency, interface error rate, trace completeness, and closed-loop action verification with an explicit explanation-quality rating of each layer-to-layer interface, since Section 5 now carries explanation quality under this construct rather than under human decision readiness. Decision reliability should combine correct-action rate, risk-weighted error, repeat inspection, unscheduled removal, disruption, and closure quality, weighting a critical false negative more heavily than a conservative low-risk alert. Assurance continuity should combine drift-detection delay, update-regression escapes, audit completeness, rollback success, and incident-learning cycle time, with every measure linked to the specific model version and approval state it describes.

11.1 Formative and reflective specification

Where a construct combines formative and reflective indicators, as Section 5 specifies for each of the eight constructs, a structural equation model should not treat all indicators as interchangeable reflective items; doing so would misestimate both the construct's reliability and its relationships to other constructs. Formative indicators should instead be combined into a composite (for example through principal component or a defined weighting scheme informed by domain judgment) before entering the structural model, while reflective indicators support internal-consistency reliability estimation in the conventional way. Because several constructs share conceptually adjacent formative indicators (physical-state observability's independent-modality count and diagnostic separability's evidence-independence indicator both depend on the same underlying sensing architecture, for instance), a structural equation model applied to this framework should report discriminant validity explicitly, using the heterotrait-monotrait ratio or an equivalent test, before interpreting any path coefficient between adjacent constructs. Common-method bias is a further risk where survey and operational indicators are collected from the same respondent or the same system log; multi-source measurement, operational indicators wherever available rather than self-report alone, and a marker-variable or measured-latent-method-factor control are the mitigations a study using this measurement model should apply.

11.2 Proposition-to-measurement crosswalk

Table 13 links each proposition to the constructs it involves, an illustrative indicator for each, a plausible data source, the unit of analysis, and a suitable empirical design, consolidating the empirical-test paragraphs in Section 7 into a single reference table.

Table 13: Proposition-to-measurement crosswalk. Source: authors' synthesis.

ID	Constructs involved	Illustrative indicator	Data source	Unit of analysis	Empirical design
P1	Physical-state observability, model validity	Independent-modality count; calibration error	Sensing logs; model validation logs	Asset or site	Prospective multilevel study
P2	Information integrity, model validity	Schema-error rate; calibration error	Data-quality logs; model logs	Site	Quasi-experiment (pre-post)
P3	Diagnostic separability	Class-specific sensitivity by model type	Fault-injection test records	Test case	Controlled injection study
P4	Model validity, diagnostic separability, decision reliability	Calibration error; misclassification cost; correct-action rate	Model and case-outcome records	Case	Mediation analysis (bootstrapped indirect effect)
P5	Human decision readiness, decision reliability	Competence score; decision-reliability outcome	Simulation trial records	Operator-trial	Factorial simulation experiment
P6	Human decision readiness, decision reliability	Override-appropriateness rate	Decision and override logs	Site	Field comparison (2x2 design)
P7	Information integrity, decision reliability	Schema-error rate; correct-action rate	Data-quality and cybersecurity audit logs	Site	Cross-sectional audit study
P8	Integration quality	Identity consistency; time to closure	Process-mining logs	Case	Process-mining study
P9	Model validity, decision reliability	Calibration error over time	Longitudinal model-performance logs	Site over time	Longitudinal study
P10	Diagnostic separability	False-diagnosis rate by ambiguity stratum	Case records with pre-registered ambiguity rating	Case	Stratified field or simulation study
P11	Decision reliability, assurance continuity	Adjudication delay; residual-risk indicator	Incident records	Organization	Comparative case study
P12	Decision reliability	Value-at-risk during outage window	Simulation scenario records	Scenario	Scenario-based simulation

11.3 Study designs

A multi-site longitudinal design offers the strongest practical test of the framework as a whole. Operators could deploy the framework in phases across matched fleets, bases, or asset classes;

interrupted time-series or stepped-wedge designs would measure change while accommodating operational rollout, with outcomes including safety precursors, maintenance yield, delays, costs, user reliance, security incidents, and fallback performance. Controlled experiments suit the narrower relationships in Table 13 that isolate one mechanism, such as P5 and P10: researchers could manipulate explanation format, uncertainty display, alert frequency, or workload in high-fidelity simulations, and diagnostic studies could inject realistic physical faults, sensor errors, integration defects, and cyber manipulations into a test environment, with correct causal classification and action as the outcome rather than anomaly detection alone. Structural equation modeling could test the construct relationships in Table 13 using survey and operational indicators together, subject to the common-method-bias and discriminant-validity controls in Section 11.1; multilevel models should account for observations nested within assets, bases, operators, and regulatory environments. Qualitative case studies remain useful for authority, coordination, and incident learning, particularly for P11, where event counts are typically too low for a quantitative comparative design.

