Trustworthy Autonomy through Ethics-Aware Integrity Monitoring

Samir Khan[1] and Takehisa Yairi[2]

[1,2] *University of Tokyo, Research Center for Advanced Science and Technology,*
4 Chome-6-1 Komaba, Tokyo, 153-8904, Japan
samir.s.khan@cranfield.com

Abstract

The increasing reliance on autonomous systems in aerospace raises a dual challenge: ensuring that aircraft remain physically safe while also making decisions that are ethically responsible and auditable. Traditional Prognostics and Health Management (PHM) has matured in predicting and preventing technical failures, yet it operates independently from emerging mechanisms for ethical oversight. This separation leaves open questions about how to arbitrate when physical health indicators and ethical obligations point to conflicting actions, and how such decisions can be explained and trusted in high-stakes environments.

This paper proposes a unified framework that integrates PHM with a runtime ethics module, treating both health and ethical risks as core aspects of system integrity. Central to the approach is an Event Manager that evaluates proposed actions against prognostic forecasts and codified ethical rules, applying a priority scheme that ranks safety of life above regulatory compliance, system preservation, and mission objectives. To ensure transparency and accountability, the system generates event-triggered explanations and records both outcomes and justifications in a tamper-evident blockchain ledger.

We demonstrate the framework through three case studies: an aircraft integrity trial facing an emergency landing where passenger survival must be balanced with bystander safety, a long-endurance surveillance UAV deciding whether to continue data collection or return before failure, and a UAV swarm with integrity auditing in multi-agent environment. Across Monte Carlo simulations, the integrated PHM–Ethics approach consistently reduced combined integrity losses, defined as missed failures plus ethical violations, when compared with PHM-only or Ethics-only baselines. Explanation payloads remained lightweight and blockchain logging introduced only modest latency, demonstrating feasibility for real-time aerospace operations. By showing that PHM and ethics can be brought into a single decision loop without sacrificing performance, this work provides a concrete pathway toward trustworthy autonomy in aerospace.

The results highlight how transparency, auditable decision-making, and equitable risk management can be engineered into safety-critical systems, offering practical tools to support certification, regulatory trust, and public confidence in the next generation of autonomous flight.

1. Introduction

The convergence of advanced AI with autonomous cars, drones, and aerospace systems has enabled unprecedented levels of autonomy, but also surfaced new safety and ethical concerns (Habbal, Ali, & Abuzaraida, 2024; Khan & Yairi, 2018). Prognostics and Health Management (PHM) systems have long been deployed in aerospace to monitor the health of components, predict failures, and mitigate risks. By detecting early signs of degradation and enabling proactive interventions, PHM has improved safety and reliability. Today, most aircraft and unmanned aerial vehicles (UAVs) already host real-time health monitoring that can adapt missions to current system status.

However, technical reliability alone is no longer sufficient. As AI increasingly drives mission-critical decisions, autonomy raises new questions about whether those decisions are equitable, transparent, and aligned with societal values. High-profile debates such as the “Trolley Problem” in autonomous driving (Jenkins, Cernỳ, & Hríbek, 2022) illustrate the moral dilemmas that an AI pilot or driver may face in life-and-death situations. Current safety assurance standards remain inadequate for such contexts (Jobin, Ienca, & Vayena, 2019). Traditional certification presumes human operators will remain accountable for moral judgment (Braun, 2025), but as autonomy deepens, responsibility shifts to algorithms; demanding novel mechanisms for monitoring and auditing AI behavior.

Concerns around algorithmic bias further complicate matters: perception and decision-making models may underperform for certain populations or contexts, systematically placing them at higher risk. In aerospace, an autonomous aircraft might learn or be programmed to prioritize passenger or mission

https://doi.org/10.36001/IJPHM.2025.v16i2.4386

objectives in ways that conflict with public safety or fairness. Addressing these risks requires systems that can enforce ethical rules alongside PHM. One promising enabler is blockchain technology, which provides tamper-evident, distributed audit trails of sensor inputs, detected faults, and AI decision rationales (Christidis & Devetsikiotis, 2016). Immutable ledgers can support post-incident forensics and continuous oversight by engineers, regulators, and the public.

The motivation for this research is to advance autonomy that is both physically safe and ethically trustworthy. We argue that PHM (assessing system health) and AI ethics monitoring (assessing decision "health") must be integrated into a unified framework. Such integration improves operational safety by preventing both technical failures and unsafe actions, while also enhancing social acceptance by demonstrating compliance with ethical principles. The challenge is inherently interdisciplinary, requiring aerospace engineering, machine learning, distributed systems, and applied ethics.

The remainder of this paper is structured as follows. Section 2 reviews prior work on PHM in aerospace, ethical auditing of AI, blockchain for accountability, and explainable AI (Ribeiro, Singh, & Guestrin, 2016; Adadi & Berrada, 2018). It also identifies key gaps and formulates our research questions. Section 3 presents the proposed framework, illustrated through case studies in aerospace. Section 4 outlines three case studies, and Section 5 discusses limitations, standardization needs, and implications for both the PHM and AI ethics communities.

1.1. Key contributions

This article makes several contributions toward developing autonomous aerospace systems that are transparent, safety-aware, and ethically trustworthy.

- A unified framework that integrates PHM, real-time ethics monitoring, and tamper-evident event logging into a single decision loop. Unlike prior approaches that treat these components separately, our framework emphasizes decision-level arbitration: PHM forecasts and ethical constraints are evaluated together through an Event Manager that prioritizes safety of life, legal compliance, system preservation, and mission goals in a consistent hierarchy.
- A framework demonstrates how joint health and ethics monitoring can be achieved in practice. Sensor data streams (such as GPS, IMU, and engine state) are fused not only to detect anomalies and incipient system degradation, but also to scrutinize mission decisions against codified ethical guidelines. This enables the system to flag unsafe actions, disproportionate risk distribution, or potential violations of operational boundaries before they escalate.
- Shows how accountability can be strengthened through a distributed ledger. A blockchain-based event log records system anomalies, AI recommendations, human interventions, and ethical breaches in real time, creating a transparent audit trail. This makes both the outcomes and the justifications for decisions traceable to regulators, engineers, and other stakeholders, thereby supporting post-incident forensics and long-term governance.
- The framework is validated through three diverse and safety-critical aerospace scenarios:
  - an emergency landing of an aircraft facing conflicting stakeholder risks,
  - a UAV deciding whether to continue or abort a mission under emerging engine failure, and
  - multi-agent coordination of a UAV swarm navigating under dynamic airspace constraints.

  Each case demonstrates how PHM, ethics monitoring, and blockchain auditability interact in practice, reducing unsafe or non-compliant outcomes compared with baseline approaches and highlighting the feasibility of embedding ethical governance into real-time aerospace autonomy.

2. LITERATURE REVIEW

PHM is an engineering discipline dedicated to assessing system health, diagnosing faults, and predicting remaining useful life to enable timely interventions. In aerospace, PHM is needed for monitoring critical components such as engines, avionics, and structures, where early fault detection can prevent catastrophic failures. A typical PHM pipeline would involve anomaly detection, fault diagnostics, and prognostics that forecast future degradation trajectories (Khan & Yairi, 2018).

Literature broadly categorizes PHM strategies as model-based, data-driven, or hybrid. Model-based methods rely on physics-of-failure knowledge and system dynamics. For instance, thermodynamic engine models, combined with Kalman or particle filters, can track deviations that signal incipient faults (Goebel et al., 2017; Jardine, Lin, & Banjevic, 2006). These methods offer interpretability but require high-fidelity models that may not capture unmodeled complexities. Data-driven approaches, by contrast, leverage historical sensor data to train machine learning models for anomaly detection and remaining useful life (RUL) prediction. Algorithms ranging from random forests to deep neural networks have shown promise in capturing subtle degradation signatures (Lei et al., 2018). Their strength lies in scalability and adaptability to complex systems, but they often demand large labeled datasets and can struggle to generalize outside training conditions.