12. Illustrative Application Sketches and Worked Examples

The eight application settings below illustrate where AIIF applies across aviation; they are sketches, not validated case studies, and are labeled as such because none has yet been tested against fielded data. Two of the eight, airframe corrosion and structural health, and aviation cybersecurity and data integrity, are developed separately in Section 12.2 as fully worked examples that follow one plausible event through every AIIF layer with concrete evidence, a stated decision threshold, and a stated fallback, which is the level of depth the sketches below deliberately do not attempt for all eight settings at once.

12.1 Brief illustrative sketches

Across all eight settings, the common failure this framework is built to prevent is the same: a high anomaly score treated as if it were already a diagnosis. In each setting, Layer 2 verifies acquisition and inspection capability, Layer 3 binds each record to asset, time, units, calibration, and configuration, Layer 4 compares physical, fault, and attack hypotheses using the decision rule in Section 8.2, Layer 5 presents the resulting uncertainty and any recommended independent confirmation, and Layer 6 controls approval and monitoring, so that setting, cyber compromise, evidence preservation, and security containment proceed alongside engineering inspection, while continued plausibility of physical damage keeps conservative operational limits in place until qualified disposition. What differs across settings is the hazard set, the evidence base, and which of the twelve propositions is most directly at stake.

Airframe corrosion and structural health. Hazard set: humidity, salts, dissimilar-metal couples, coatings, joints, and hidden moisture paths. Evidence base: electrochemical probes, thickness readings, eddy-current signals, maintenance findings, and environmental exposure records (Atta *et al.*, 2018, 2020). This setting is the basis for the first worked example in Section 12.2.

Weld and bonded-joint inspection. Hazard set: manufacturing discontinuities, fatigue cracks, porosity, lack of fusion, debonding, and repair quality. Evidence base: ultrasonic, radiographic, thermographic, acoustic, visual, and digital inspection records (Atta *et al.*, 2021; Atta & Adenuga, 2022). Diagnostic separability here rests mainly on modality disagreement among method-specific findings, the P3 mechanism most directly.

Engine and rotating-component maintenance. Hazard set: thermal cycling, vibration, wear, creep, lubrication changes, and foreign-object damage. Evidence base: engine health monitoring, oil

debris, vibration spectra, borescope images, trend data, and maintenance actions (Jardine *et al.*, 2006; Lei *et al.*, 2018). Integration quality across vendor systems, the Layer 3 concern in Section 6.3, is this setting's binding constraint more often than sensing coverage itself.

Avionics and software assurance. Hazard set: sensor faults, software defects, configuration errors, timing anomalies, corrupted interfaces, and malicious inputs. Evidence base: flight data, built-in tests, software logs, network telemetry, configuration baselines, and assurance evidence (RTCA, 2011, 2014). The time-critical Layer 5 constraint described in Section 6.5 dominates this setting; there is little time for extended explanation before a decision is needed.

Airport and airline operations. Hazard set: turnaround variability, congestion, weather disruption, baggage failures, staffing constraints, and cascading delays. Evidence base: operational databases, surveillance feeds, schedules, passenger flows, weather services, and resource status (ICAO, 2018; EASA, 2023). This setting is the framework's clearest test of Proposition 12: efficiency gains here are only as good as the fallback available when the operational data feed itself degrades.

Air traffic and trajectory services. Hazard set: demand-capacity imbalance, uncertainty, loss-of-separation precursors, communication faults, and automation surprise. Evidence base: surveillance tracks, flight plans, controller inputs, weather products, and decision-support outputs (Endsley, 1995; Parasuraman *et al.*, 2000). Proposition 6's concern, automation authority outrunning communicated uncertainty, is this setting's central risk.

Aviation cybersecurity and data integrity. Hazard set: false-data injection, spoofing, ransomware, insider misuse, supply-chain compromise, and model poisoning. Evidence base: identity events, network flows, signatures, logs, provenance records, model versions, and security alerts (ICAO, 2019; NIST, 2024). This setting is the basis for the second worked example in Section 12.2.

Airport renewable energy and sensing infrastructure. Hazard set: module corrosion, recombination loss, thin-film nonuniformity, nanowire instability, inverter faults, and manipulated telemetry. Evidence base: irradiance, current-voltage traces, electroluminescence images, climate exposure, deposition records, and security logs (Atta *et al.*, 2024, 2025; Atta, 2025, 2026). This setting illustrates Layer 4's hybrid physics-data modeling requirement most directly, as Section 6.4 develops.

Across all eight, success is measured by risk-weighted diagnostic accuracy, time to correct action, inspection yield, user-reliance quality, and safe recovery after degraded data or service, with cost and availability considered only after safety constraints are met; the framework predicts that integration quality shortens resolution time because the required evidence and authority paths are prearranged rather than assembled during the event itself, which is Proposition 8 stated once here rather than repeated after each sketch.

12.2 Worked example 1: airframe corrosion and structural health

Input evidence. A scheduled eddy-current inspection on a wing-skin lap joint returns a thickness estimate 6 percent below the previous reading, exceeding the site's routine variation band. The asset's Layer 1 record shows the joint sits in a documented moisture-trap location with a known dissimilar-metal couple, consistent with a plausible corrosion mechanism.

Competing hypotheses. Physical degradation (accelerated corrosion at this joint) competes against a sensor or acquisition fault (probe coupling loss, common at this joint geometry) and, given the joint's location near an access panel serviced two days earlier, a process or human error (incorrect

probe reference calibration after panel reinstallation). Malicious manipulation is assigned a low prior for this asset class and access profile but is not excluded outright.

Model output and uncertainty. The Layer 4 hybrid model, combining a corrosion-rate physics model with the acquisition-pattern classifier described in Section 6.1, returns a posterior concentrating 58 percent on physical degradation, 31 percent on sensor or acquisition fault, and 11 percent split across the remaining classes, with the two leading hypotheses separated by a log-likelihood ratio below the site's confidence threshold for this consequence tier. Under Section 8.4, this is an ambiguous case, not a resolved one.

Independent evidence requested. Following the high-consequence disposition in Section 8.4, the system requests a second, independent reading using an ultrasonic thickness gauge rather than a repeat eddy-current reading, because a repeat reading with the same instrument would not be conditionally independent of the first under Section 8.1's assumption and so would not usefully update the posterior.

Decision threshold and responsible authority. The site's evidence plan (Section 10.3) sets the action threshold for this consequence tier at 85 percent posterior concentration on a single cause before returning the aircraft to unrestricted service. The duty structures engineer, acting within the Layer 5 authority matrix, is the responsible authority for the interim restriction; escalation to the safety-security governance board (Layer 6) is required only if the independent reading fails to resolve the ambiguity.

Fallback action. Pending the independent reading, the aircraft is restricted to a reduced inspection interval on that joint rather than grounded outright, consistent with the moderate, not high, immediate consequence of a joint already inside its corrosion allowance.

Closure evidence. The ultrasonic reading confirms the eddy-current finding within instrument tolerance, moving the posterior to 93 percent concentration on physical degradation. The joint is scheduled for disposition under the site's corrosion-control program, the interim restriction is lifted, and the case record, including both readings, the posterior trajectory, and the authority decision, is retained as the assurance-case evidence Section 6.6 requires.

12.3 Worked example 2: aviation cybersecurity and data integrity

Input evidence. An airport ground-systems monitoring feed reports a sequence of sensor readings from a runway surface-condition sensor array that are individually plausible but collectively inconsistent with the concurrent weather-service report and with an adjacent, independently powered sensor of the same type.

Competing hypotheses. A genuine, highly localized surface condition affecting only one sensor's footprint competes against a sensor or acquisition fault (a known failure mode for this sensor model under specific temperature ranges) and against false-data injection targeting the monitoring feed, motivated by the operational disruption a false runway-condition alert would cause.

Model output and uncertainty. Layer 4 combines the physical plausibility check (is the reported condition consistent with the concurrent weather report and adjacent sensor) with the security-telemetry channel from Layer 6. The physical-plausibility check alone would favor a sensor fault; adding the security-telemetry channel, which flags an unusual authentication pattern on the feed's ingestion gateway in the same window, shifts the posterior toward false-data injection, illustrating

Proposition 10's prediction that independent evidence has its largest effect exactly in a case this ambiguous.

Independent evidence requested. The joint safety-security governance function (Section 6.6) requests network-forensic evidence on the ingestion gateway rather than relying on the sensor readings alone, since sensor-side evidence cannot distinguish a fault from an injection at the gateway.

Decision threshold and responsible authority. Because a false runway-condition alert and a suppressed genuine one carry materially different but both significant safety consequences, the applicable consequence tier requires joint sign-off from safety and cybersecurity leads, operationalizing Proposition 11, before either accepting the reading or overriding it. Containment (isolating the affected feed) is evaluated against the specific constraint named in Section 6.6: isolating a runway-condition feed during active operations creates its own availability risk, so containment proceeds only alongside a validated manual reporting fallback for that runway segment, not before it is in place.

Fallback action. The affected sensor's readings are flagged as unconfirmed and excluded from the automated advisory pending investigation; the runway remains in service under the manual reporting fallback rather than being closed on unconfirmed evidence, reflecting the consequence-weighted decision rule in Section 8.2 rather than a default toward the most conservative possible action.

Closure evidence. Network-forensic analysis confirms an authentication anomaly consistent with credential misuse rather than a sophisticated data-poisoning attempt, the affected credential is revoked, the ingestion gateway is patched, and the sensor array is independently re-verified before the automated feed resumes. The full evidence chain, physical, sensor, and security telemetry together with the joint governance decision, is retained in the assurance case, and the incident enters the Layer 6 incident-learning process to check whether the same authentication weakness exists on comparable gateways elsewhere in the network.