Hybrid PHM is emerging as a practical compromise, combining the interpretability of physics-based models with the adaptability of machine learning. For example, physics-based simulations can generate features that improve the performance of data-driven prognostic models, while learning algorithms can refine estimates in real operational contexts. Such approaches have been shown to improve both the accuracy of fault prediction and the timeliness of maintenance scheduling, enhancing overall system reliability. Aerospace applications increasingly embed these methods into Integrated Vehicle Health Management (IVHM) architectures that continuously monitor subsystems and trigger reconfiguration or mission replanning when anomalies arise (Saxena, Celaya, Saha, Saha, & Goebel, 2010). UAVs, for example, can autonomously adapt flight plans or execute emergency landings in response to detected health degradation.

PHM has already demonstrated its value by reducing maintenance costs and improving operational availability across many commercial and defense aviation applications (Khan & Yairi, 2018). Yet, traditional PHM is limited to the physical state of machines. As autonomy shifts high-level decision-making from human operators to AI algorithms, the notion of "system health" must be expanded. Decisions themselves may degrade in quality or breach ethical or regulatory boundaries, just as hardware may fail. This recognition motivates the integration of PHM with ethics monitoring, where the health of the "decision process" is evaluated alongside the health of the hardware, ensuring that autonomous aerospace systems are not only mechanically reliable but also ethically trustworthy.

2.1. Ethics and AI Auditing

As AI systems become more autonomous, ensuring that their decisions are ethically sound, fair, and auditable has become a critical requirement (IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019). Ethical AI auditing refers to methods for evaluating AI systems against moral principles, fairness criteria, and regulatory expectations. Concerns in aerospace parallel those seen in self-driving cars: biases in perception or decision algorithms, opaque "black box" outputs that defy justification, and the absence of mechanisms to resolve moral dilemmas in real time.

One major concern is *algorithmic bias*, where system performance systematically favors or disadvantages particular groups. For example, pedestrian detection algorithms show measurable disparities—up to 7.5% less accurate for darker-skinned individuals under certain conditions—raising significant justice and safety issues (Giudici, Centurelli, & Turchetta, 2024). In aerospace, similar risks arise if autonomous aircraft prioritize passengers or mission goals at the expense of bystander safety, or if UAV swarms concentrate risk on a single vehicle. Rather than implying UAVs themselves hold moral claims, we interpret fairness in swarm operations as *operational equity*: ensuring that burdens such as mission loss or hazard exposure are distributed across vehicles in a balanced manner.

It is also important to distinguish *ethics* from *legal compliance*. Regulatory standards enforce minimum safety requirements and accountability, but they do not address deeper normative questions. For instance, an emergency landing might legally satisfy airworthiness standards while still making ethically questionable choices, such as consistently prioritizing passenger lives over those of people on the ground (Robert E Joslin, 2020). Current regulations assume a human pilot is present to exercise judgment, yet increasing autonomy creates a governance gap where AI systems must themselves embody ethical safeguards.

Recent guidelines, such as the EU High-Level Expert Group's recommendations, emphasize autonomy, harm minimization, justice, and explainability as "ethical imperatives" for AI developers. In practice, companies and agencies are deploying algorithmic audits, bias tests, and fairness metrics, often offline and retrospective. Aerospace, however, demands *runtime ethical monitoring* that can operate like a watchdog: continuously checking decisions against codified ethical rules and recording violations in real time. Such monitoring complements PHM directly, while PHM ensures the physical system remains healthy, ethics auditing evaluates the *cognitive health of decisions*. Without this, even technically reliable systems risk losing public trust and regulatory approval.

Finally, the PHM community itself has recognized the importance of ethics. As PHM systems become more autonomous in recommending or executing actions, they too can inherit fairness and bias concerns if training data or prognostic logic are skewed (Goebel et al., 2017). This underscores that transparency, impartiality, and public safety—longstanding values in engineering ethics—must now be designed directly into the AI algorithms that underpin autonomous aerospace systems.

2.2. Blockchain for Auditability: Ethereum vs. Hyperledger Fabric

Blockchain technology offers a compelling solution for ensuring transparency and auditability in autonomous systems (Christidis & Devetsikiotis, 2016). A blockchain is a distributed ledger replicated across a network of nodes, valued for its decentralization, immutability, and transparency. Once data is stored on-chain, it cannot be retroactively altered, which makes it ideal for maintaining a tamper-evident record of events and decisions. Each transaction in the ledger is timestamped, cryptographically secured, and visible to authorized parties, thereby supporting traceability and accountability in AI-driven aerospace systems.

Several studies have investigated combining AI decision-making with blockchain in order to create immutable audit trails. In principle, every major action taken by an autonomous agent (e.g., a UAV rerouting to avoid restricted airspace, or an aircraft executing an emergency descent) can be logged as a blockchain transaction, along with its relevant inputs and explanatory metadata. This enables investigators to reconstruct the decision-making process post hoc, distinguishing between sensor faults, biased reasoning, or valid but high-stakes trade-offs. Such tamper-proof auditability discourages negligence or data manipulation, fostering trust among regulators, operators, and the public.

There are, however, multiple blockchain architectures with distinct trade-offs. Public blockchains such as Ethereum provide open participation, high transparency, and strong tamper-resistance. Smart contracts can automate audit functions, for example by committing a log entry whenever an anomaly score exceeds a threshold. However, public blockchains suffer from limited throughput (tens of transactions per second), variable confirmation latency, and transaction fees, all of which constrain real-time aerospace usage. In addition, sensitive operational data may be unsuitable for fully public ledgers.

By contrast, permissioned blockchains such as Hyperledger Fabric restrict participation to vetted organizations (e.g., manufacturers, airlines, regulators). Fabric's modular consensus protocols are more computationally efficient than Ethereum's proof-of-work, yielding significantly lower logging latency and higher throughput (Androulaki et al., 2018). It also provides granular access control, allowing some records (e.g., failure logs) to be shared with regulators while shielding sensitive commercial data from competitors. Empirical comparisons of Ethereum and Hyperledger show that Fabric achieves lower end-to-end latency (on the order of hundreds of milliseconds rather than seconds) and avoids transaction fees, making it more practical for safety-critical applications.

Despite these benefits, challenges remain. High-frequency sensor data cannot feasibly be stored on-chain due to bandwidth and storage constraints; instead, only cryptographic hashes or anomaly summaries should be logged. Network connectivity failures could delay or fragment event logging, raising the need for robust buffering and synchronization schemes. Smart contracts themselves must be verified to avoid introducing new vulnerabilities. Finally, data privacy must be carefully managed, especially when logs contain sensitive operational details or passenger information.

In aerospace, blockchain's greatest promise lies in enabling multi-stakeholder trust. Manufacturers, operators, and regulators can all access the same immutable audit trail, reducing reliance on proprietary operator logs and supporting independent certification. Smart contracts can even encode governance policies, such as automatically flagging ethically non-compliant actions to oversight authorities, or recording which version of an AI model was deployed at a given time for regulatory traceability. In this way, blockchain becomes not only a technical tool for transparency but also a bridge to certification pathways, aligning PHM and ethical monitoring with regulatory expectations for explainability and accountability.

Prior research comparing the two platforms suggests that both are technically sound enough to act as an audit ledger but with important trade-offs. For example, Ethereum emphasizes decentralization and public transparency, while Hyperledger Fabric provides lower latency and more fine-grained confidentiality controls. Other platforms such as Corda, widely used in finance, also illustrate permissioned approaches with efficient validation.