13. Governance, Ethics, and Organizational Design

AIIF assigns accountability through decision rights rather than through a generic responsible-AI statement. The asset owner defines performance needs; engineering defines technical acceptability; safety accepts residual safety risk; cybersecurity manages threat risk; IT maintains service controls; quality verifies process conformity; operations owns real-time action within approved limits. Data scientists build and monitor models but do not accept operational risk on behalf of accountable functions, which is the specific allocation problem Layer 6 in Section 6.6 is built to enforce.

This allocation is not novel to aviation; it extends an existing line of aviation safety-management research on organizational capability rather than treating governance as a software configuration choice. Adaptive safety-management research frames aerodrome implementation as an organizational capability question, not an installation exercise (Adeyelu, 2019), and predictive compliance monitoring that links complaint data to systemic safety indicators shows the same evidence a governance function like Layer 6 needs already exists in aviation practice, if fragmented across systems that do not currently talk to each other (Adeyelu, 2018). Maturity-based evaluation of airport safety audits makes the same point from a different angle: a model that predicts audit outcomes should remain subordinate to documented evidence, inspector judgment, corrective-

action status, and independent verification, never replacing direct testing of critical safeguards (Adeyelu & Dagodzo, 2024), which is the reason Section 6.6 defines Layer 6's evidence as audit reports and change-approval records rather than a predictive score. Risk-tiered oversight research on unmanned aircraft operations offers a transferable allocation principle for a mixed-authority environment: surveillance and regulatory attention should follow consequence, exposure, and control maturity rather than treating every operation as equal (Adeyelu, 2023), a principle Section 10.8's consequence-tiered review gates apply directly to AI-supported decisions rather than to aircraft operations.

Ethical governance includes transparency, competence, proportionality, fairness, and contestability. Passenger-facing systems require privacy and nondiscrimination analysis. Workforce systems require protection against opaque surveillance or disciplinary use. Safety-critical systems require truthful uncertainty and accessible challenge. EASA's human-centric approach and the NIST AI RMF support governance across the lifecycle (EASA, 2023; NIST, 2023).

Supplier management must cover datasets, software dependencies, model artifacts, hosting, security, support, and update policies. Contracts should grant audit evidence, vulnerability-disclosure, incident-cooperation, portability, and termination support. Supply-chain risk is highest when a single supplier controls both the model and the only available interpretation of its performance, which is why Section 6.6 requires supplier oversight to sit inside the same governance function as safety and security risk acceptance rather than as a separate procurement control.

Assurance cases should link claims to evidence. A top claim might state that the AI-supported decision keeps risk within accepted bounds for a defined operating domain. Subclaims address data, models, humans, cybersecurity, integration, fallback, and monitoring. Evidence includes tests, analyses, training records, process audits, operational trials, and incident exercises, and version control preserves the relationship between a claim and the evidence that supported it after change, which is the specific record Table 6's Layer 6 feedback boundary requires before a change takes effect.

Organizational learning depends on a just reporting culture. Users should report confusing explanations, false alerts, workarounds, and unsafe recommendations without fear when their conduct was responsible, and reports become structured evidence for model and workflow improvement rather than being absorbed informally. Governance should prevent raw feedback from changing production behavior without review, which is the same discipline Section 10.8 applies to drift and override patterns: a signal that authority should be reduced is acted on through the defined review gate, not by an individual operator's informal workaround becoming the new default.

14. Related Evidence Supporting Specific Constructs and Layers

This section states, for each retained cross-sector or foundational source not already placed in Sections 4, 6, 9, or 13, which specific AIIF construct, layer, or proposition it supports. Sources considered during the synthesis but not tied to a specific claim in this manuscript were excluded under the screening test in Section 2.3, consistent with the source-pool disposition reported in Table 1; this section is deliberately short because that filter, not a change in topic, is what removed the bulk of the originally assembled cross-sector literature.

Layer 3's information-integrity and configuration controls draw on cloud-configuration governance research showing how approved baselines, automated checks, and reviewable changes are maintained at scale (Aliliele *et al.*, 2023), and on connected-infrastructure research showing that airport surface systems and operational data exchanges need the same asset inventory, authenticated telemetry, and segmentation that shared IoT-driven transportation infrastructure requires (Jimoh *et al.*, 2023). Because aviation systems are long-lived relative to the cryptographic algorithms protecting them, Layer 3's configuration-management control should anticipate the kind of crypto-agile migration capability documented for enterprise systems more broadly, the ability to replace an algorithm or protocol without redesigning every dependent service (Nwakamma *et al.*, 2026).