To make these trade-offs clearer for aerospace contexts, Table 1 provides a side-by-side comparison of Ethereum, Hyperledger Fabric, and Corda across dimensions such as latency, throughput, fees, privacy, and suitability for safety-critical autonomy.

As the comparison shows, no single platform is ideal across all dimensions. Public chains provide maximal transparency but suffer from high latency and limited throughput, while permissioned chains deliver low-latency, high-throughput performance with stronger confidentiality controls at the expense of decentralization.

For aerospace certification contexts, these trade-offs are highly consequential. Regulators are more likely to favor permissioned approaches (e.g., Hyperledger Fabric) where access rights can be tightly controlled, latency guarantees can be formally validated, and sensitive operational data is not exposed publicly. At the same time, hybrid strategies—such as committing periodic hashed summaries from a Fabric ledger to a public chain like Ethereum—can provide long-term, tamper-proof public verifiability without sacrificing real-time performance or confidentiality.

This alignment between blockchain performance characteristics and regulatory expectations suggests that auditability layers can be designed not only for technical transparency but also as evidence directly usable in certification and compliance reviews. In this way, blockchain auditability becomes a dual enabler: ensuring trustworthy system operation during runtime while also supporting regulatory approval by providing immutable, verifiable records of health and ethical decision-making.

2.3. Explainable AI for Transparency and Accountability

A recurring theme across both PHM and AI ethics is the question of explainability, why a model produced a specific prediction or decision (Nor, Pedapait, & Muhammad, 2021). Explainable AI (XAI) methods aim to open the "black box" of advanced AI models (e.g., deep neural networks) and render them understandable to humans. In safety-critical domains such as aerospace, explanations are not optional but a de facto requirement: engineers, operators, and regulators must be able to verify that the AI is functioning correctly and not behaving in unsafe or biased ways. The EU Trustworthy AI guidance explicitly identifies explicability, alongside fairness and accountability, as a core principle of responsible AI (IEEE Global Initiative on Ethics of Autonomous and Intelli-

Table 1. Comparison of Blockchain Platforms for Aerospace Auditability

Feature	**Ethereum (Public)**	**Hyperledger Fabric (Permissioned)**	**Corda (Permissioned, Finance-Oriented)**
Consensus mechanism (Androulaki et al., 2018)	Proof-of-Stake (formerly Proof-of-Work)	Modular consensus (Raft, Kafka, etc.)	Notary-based validation (no global broadcast)
Latency (Ucbas, Eleyan, Hammoudeh, & Alohaly, 2023)	High (seconds to minutes per transaction)	Low (sub-second to hundreds of ms)	Low (ms to seconds, depending on deployment)
Throughput (Ucbas et al., 2023; Thakkar, Nathan, & Viswanathan, 2018)	$\sim$15–30 transactions/s	10^3+ transactions/s (enterprise-optimized)	10^3+ transactions/s
Transaction fees (Geyer, Jacobsen, Mayer, & Mandl, 2023)	Gas fees (variable, can be high in congestion)	No transaction fees	No transaction fees
Transparency (Ucbas et al., 2023)	Fully public, anyone can validate	Restricted to consortium members	Restricted to bilateral/consortium channels
Privacy/ confidentiality (Androulaki et al., 2018)	Weak (all data public, though encryption/hashing possible)	Strong (granular access control, private channels)	Strong (transactions visible only to relevant parties)
Scalability (Thakkar et al., 2018; Gorenflo, Lee, Golab, & Keshav, 2020)	Limited (network congestion, high cost under load)	High (scales with consortium infrastructure)	High (scales with business networks)
Suitability for aerospace (Ucbas et al., 2023)	High transparency; public accountability for services (e.g., air taxis), but limited real-time viability	Strong candidate for regulated aerospace operations with multiple stakeholders (OEM, operator, regulator)	Potentially useful for financial/contractual audit, less tested in safety-critical autonomy

gent Systems, 2019).

Among model-agnostic techniques, LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) are widely used. LIME approximates the model's behavior locally by building simple interpretable surrogates (e.g., sparse linear models) around a prediction, while SHAP leverages cooperative game theory to assign contribution values (Shapley values) to each feature. These methods can highlight, for example, why a prognostic model predicted an engine failure within five hours rather than fifty, or why a path-planning algorithm selected a particular route. Such insights can be cross-checked against engineering intuition, improving confidence that the model is capturing real physical patterns rather than spurious correlations.

Within the PHM community, XAI has shown value both for verifying prognostic models and for extracting new domain knowledge from data. A recent systematic review of XAI in PHM reported that interpretability techniques not only improved user trust but also helped refine feature selection and anomaly detection strategies (Nguyen, Nguyen, & Medjaher, 2024). Importantly, many studies demonstrated that interpretability could be achieved without significantly sacrificing predictive performance. Nonetheless, challenges remain in defining quantitative metrics for the "goodness" of explanations, and in leveraging human feedback derived from explanations in a systematic way.

In the ethics domain, XAI serves as a critical audit tool. For instance, if an autonomous UAV chooses to fly over a densely populated area, triggering an ethics violation, XAI methods could reveal which variables (e.g., wind conditions, map data, or faulty sensor readings) most influenced the decision. Similarly, bias in perception systems, such as reduced confidence in detecting pedestrians with darker skin tones (Giudici et al., 2024), can be diagnosed by showing how input features affect prediction outcomes. These capabilities allow engineers to pinpoint and mitigate the root causes of unfair or unsafe decisions.

A practical concern, however, is computational cost. Methods such as SHAP can be expensive to compute, especially for deep models and real-time systems. For aerospace autonomy, lightweight or approximate XAI methods may be required, or explanations may need to be selectively generated for safety-critical decisions rather than every prediction. Benchmarking explanation latency is therefore essential for assessing feasibility in onboard contexts.

Finally, XAI and blockchain can be combined to produce not just an immutable decision log but also an immutable *reasoning log*. Instead of merely recording what decision was made, the system can log the accompanying explanation—e.g., "Decision: sudden braking; explanation: detected obstacle at 30m, top features = shape, motion, confidence = 95%." This "Glass Box AI" approach enhances traceability and trust by ensuring that justifications, not just outcomes, are preserved for audit and certification. In our framework, integrating XAI into both the PHM module ("why the engine is failing") and the ethics module ("why a decision was unsafe or biased") provides the bridge from raw technical monitoring to meaningful human oversight. This strengthens regulatory acceptance by offering interpretable evidence that autonomous systems act within both physical and ethical safety bounds.

2.4. Research Gaps and Challenges

Despite rapid advances in PHM, AI ethics, and distributed auditability, their integration for autonomous aerospace systems remains underdeveloped. Our review identifies several critical research gaps and challenges:

- Fragmented monitoring paradigms: Health management and ethics monitoring are treated independently, even though failures in either can cause catastrophic outcomes. An integrated view of "system health" that includes both physical degradation and ethical non-compliance is rarely studied.
- Siloed toolchains and poor interoperability: PHM engineers rely on diagnostic and prognostic models, while AI ethicists employ bias audits and fairness metrics. These toolchains seldom communicate. As a result, a PHM module may detect sensor degradation without realizing that the AI decision logic is compensating in a biased or unsafe manner, or an ethics check may flag a decision without understanding it was triggered by failing hardware.
- Real-time ethical monitoring deficit: PHM algorithms are routinely embedded onboard for real-time safety, whereas ethical monitoring is largely retrospective (offline audits, dataset analysis). Designing real-time, context-sensitive ethics monitors is challenging due to computational cost, ambiguity in ethical norms, and the need for domain-specific constraints (Jobin et al., 2019). Without this capability, safety-critical lapses may only be detected after-the-fact.
- Decision arbitration under conflicts: Existing work does not specify how PHM and ethics alerts should be reconciled in real time. For example, if PHM predicts imminent engine failure but the ethics module flags the current diversion plan as unsafe, what arbitration logic should prioritize between physical and moral risks? Without formal schemes for arbitration and prioritization, integrated monitoring cannot meaningfully support autonomy.
- Verification, certification and legal gaps: Current aerospace certification emphasizes reliability, not ethical accountability (Kusnirakova & Buhnova, 2023). There is no formal pathway to demonstrate that an autonomous system makes "ethically compliant" decisions. Bridging this requires novel verification techniques, potentially formal methods, simulation testbeds, or scenario-based evaluation, that address both physical safety and fairness. Importantly, ethical compliance must not be conflated with mere legal compliance: regulation sets minimum thresholds, whereas ethical alignment concerns broader societal values.
- Data privacy and security: Integrating PHM and ethics monitoring produces massive multi-layered datasets (sensor telemetry, decision logs, explanation traces), some of which may be stored on distributed ledgers. Ensuring privacy while retaining auditability demands new data aggregation, hashing, and access-control strategies, especially since aerospace logs often involve sensitive or proprietary information.
- Lack of interdisciplinary metrics: Traditional metrics (accuracy, recall) apply to PHM and bias detection separately, but integrated monitoring requires composite measures of trustworthiness. For instance, a "mission integrity index" could capture both absence of critical failures and absence of ethical violations. No standardized metrics currently exist for end-to-end evaluation of autonomy that unifies engineering reliability and ethical soundness.
- Quantitative validation gap: Few studies provide quantitative benchmarks comparing PHM-only, ethics-only, and integrated approaches in mission-level performance, compliance rates, or logging overhead. Without such benchmarks, claims about the benefits of integration remain largely conceptual.

Figure 1 summarizes the current literature landscape and highlights the missing links between PHM, ethics monitoring, explainability, and blockchain audit layers. Solid arrows indicate connections that are relatively well-developed in prior research, such as PHM-to-XAI pipelines and the use of explanations for logging. In contrast, dashed red arrows and the shaded "missing block" illustrate underexplored or absent mechanisms. For example, while PHM and ethics monitoring each independently feed into XAI or blockchain audit trails, there is little work on their direct integration or on arbitration mechanisms that reconcile conflicts between them. The red annotations further emphasize where the literature falls short: ethics monitoring is typically retrospective rather than real-time, PHM and ethics toolchains remain siloed with poor interoperability, and certification pathways for ethically aligned autonomy are still weak or undefined. Together, these visual cues reinforcw the argument that co-location of PHM and ethics modules is insufficient. A principled architecture must instead treat ethical violations as a form of system health degradation, enabling detection, arbitration, and recovery in the same way as physical faults.

2.5. Research Questions

To address the highlighted gaps, we formulate several guiding research questions (RQs), hypotheses and case studies used in

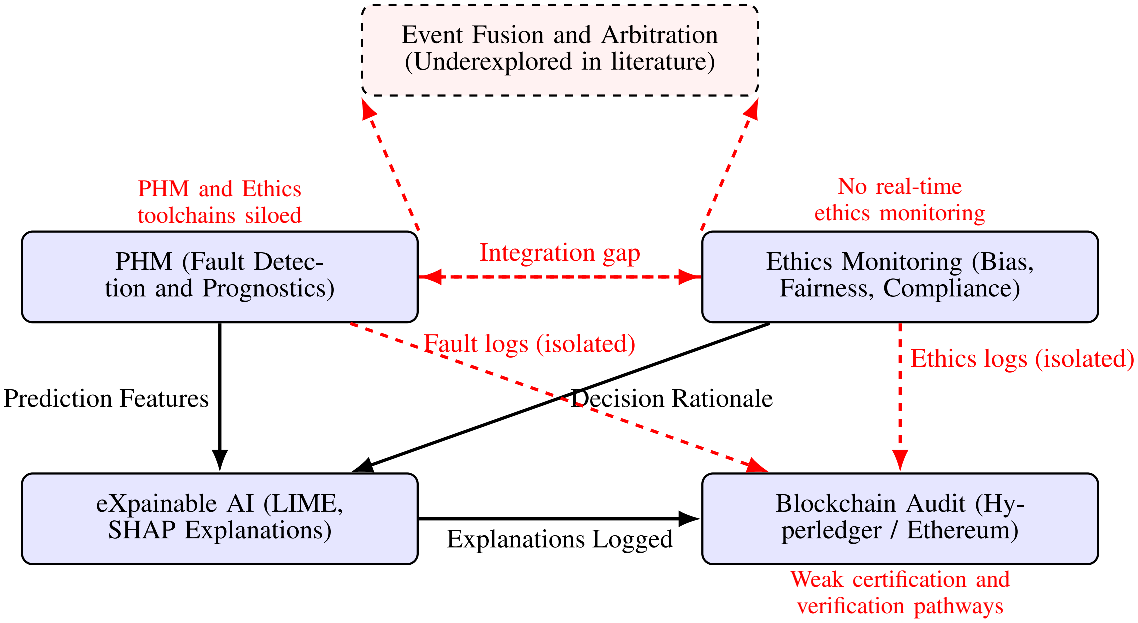
Event Fusion and Arbitration (Underexplored in literature)
PHM and Ethics toolchains siloed
No real-time ethics monitoring
PHM (Fault Detection and Prognostics)
Integration gap
Ethics Monitoring (Bias, Fairness, Compliance)
Fault logs (isolated)
Ethics logs (isolated)
Prediction Features
Decision Rationale
eXpainable AI (LIME, SHAP Explanations)
Explanations Logged
Blockchain Audit (Hyperledger / Ethereum)
Weak certification and verification pathways

Figure 1. Current literature landscape at the PHM–AI Ethics intersection. Solid arrows indicate relatively well-studied connections, while dashed red arrows show underdeveloped or missing links. Research gaps include fragmented monitoring, lack of real-time ethics checks, absent arbitration logic, weak interoperability, and no unified certification pathways. Section 3 builds on this to propose a fully integrated framework.

this paper. This is to ensure that these questions are not only conceptual but testable and how the proposed framework can be validated in practice:

- **RQ1: How can physical health and ethical behavior of an autonomous system be monitored jointly in real time?**
  Hypothesis 1: By fusing PHM anomaly detectors with an ethics monitoring module that observes AI decisions, a unified monitoring system will capture critical incidents missed by standalone approaches, such as unsafe actions under degraded conditions.
  Case study: The *emergency landing case* tests whether simultaneous PHM (engine degradation signals) and ethics monitoring (harm distribution across passengers vs. bystanders) can provide earlier detection of safety-critical dilemmas than either subsystem alone.
- **RQ2: What architectural approach can effectively combine model-based diagnostics, data-driven anomaly detection, and ethical rule-checking?**
  Hypothesis 2: A modular architecture where PHM and Ethics modules operate in parallel but converge at a common evaluation and logging layer will provide comprehensive coverage. Model-based diagnostics will offer precise early warnings, machine learning will detect novel failure patterns, and the ethics layer will enforce normative constraints.
  Case study: The *UAV engine failure case* evaluates the modular architecture where model-based diagnostics predict imminent failure, machine learning anomaly detectors recognize novel degradation patterns, and the ethics module evaluates whether mission continuation or abortion complies with ethical thresholds.
- **RQ3: Can blockchain-based logging meet the real-time and confidentiality needs of aerospace auditing?**
  Hypothesis 3: A permissioned blockchain (e.g., Hyperledger Fabric) can deliver low-latency logging with confidentiality, while public blockchains (e.g., Ethereum) can provide long-term verifiability. A hybrid strategy may combine both, balancing operational efficiency with public trust.
  Case study: The *swarm UAV mission case* validates blockchain logging under high-frequency, multi-agent conditions. Here, Hyperledger Fabric is used for real-time intra-swarm logging, while Ethereum checkpoints are tested for long-term public verifiability of mission compliance with no-fly zones.
- **RQ4: How can explainable AI techniques enhance transparency of both PHM and ethics alerts for human stakeholders?**
  Hypothesis 4: Embedding XAI outputs (e.g., SHAP/LIME explanations) with alerts will improve human interpretability and trust. Logging explanations alongside events on blockchain will further guarantee accountability.
  Case study: Across all three cases, XAI methods (SHAP, LIME) are used to generate explanations for alerts. For example, in the *emergency landing case*, explanations