Layer 6's threat model for the malicious-manipulation cause class in Section 8 draws on adversarial-machine-learning threat modeling for AI-enabled operational-technology systems (Adebayo *et al.*, 2022) and on zero-trust architecture research for operational technology, which grounds the framework's requirement that a Layer 6 containment action be evaluated against the availability risk it could itself create (Ahmed *et al.*, 2021). An AI-driven security-operations architecture integrating threat intelligence, behavioral analytics, and automated response illustrates the kind of monitoring capability Layer 6 draws on for its own detection function (Adegbite *et al.*, 2024), and cyber-risk quantification research supports the consequence-weighting that Section 8.2's decision rule requires, provided the aviation safety consequence is modeled explicitly rather than folded into a generic financial or technical score (Adegbite *et al.*, 2025). Incident-response and digital-forensics research grounds the network-forensic evidence step in Worked Example 2 (Dosunmu & Ogundele, 2021), and security-awareness research linking training to insider-threat behavior supports Layer 6's people-and-process control, the element of governance that a purely technical control set cannot cover on its own (Onche *et al.*, 2024).

Layer 4's hybrid physics-data modeling requirement, formalized as Proposition 3 in Section 7.3, is supported by physics-informed neural network research showing that combining physical constraints with learned components improves defect and internal-structure analysis over either approach alone (Zhang *et al.*, 2022). Digital-twin terminology used in Section 6.4 follows the established distinction between a digital model, a digital shadow, and a digital twin by the direction and automation of data exchange with the physical asset (Kritzinger *et al.*, 2018; Tao *et al.*, 2019), and a closed-loop digital-twin architecture for real-time process control in deposition equipment illustrates the same physics-data integration this paper's photovoltaic and thin-film examples draw on (Olodo *et al.*, 2025). Layer 2's structural-sensing requirement draws on a machine-learning overview of structural health monitoring that identifies sensor placement, environmental compensation, and independent confirmation as the recurring design concerns AIIF's Layer 2 controls in Section 6.2 are built to address (Malekloo *et al.*, 2021), and the aleatoric-epistemic uncertainty distinction used throughout Section 8's formalization follows established practice for separating irreducible measurement noise from model uncertainty that more data could resolve (Hüllermeier & Waegeman, 2021).

Layer 5's bounded local authority and validated fallback, developed in Section 6.5 and formalized as Proposition 12 in Section 7.12, has a transferable design pattern in edge-intelligence research for resilient microgrid control when central services degrade (Komi & Ganiu, 2023), and in general human-in-the-loop research on iterative collaboration between people and learning systems (Ladapo *et al.*, 2022), which supports treating the human decision step as an active control function

rather than a passive recipient of model output, consistent with Section 6.5's failure-mode discussion.

Layer 6's organizational design, developed in Section 13, follows responsible-AI governance research that allocates oversight, risk, documentation, and accountability across roles rather than concentrating them in a single function (Ominyi & Anichukwueze, 2024), and Layer 6's incident-learning requirement in Section 10.8 follows safety research on structured incident investigation and prevention-oriented organizational learning (Obriki & Arumosoye, 2024), extended here from industrial maintenance environments to the AI-supported decision context this framework addresses.

15. DISCUSSION

AIIF shifts the unit of analysis from model accuracy to decision reliability. This shift integrates technical and organizational factors without reducing one to the other: it explains why a statistically strong model may fail in operations when asset identity, provenance, user authority, or fallback is weak, and why a modest model may deliver value when embedded in disciplined inspection and decision processes.

Diagnostic separability is the framework's central and, per the comparison in Section 3, its most novel contribution. Aviation systems need to distinguish degradation, sensor fault, process error, software defect, and attack because each requires a different response, and Section 8's formalization gives that requirement a decision rule rather than leaving it as a taxonomy. Independent modalities, provenance, security telemetry, and causal constraints improve diagnosis when symptoms overlap, precisely because they add conditionally independent evidence to the posterior computation in Section 8.1, not because more evidence is generically better.

The framework connects safety and efficiency through bounded authority rather than treating them as competing objectives. AI may reduce unnecessary inspections, delays, and component removals, but Proposition 12 makes the condition explicit: efficiency gains count only when risk stays within approved limits, and fallback design is what keeps that ordering intact when confidence or integrity falls.

15.1 Limitations

AIIF is a conceptual synthesis and every claim in Section 7 is a hypothesis, not a demonstrated effect; Table 13 exists precisely because none of the twelve propositions has yet been tested against fielded aviation data. Constructs may interact differently across maintenance, air traffic, airports, and passenger services than the causal chain in Figure 2 assumes, and the framework has not been tested for whether that chain holds, or holds with the same strength, across those functions. Regulatory jurisdictions differ in ways this synthesis has not attempted to map exhaustively. Some measures in Table 12 need proprietary operational data that may not be available to independent researchers, which is a practical constraint on testing the framework, not a limitation of the constructs themselves. Future studies should test parsimonious subsets of the model, for example the P1 through P4 chain alone, before attempting to estimate the complete twelve-proposition structure, since a partial test that succeeds gives more confidence than a complete test that cannot be adequately powered. The materials, inspection, and energy-system research the framework draws on for mechanism-specific grounding (Section 9) remains transfer evidence, not aviation validation, and every AIIF construct that touches airworthiness or certification remains subordinate to the standards compared in Section 3, not a substitute for them.