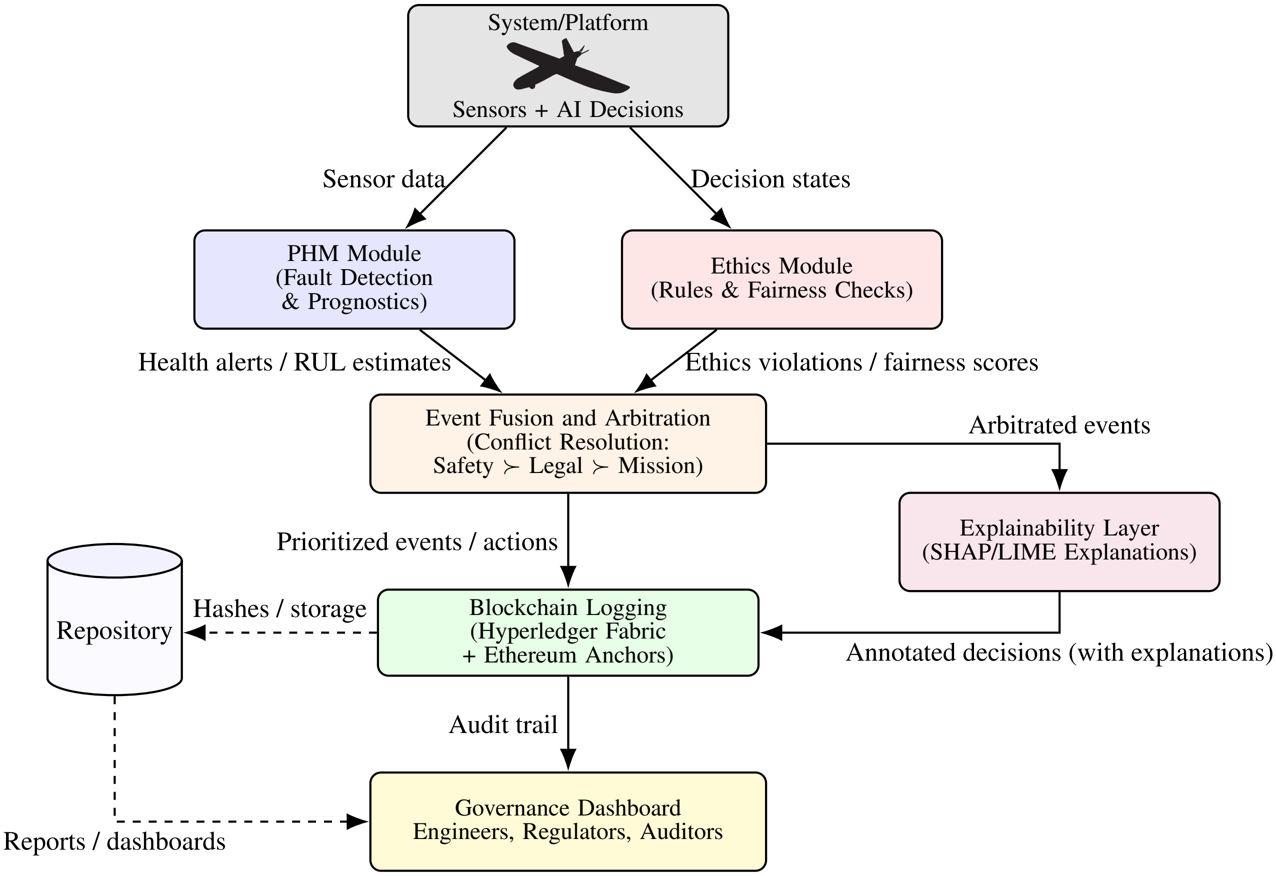
System/Platform
Sensors + AI Decisions
Sensor data
Decision states
PHM Module (Fault Detection & Prognostics)
Ethics Module (Rules & Fairness Checks)
Health alerts / RUL estimates
Ethics violations / fairness scores
Event Fusion and Arbitration (Conflict Resolution: Safety ≻ Legal ≻ Mission)
Arbitrated events
Explainability Layer (SHAP/LIME Explanations)
Prioritized events / actions
Repository
Hashes / storage
Blockchain Logging (Hyperledger Fabric + Ethereum Anchors)
Annotated decisions (with explanations)
Audit trail
Governance Dashboard Engineers, Regulators, Auditors
Reports / dashboards

Figure 2. High-level framework for trustworthy UAV autonomy. Data from sensors and AI decisions flow into PHM and Ethics monitoring, converge through arbitration, and are logged with blockchain support. An explainability layer provides reasoning for decisions, while a governance interface enables human oversight and regulatory auditing.

clarify why a specific landing zone was prioritized; in the *UAV case*, they reveal which sensor anomalies drove the failure prediction.

- **RQ5: How should conflicting PHM and ethics alerts be arbitrated in mission-critical contexts?**
  Hypothesis 5: Coupled arbitration schemes that weigh both physical safety and ethical compliance can mitigate risks more effectively. For example, if a UAV faces both engine failure and an ethics violation in its flight path, integrated arbitration can propose safer alternatives than either subsystem alone.
  Case study: The *UAV engine failure case* explicitly tests conflict resolution: the PHM module may recommend immediate diversion, while the ethics module flags that the planned landing route crosses restricted civilian zones. Arbitration logic will be evaluated for its ability to balance safety and fairness.

- **RQ6: What metrics can evaluate the overall trustworthiness of an integrated framework?**
  Hypothesis 6: Composite performance metrics—such as "mission success with integrity"—will better capture end-to-end system reliability than traditional PHM or bias-only metrics. User studies with engineers and regulators can validate the perceived utility of these measures.
  Case study: All three case studies will be assessed using composite mission-level metrics such as "mission success with integrity." For example, in the *swarm UAV case*, integrity is defined not only by mission completion but also by avoidance of no-fly zone violations.

Answering these research questions will establish whether integration of PHM and AI ethics monitoring can yield autonomous aerospace systems that are not only technically reliable but also ethically transparent and certifiable for real-world deployment.

3. A FRAMEWORK FOR INTEGRATED PHM AND ETHICS MONITORING

To address the identified gaps, we propose a unified framework that integrates PHM and AI ethics monitoring, supported by a blockchain-based audit layer and a governance interface (Arkin, Ulam, & Duncan, 2009). Figure 2 shows the main modules and data flows. Solid lines represent real-time signals, decisions, or logs; dotted lines represent transparency and oversight interactions. The modular design ensures scalability across aerospace platforms (e.g., UAV swarms, UAVs) and can accommodate future extensions such as cybersecurity checks or fairness audits.

Integration here refers to the co-existence of PHM and ethics modules within a single governance and logging framework, enabling shared data, coordinated alerts, and unified audit trails, while retaining their domain-specific algorithms. This avoids the ambiguity of simple co-location and emphasizes a common event management and oversight layer that couples physical health and ethical monitoring.