16. Future Research

Researchers should build datasets containing paired physical faults, acquisition faults, process errors, software defects, and cyber manipulations. Current datasets often label all abnormal behavior as one class, which is precisely what prevents the causal evaluation Section 8 requires; testbeds should preserve realistic protocols, sensor physics, maintenance workflows, and user decisions rather than synthetic anomalies alone.

Physics-informed learning deserves comparative trials directly addressing Proposition 3: studies should compare unconstrained models, soft constraints, hard constraints, and digital-twin residual methods across sparse failures and changing fleets, with calibration, extrapolation, adversarial behavior, and sensitivity to incorrect physics as explicit outcomes rather than aggregate accuracy alone.

Federated learning offers multi-operator scale without centralizing raw data, yet heterogeneity and poisoning create risk that Section 8's malicious-manipulation cause class must extend to cover. Research should test secure aggregation, local validation, contribution weighting, rollback, and operator-specific accountability; a global model should not silently weaken safety for a minority fleet, which is a distributional concern current federated-learning evaluation practice does not typically report.

Human-factors research should examine explanation timing and role directly, testing Proposition 5 rather than assuming it. Maintenance, flight operations, air traffic, security, and management require different detail and different time budgets, as Section 6.5's avionics illustration makes concrete; experiments should measure appropriate reliance, diagnostic accuracy, workload, and recovery rather than user preference alone.

Continuous assurance needs practical tooling that operationalizes Section 10.8 rather than leaving it as a procedural description. Research should automate trace links among hazards, requirements, datasets, models, tests, incidents, and releases while keeping accountable review in the loop; the goal is faster evidence maintenance, not automatic risk acceptance, a distinction Section 6.6 treats as a governance boundary rather than an implementation detail.

17. CONCLUSIONS

The Aviation Intelligence and Integrity Framework integrates physical condition, trusted sensing, governed information infrastructure, hybrid AI, human authority, and safety-security governance into eight constructs, six layers, and twelve testable propositions, with a formalized diagnostic separability model and an eight-step implementation process. Its central claim is a hypothesis, not a finding: that aviation AI is more likely to support reliable decisions where organizations preserve physical observability, information integrity, model validity, human readiness, integration quality, and assurance continuity together, rather than any one of them alone. Corrosion, NDT, welding, and photovoltaic-degradation research provide mechanism-based transfer evidence for selected aviation assets and sensors, not aviation validation in themselves. Future empirical work should test the propositions in Table 13, beginning with the more tractable subsets identified in Section 15.1, across operators and functions before the complete model is evaluated as a whole. Practitioners may use the implementation process in Section 10 to structure pilots, assurance cases,

supplier requirements, deployment gates, monitoring, and incident exercises, with the explicit understanding that the framework keeps operational efficiency subordinate to bounded risk rather than treating the two as automatically aligned.

List of abbreviations

AI, artificial intelligence; AIIF, Aviation Intelligence and Integrity Framework; IT, information technology; NDT, nondestructive testing; UAV, unmanned aerial vehicle.

REFERENCES

1. Adebayo, A., Adegbite, M. P., & Ahmed, M. O. (2022). Adversarial machine learning in critical infrastructure: A conceptual framework for threat modeling AI-enabled OT systems. World Journal of Innovation and Modern Technology, 6(1), 184-234. https://doi.org/10.56201/wjimt.v6.no1.2022.pg184.234
2. Adegbite, M. P., Adebayo, A., & Ahmed, M. O. (2024). An AI-driven security operations architecture for utility-sector SOCs: Integrating threat intelligence, behavioral analytics, and automated response. International Journal of Engineering and Modern Technology, 10(11), 197-256. https://doi.org/10.56201/ijemt.v10.no11.2024.pg197.256
3. Adegbite, M. P., Adebayo, A., & Ahmed, M. O. (2025). A cyber risk quantification and governance architecture for critical infrastructure: From posture measurement to executive reporting. International Journal of Computer Science and Mathematical Theory, 11(12), 232-293. https://doi.org/10.56201/ijcsmt.vol.11.no12.2025.pg232.293
4. Adeyelu, O. O. (2018). A predictive compliance monitoring framework for detecting systemic safety risks through aviation consumer complaint data. Iconic Research and Engineering Journals, 1(8), 250-276. https://doi.org/10.64388/IREV1I8-1718478
5. Adeyelu, O. O. (2019). An adaptive safety management system implementation model for aerodromes in developing economy contexts. Iconic Research and Engineering Journals, 3(4), 628-654. https://doi.org/10.64388/IREV3I4-1718479
6. Adeyelu, O. O. (2023). A risk-tiered oversight model for unmanned aircraft operations in shared controlled airspace over Nigerian airports. International Journal of Scientific Research in Civil Engineering, 7(4), 147-184. https://doi.org/10.32628/IJSRCE237419
7. Adeyelu, O. O., & Dagodzo, D. (2024). A maturity model for predicting airport safety audit outcomes in resource-constrained regulatory environments. International Journal of Scientific Research in Civil Engineering, 8(4), 132-170. https://doi.org/10.32628/IJSRCE248423
8. Ahmed, M. O., Adegbite, M. P., & Adebayo, A. (2021). Zero trust architecture for operational technology in North American critical infrastructure: A framework for implementation and resilience optimization. International Journal of Engineering and Modern Technology, 7(1), 66-113. https://doi.org/10.56201/ijemt.vol.7.no1.2021.pg66.113
9. Aliliele, C., Mbonu, I. S., & Iwuanyanwu, U. (2023). A conceptual framework for continuous cloud misconfiguration monitoring and enterprise risk mitigation strategies. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 9(10), 373-394. https://doi.org/10.32628/CSEIT2361071
10. Atta, S. (2025). Unraveling spray pyrolysis mechanisms: Droplet dynamics and film formation in CdTe thin-film photocathodes. Current Journal of Applied Science and Technology, 44(9), 89-99. https://doi.org/10.9734/cjast/2025/v44i94607
11. Atta, S. (2026). Instability patterns in optoelectronic heterojunctions in GeS nanowires. MRS Bulletin, 51, 643. https://doi.org/10.1557/s43577-026-01126-7
12. Atta, S., & Adenuga, O. M. (2022). Advances in welding inspection standards and quality assurance practices in large-scale power generation projects. International Journal of Engineering and Modern Technology, 8(5).
13. Atta, S., Asomani, E., & Adenuga, O. M. (2018). A systematic review of green corrosion inhibitors from agricultural extracts for mild steel protection in acidic and alkaline media. International Journal of Engineering and Modern Technology, 4(1).
14. Atta, S., Kerubo, M. L., Sabisa, N. E., Adu-Gyamfi, D., Appiah, S., & Onyishi, E. (2025). Environmental corrosion and long-term degradation of crystalline silicon solar cells: Mechanisms, climate effects and mitigation strategies. Current Journal of Applied Science and Technology, 44(10), 9-18.

15. Atta, S., Onyishi, E. E., & Appiah, S. A. (2024). Theoretical analysis of charge transport and recombination losses in thin film solar cell architectures. International Journal of Engineering and Modern Technology, 10(10).
16. Atta, S., Tofah, P., Afrane Kankam, M., & Adenuga, O. M. (2020). A critical review of NDT-based integrity management and corrosion risk assessment strategies for refinery process equipment. International Journal of Engineering and Modern Technology, 6(2).
17. Atta, S., Tofah, P., & Adenuga, O. M. (2021). Advances in non-destructive testing methods for weld integrity evaluation in high-capacity power infrastructure. International Journal of Engineering and Modern Technology, 7(1).
18. Dosunmu, A. A., & Ogundele, P. O. (2021). Incident response and digital forensics strategies for rapid cyberattack containment. Gyanshauryam, International Scientific Refereed Research Journal, 4(4), 239-258.
19. Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1), 32-64. https://doi.org/10.1518/001872095779049543
20. European Union Aviation Safety Agency. (2023). EASA artificial intelligence roadmap 2.0: A human-centric approach to AI in aviation. https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-roadmap-20
21. European Union Aviation Safety Agency. (2024). Machine learning application approval research project: Final report. https://www.easa.europa.eu/en/research-projects/machine-learning-application-approval
22. Federal Aviation Administration. (2023). National aviation research plan, fiscal years 2024-2028. https://www.faa.gov/sites/faa.gov/files/FY_2024-2028_National_Aviation_Research_Plan.pdf
23. Gama, J., Zliobaite, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), Article 44. https://doi.org/10.1145/2523813
24. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. International Conference on Learning Representations. https://arxiv.org/abs/1412.6572
25. Humayed, A., Lin, J., Li, F., & Luo, B. (2017). Cyber-physical systems security: A survey. IEEE Internet of Things Journal, 4(6), 1802-1831. https://doi.org/10.1109/JIOT.2017.2703172
26. Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. Machine Learning, 110(3), 457-506. https://doi.org/10.1007/s10994-021-05946-3
27. International Civil Aviation Organization. (2018). Safety management manual (Doc 9859, 4th ed.). https://www.icao.int/safety-management/SMI/SMM
28. International Civil Aviation Organization. (2019). Aviation cybersecurity strategy. https://www.icao.int/aviationcybersecurity/Pages/Aviation-Cybersecurity-Strategy.aspx
29. International Organization for Standardization. (2014). ISO 55000:2014 asset management, overview, principles and terminology. https://www.iso.org/standard/55088.html
30. International Organization for Standardization. (2018). ISO 31000:2018 risk management, guidelines. https://www.iso.org/standard/65694.html
31. Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, 20(7), 1483-1510. https://doi.org/10.1016/j.ymssp.2005.09.012
32. Jimoh, H. O., Abolle-Okoyeagu, C. J., Ahmed, M. O., & Lawal, N. O. (2023). Advancing security in IoT-driven critical infrastructure: A focus on smart transportation systems. American Journal of Engineering Research, 12(12), 33-46.
33. Komi, N. M., & Ganiu, O. S. (2023). Edge intelligence for resilient microgrid control: Advances, energy sovereignty, and open challenges. Gyanshauryam, International Scientific Refereed Research Journal, 6(3), 525-586. https://doi.org/10.32628/GISRRJ236339
34. Kritzinger, W., Karner, M., Traar, G., Henjes, J., & Sihn, W. (2018). Digital twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine, 51(11), 1016-1022. https://doi.org/10.1016/j.ifacol.2018.08.474
35. Ladapo, O. O., Dosunmu, A. A., Jooda, D., & Abolaji, T. O. (2022). Human-in-the-loop machine learning: A state of the art. Journal of Frontiers in Multidisciplinary Research, 3(1), 656-669. https://doi.org/10.54660/.JFMR.2022.3.1.656-669
36. Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical Systems and Signal Processing, 104, 799-834. https://doi.org/10.1016/j.ymssp.2017.11.016