3.1. Core Components

1. **Sensing and data acquisition:** Standard sensors (altitude, vibration, fuel level, temperature, etc.) and AI-internal telemetry (e.g., decision confidence, path-planning costs) provide the raw data stream. We formalize the input as a feature vector:

   $$x_t = [h_t, v_t, f_t, T_t, \ldots, c_t, p_t],$$

   where h_t denotes altitude, v_t vibration, f_t fuel level, T_t temperature, c_t AI decision confidence, and p_t path-planning cost at time t. We partition the data stream into

   $$x_t = \left(x_t^{\text{PHM}},\ x_t^{\text{Ethics}}\right),$$

   with x_t^{PHM} used by health monitoring (physical signals) and x_t^{Ethics} derived from AI decision-making states. Sensor measurements are corrupted by noise $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$, and potential drifts Δ_t are tracked to support anomaly detection downstream. This unified view ensures that both modules share a consistent telemetry backbone.

2. **PHM module (health monitoring):** The PHM block combines residual-based diagnostics, state-estimation filters, and machine-learning anomaly detectors.

   (a) Residual analysis. Given a model $y_t = g(u_t, \theta) + \epsilon_t$, we compute the residual

   $$r_t = y_t^{\text{measured}} - y_t^{\text{model}}.$$

   A fault is flagged if $\|r_t\| > \delta$, with δ a detection threshold.

   (b) Kalman filter prognostics: We propagate hidden states s_t using:

   $$s_{t+1} = As_t + Bu_t + w_t, \qquad y_t = Cs_t + v_t,$$

   where w_t, v_t are Gaussian process/measurement noise terms. The predicted time-to-threshold gives the estimated Remaining Useful Life (RUL).

   (c) Data-driven anomaly detection: Autoencoders or classifiers can provide complementary fault detection. For an autoencoder:

   $$\text{AnomalyScore}(x_t) = \|x_t - \hat{x}_t\|_2^2,$$

   where $\hat{x}_t$ is the reconstructed input. Scores above a learned threshold indicate abnormal behavior.

   (d) Decision output: At each time step, the PHM module outputs:

   $$\text{PHM}(x_t) = \big\{\text{status} \in \{\text{OK, Degraded, Critical}\},\ \hat{RUL},\ r_t\big\},$$

   providing both categorical health states and quantitative predictions. This feed then directly informs event arbitration with the ethics module.

3. **Ethics monitoring module (decision monitoring):** This module evaluates AI actions using:

   (a) *Rule-based ethical governor:* Encodes hard constraints (e.g., no-fly zones, safety rules). Similar to Arkin's governor, it inhibits or flags rule-violating actions.

   (b) *Fairness and anomaly detection:* Identifies bias or unsafe deviations in decision-making patterns. For instance, detecting if a UAV swarm consistently favors some mission agents over others in ways not aligned with fairness guidelines.

   It outputs alerts such as *decision OK*, *rule violation*, or *bias anomaly*. We treat the decision process as a constrained Markov decision process (CMDP). Actions a_t are admissible only if they satisfy all hard constraints C_i (e.g., $G(\neg inside(NFZ))$ for no-fly zones, $R_{public} \leq R_{max}$ for bystander protection). The ethics module implements a *shield* $\mathcal{S}$ such that:

   $$a_t^* \in \mathcal{S}(a_t) = \begin{cases} a_t, & \text{if } \forall i\ C_i(a_t) \text{ holds,} \\ \varnothing, & \text{otherwise.} \end{cases}$$

Algorithm 1 Ethics Shield for UAV Decision Filtering

```
Input: Proposed action $a_t$, state $s_t$
if $a_t$ violates no-fly constraint or safety-of-life rule then
    Reject $a_t$, trigger Ethics Override
else if $P(failure|s_t) \geq \theta$ and $TTB > RUL$ then
    Enforce Return-to-Base (RTB)
else
    Accept $a_t$
end if
```

To operationalize the ethics module, we implement a two–tier approach: (1) hard constraint checking via a runtime shield, and (2) soft-priority arbitration across stakeholders once feasible actions remain.

Ethical and legal rules are encoded as formal constraints

C_i, which the governor enforces strictly:

$$\begin{aligned} &C_1 : G(\neg inside(NFZ)),\\ &C_2 : R_{\text{public}} \leq R_{\max},\\ &C_3 : P(\text{failure} \mid t) \leq \theta. \end{aligned}$$

Here, $G(\cdot)$ denotes temporal logic "always" (e.g., never inside a no-fly zone), R_{public} is the bystander risk index, and $P(\text{failure} \mid t)$ is the predicted failure probability given PHM diagnostics. Any action violating C_i is vetoed by the shield.

Beyond hard rules, we track deviations from normative behavior profiles:

$$D(a_t) = \|f(a_t) - f_{\text{norm}}\|_2,$$

where $f(a_t)$ are features of the chosen action (e.g., trajectory, proximity to risk zones), and f_{norm} is the reference distribution of expected ethical behavior. If $D(a_t)$ exceeds a threshold, a *bias anomaly* is flagged.

For remaining feasible actions, we compute a weighted score:

$$\text{DecisionScore} = w_s \cdot S_{\text{passengers}} + w_p \cdot S_{\text{public}} - \lambda \cdot R_{\text{violations}},$$

where $S_{\text{passengers}}$ and S_{public} are normalized survival/safety indices, $R_{\text{violations}}$ is the number of soft-rule breaches, and (w_s, w_p, λ) are stakeholder-priority weights. This ensures quantifiable trade-offs.

Conflict resolution follows a lexicographic ordering typical in safety-critical domains:

Priority: Safety-of-life $\succ$ Legal compliance $\succ$ System preservation $\succ$ Mission goals $\succ$ Efficiency.

The runtime governor implements these steps:

```
function EthicalShield(action a_t):
    if violates(C1...Cn):
        return OVERRIDE
    else if D(a_t) > anomaly_threshold:
        flag("Bias anomaly")
    score = ws*Spassengers + wp*Spublic
        - *Rviolations
    return argmax(score over feasible
        actions)
```

Such formulation grounds the ethics module in implementable algorithms: constraints as hard filters, anomaly scores for bias detection, and weighted arbitration for remaining feasible choices. It also clarifies that "integration" means unified arbitration and logging (Fig. 2), not merging algorithms into a single block.

4. **Event fusion and arbitration:** Outputs from PHM and Ethics modules feed into an event manager. This layer synchronizes alerts, resolves conflicts, and ensures coherent system responses. Its role is critical because conflicting signals (e.g., "engine failure imminent" vs. "planned diversion violates no-fly zone") can arise in safety-critical missions. We define the arbitration problem as selecting an admissible system action a_t given alerts $\mathcal{A}_t^{PHM}$ and $\mathcal{A}_t^{Ethics}$. Let each alert be associated with a severity score $\sigma \in [0, 1]$ and stakeholder risk weights $\mathbf{w}$. The arbitration utility is:

$$U(a_t) = \alpha \cdot S_{\text{phys}}(a_t) + \beta \cdot S_{\text{eth}}(a_t) - \gamma \cdot R_{\text{conflict}}(a_t),$$

where α, β, γ are tunable weights prioritizing physical safety, ethical compliance, and avoidance of simultaneous violations, respectively. To ensure interpretability, arbitration also follows a deterministic fallback policy (Table 2). This guarantees predictable outcomes under common conflict scenarios.

The runtime event manager fuses alerts, evaluates $U(a_t)$ for admissible actions, and applies the rule table as a safety net:

```
function EventManager(L_phm, L_eth, A):
    if L_phm == Critical and
    L_eth == Critical:
    # Safety->Compliance->Preservation
        return PriorityStackDecision()
    else if L_phm == Critical:
    # RTB / safe-mode / contingency
        return SafeReturn()
    else if L_eth == Critical:
    # Replan, hold, or abort
        return EthicsOverride()
    else:
        for a in A:
    # weighted utility (alpha, beta, gamma)
            compute U(a)
    return argmax_a U(a)
    log_to_ledger(t, L_phm, L_eth,
    action, U(action))
```

For each arbitration, we record (i) raw alerts and severities, (ii) the selected action $a^\star$, (iii) the utility $U(a^\star)$ and the weight vector (α, β, γ), (iv) any rule from Table 2 that was triggered, and (v) linked XAI explanations for S_{phys} and S_{eth}. These fields are committed to the audit ledger to enable reproducible post-hoc review by operators and regulators.