37. Leveson, N. G. (2011). Engineering a safer world: Systems thinking applied to safety. MIT Press. https://mitpress.mit.edu/9780262533690/engineering-a-safer-world/
38. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
39. Malekloo, A., Özer, E., AlHamaydeh, M., & Girolami, M. (2021). Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. Structural Health Monitoring, 21(4), 1906-1955. https://doi.org/10.1177/14759217211036880
40. National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1
41. National Institute of Standards and Technology. (2024). The NIST cybersecurity framework (CSF) 2.0 (NIST CSWP 29). https://doi.org/10.6028/NIST.CSWP.29
42. Nwakamma, S., Ojukwu, J., & Oyesiji, S. O. (2026). Crypto-agile migration to hybrid post-quantum cryptography in enterprise TLS, PKI, VPN, and cloud-native systems and services. Gulf Journal of Computer Sciences, 1(1). https://doi.org/10.51594/gjcs.v1i1
43. Obriki, O. D., & Arumosoye, O. M. (2024). Systematic review of incident investigation approaches and prevention-oriented learning in industrial operations. Shodhshauryam, International Scientific Refereed Research Journal, 7(4), 239-263. https://doi.org/10.32628/SHISRRJ247163
44. Olodo, A., Akanbi, O., & Adenuga, O. (2025). Conceptualizing a closed-loop digital twin for real-time plasma control in large-area magnetron sputtering systems. International Journal of Engineering and Modern Technology, 11(12), 205-227. https://doi.org/10.56201/ijemt.vol.11.no12.2025.pg205.227
45. Ominyi, M., & Anichukwueze, C. C. (2024). Conceptualizing responsible AI governance structures for automated decision systems in regulated industries. International Journal of Social Sciences and Management Research, 10(11), 403-430. https://doi.org/10.56201/ijssmr.v10.no11.2024.pg.403.430
46. Onche, V. O., Ogbonna, C. S., & Adegbite, M. P. (2024). Effectiveness of cybersecurity awareness training on insider-threat behavior among university administrators: An integrative academic review. International Journal of Computer Science and Mathematical Theory, 10(2), 117-144. https://doi.org/10.56201/ijcsmt.v10.no2.2024.pg117.144
47. Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics Part A, 30(3), 286-297. https://doi.org/10.1109/3468.844354
48. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144. https://doi.org/10.1145/2939672.2939778
49. RTCA. (2011). DO-178C: Software considerations in airborne systems and equipment certification. https://www.rtca.org/standards/
50. RTCA. (2014). DO-326A: Airworthiness security process specification. https://www.rtca.org/standards/
51. SAE International. (2010). ARP4754A: Guidelines for development of civil aircraft and systems. https://www.sae.org/standards/content/arp4754a/
52. Saltzer, J. H., & Schroeder, M. D. (1975). The protection of information in computer systems. Proceedings of the IEEE, 63(9), 1278-1308. https://doi.org/10.1109/PROC.1975.9939
53. Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2019). Digital twin in industry: State-of-the-art. IEEE Transactions on Industrial Informatics, 15(4), 2405-2415. https://doi.org/10.1109/TII.2018.2873186
54. Zhang, E., Dao, M., Karniadakis, G. E., & Suresh, S. (2022). Analyses of internal structures and defects in materials using physics-informed neural networks. Science Advances, 8(7), eabk0644. https://doi.org/10.1126/sciadv.abk0644