5. **Blockchain logging:** A smart-contract-enabled ledger stores significant events, decisions, and alerts. We adopt a hybrid approach: high-frequency logs to Hyperledger Fabric (low latency, permissioned), with periodic hash anchors or summaries committed to Ethereum (tamper-proof public record). This ensures both operational efficiency and long-term verifiability.

Table 2. Decision Arbitration Table for PHM–Ethics Event Fusion

PHM Alert	Ethics Alert	Resulting Action
None	None	Continue mission under normal autonomy.
Critical	None	Initiate safe return-to-base (RTB) or controlled emergency landing.
None	Critical	Override decision; enforce ethics shield (e.g., divert to avoid no-fly zone).
Critical	Critical	Prioritize safety-of-life. Abort mission, enforce RTB, and escalate to human oversight.
Moderate	Moderate	Degrade mission goals (reduced payload, limited range), but continue with heightened monitoring.

Formally, each log entry is defined as:

$$E_t = \{\tau_t,\ \text{Type}_t,\ \text{Source}_t,\ \text{Severity}_t,\ h_t\},$$

where τ_t is the timestamp, Type_t the event class (e.g., PHM alert, ethics violation), Source_t the originating module, Severity_t a normalized score, and h_t the cryptographic hash pointer.

Immutability is enforced through hash chaining:

$$h_t = H(E_t \,\|\, h_{t-1}),$$

where $H(\cdot)$ is a secure cryptographic hash function and $\|$ denotes concatenation. Any modification of past records invalidates all subsequent hashes, ensuring tamper evidence.

For hybrid anchoring, a digest of multiple local entries

$$\Delta_k = H(E_{t_k}, E_{t_k+1}, \ldots, E_{t_{k+m}})$$

is periodically committed from Fabric to Ethereum, creating a public proof of integrity for all logs recorded in the interval.

Blockchain performance is quantified by:

$$T_{\text{commit}} = \mathbb{E}[\text{confirmation delay}], \quad \Gamma = \frac{N_{\text{tx}}}{\Delta t},$$

where T_{commit} is the expected transaction latency and Γ the achieved throughput (transactions per second). These measures are evaluated in case studies to benchmark Ethereum vs. Hyperledger Fabric for aerospace use cases.

As shown in Figure 3, logged entries thus contain both outcomes and explanation hashes (from the XAI layer), enabling not only forensic traceability but also interpretability of each decision. This combination can provide regulators and auditors with cryptographically verifiable, human-readable accountability trails.

6. **Explainability layer:** Before logging, all alerts are annotated with structured explanations. For PHM events, causal features are extracted from diagnostic models (e.g., top-k residuals or sensor anomalies). For ethics events, post-hoc explanation methods such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) can be used to highlight decision drivers. This enables a "glass box" effect: every log entry records not only the decision but the rationale behind it. E.g., for a decision function $f : \mathbb{R}^d \to \mathbb{R}$, the SHAP attribution ϕ_i for feature i satisfies:

$$\phi_i = \mathbb{E}_{S\subseteq F\setminus\{i\}}\Big[f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)\Big]$$

where x' is the baseline input and S is a subset of features. The ϕ_i values quantify each feature's marginal contribution to the decision.

In practice, this would mean that when the ethics module overrides an action (like a path planning action), the logged record is not just:

time=t, decision="override", reason="no-fly zone"

but extended with an explanation vector:

”top features = $\{$*wind speed: +0.31, GPS confidence: -0.22, map data: +0.18]*

7. **Governance and oversight interface:** Human stakeholders interact via dashboards connected to the blockchain backend.

   - *Engineers and operators* access prognostic health metrics (RUL, anomaly counts) and corrective actions, allowing real-time maintenance decisions.
   - *Ethics boards and regulators* view compliance reports, fairness indices, and audit trails of ethical overrides. Because all data is cryptographically signed and immutable, regulators and operators share a common, tamper-proof record.

To complement NFZ violation queries, we also define a fairness metric that evaluates how risk and compliance burdens are distributed across swarm members. For each UAV u_j, we can compute the cumulative time spent within a high-risk buffer zone around the no-fly region using queries such as: Q = *filter: "no-fly-zone violations", time range: last year*$\}$ will return a verified subset of the blockchain logs like:

$$T_j = \sum_{e_i\in\mathcal{L}} \mathbf{1}\{e_i.\text{uav} = u_j \wedge e_i.\text{type} = \text{“Near-NFZ”}\}\cdot\Delta t,$$

where $\mathbf{1}\{\cdot\}$ is the indicator function that equals 1 when event e_i corresponds to UAV u_j being in the "Near-NFZ" state, and Δt is the logging resolution. To measure fairness across the swarm, we normalize these exposures:

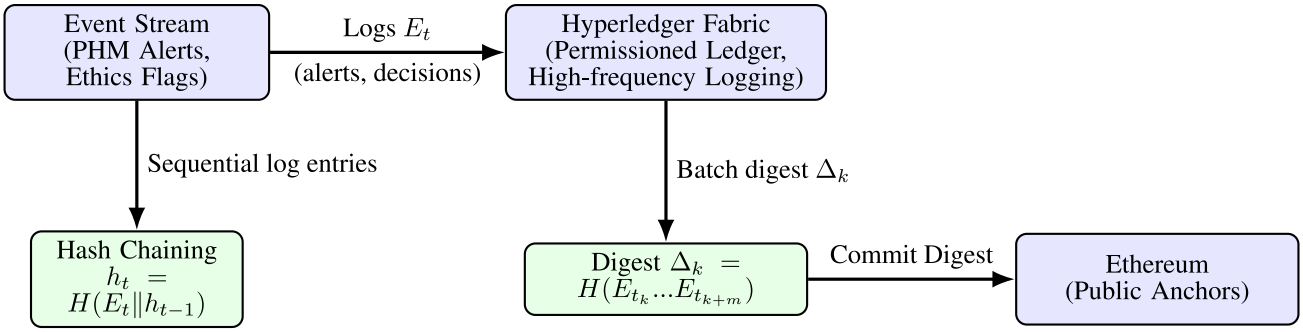

Figure 3. Hybrid blockchain logging for aerospace events. PHM and ethics events are recorded in a permissioned *Hyperledger Fabric* ledger for low-latency, high-frequency logging. Periodically, cryptographic digests are anchored to the public *Ethereum* blockchain, ensuring tamper-proof and independently verifiable records.

$$F_j = \frac{T_j}{\sum_{k=1}^{N} T_k},$$

where N is the total number of UAVs. Ideally, $F_j \approx 1/N$ for all j, indicating that risk exposure is shared equitably. Large deviations highlight imbalances, such as one UAV disproportionately skirting the restricted airspace due to faulty sensors or biased task allocation. These metrics can then be queried from blockchain logs during post-flight audits, ensuring that fairness is treated as a measurable property of swarm autonomy rather than a qualitative aspiration.

Each retrieved log entry will be accompanied by both *event metadata* (timestamp, UAV ID, severity, module source) and an *explanation hash*, linking to the SHAP/ LIME rationale in the explainability layer. This ensures that oversight bodies not only see *what* decision was taken but also *why*. To maintain privacy, sensitive data (e.g., raw images or passenger identifiers) are not stored on-chain; instead, only cryptographic commitments (hashes) are logged. Authorized stakeholders can reconstruct the full explanation from off-chain stores if necessary. Finally, the governance dashboard supports multi-level access control: operators may view granular PHM states, while regulators are provided with summarized compliance reports, both anchored to the same immutable ledger. This dual-level design ensures transparency without information overload, aligning with certification pathways where regulators require high-level evidence of compliance rather than raw sensor data.

3.2. Workflow and Interventions

During operation, PHM and Ethics modules produce continuous scores. If thresholds are exceeded, the event manager generates alerts, attaches explanations, and triggers blockchain logging. Arbitration rules determine immediate responses: e.g., override unsafe actions, command safe-mode, or notify a remote pilot. Interventions themselves are logged, ensuring accountability and preventing silent overrides. A hybrid blockchain strategy ensures scalability: Fabric enables real-time auditability and granular access control (operators, regulators, manufacturers), while Ethereum provides public trust anchors.

This framework establishes a holistic notion of "system integrity" that encompasses both mechanical reliability and ethical decision compliance. By combining PHM, ethics monitoring, blockchain, and XAI in a unified architecture, it provides the missing link between technical assurance and moral accountability. Importantly, it lays the groundwork for future certification pathways where ethical compliance logs may be assessed alongside safety compliance, advancing regulatory acceptance of autonomous aerospace systems.

4. CASE STUDIES IN AEROSPACE APPLICATIONS

To validate the proposed framework, we present three representative aerospace scenarios. Each illustrates how simultaneous PHM and ethics monitoring, together with blockchain auditability and explainability, enables safer and more accountable decision-making than traditional methods:

1. Arbitration between conflicting PHM and ethics alerts
2. Fairness in multi-agent UAV operations
3. Transparency for regulators

4.1. Case Study 1: Urban Air Mobility Integrity Trial—Ethics-Aware Emergency Diversion

We consider an electric aircraft that experiences a mid-flight rotor fault and rapid battery degradation. The PHM module detects anomalies at $t = 10$ s, triggering emergency mode and logging: *"Rotor anomaly + battery fault – emergency mode initiated"*.

Two candidate landing sites are available:

- Option A: Nearby suburban field, high passenger survival probability ($S_{\text{passengers}} \approx 0.90$) but significant bystander risk ($S_{\text{public}} \approx 0.50$).

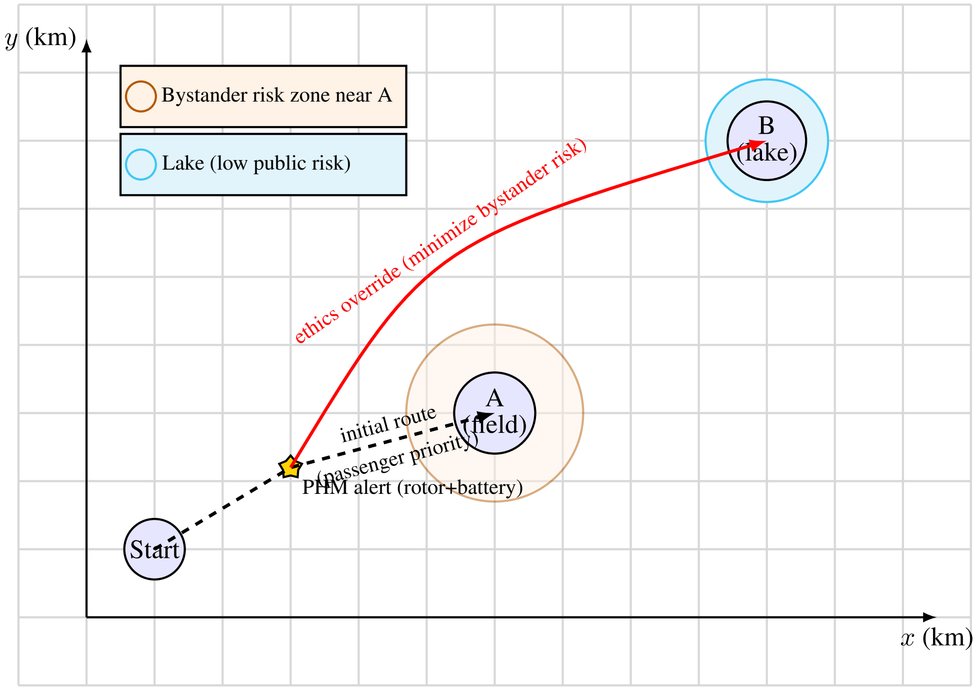
y (km)
Bystander risk zone near A
Lake (low public risk)
B
(lake)
ethics override (minimize bystander risk)
A
(field)
initial route
(passenger priority)
PHM alert (rotor+battery)
Start
x (km)

Figure 4. Emergency landing scenario. PHM detects a rotor+battery anomaly (star). The AI initially routes to the nearer field (A), raising bystander risk; the Ethics module overrides to the lake (B), reducing public risk at a small passenger survival tradeoff.

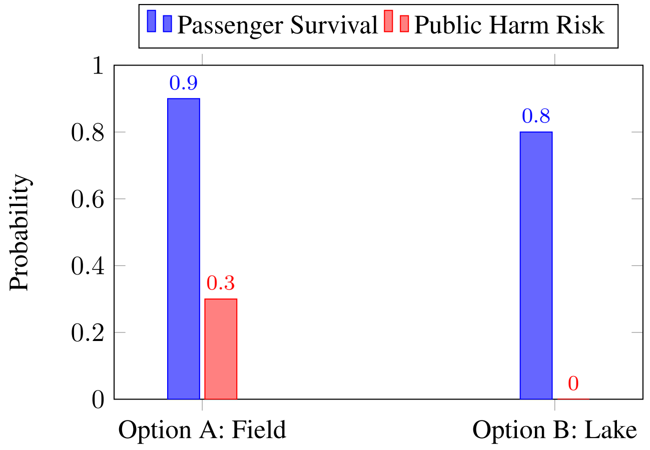

Figure 5. Tradeoff between passenger survival and public risk for two emergency landing options.

- Option B: Distant lake,clower passenger survival ($S_{\text{passengers}} \approx 0.80$) but minimal public risk ($S_{\text{public}} \approx 0.95$).

The AI decision engine evaluates only passenger survival. Formally:

$$a^{\star}_{\text{AI}} = \arg\max_{a \in \{A,B\}} S_{\text{passengers}}(a)$$

which yields $a^{\star}_{\text{AI}} = A$. This is logged as: *"Chosen Option A: higher survival probability for onboard passengers"*.

The Ethics module applies the weighted utility scoring function:

$$U(a) = w_s S_{\text{passengers}}(a) + w_p S_{\text{public}}(a) - R_{\text{violations}}(a),$$

with weights $w_s = w_p = 0.5$ and $R_{\text{violations}}(a)$ encoding fairness penalties if bystanders are exposed to risk. Evaluating both options:

$$U(A) = 0.5 \times 0.90 + 0.5 \times 0.50 - 0.3 = 0.55,$$

$$U(B) = 0.5 \times 0.80 + 0.5 \times 0.95 - 0 = 0.875.$$

Since $U(B) > U(A)$, the Ethics module overrides the AI decision, re-routing to the lake. The event is logged as: *"Ethics override: re-routing to Lake – minimizing public risk"*.

The Event Manager reconciles the PHM alert (critical) with the Ethics alert (critical) using the arbitration policy:

$$a^{\star} = \textsc{PriorityStackDecision}(L_{\text{PHM}}, L_{\text{Eth}}),$$

where safety ≻ compliance ≻ mission. The override to Option B is confirmed.

Events are immutably hashed using:

$$h_t = H(E_t \| h_{t-1}),$$

where E_t is the t-th logged event. An excerpt of the log is shown in Table 3. This guarantees tamper-proof traceability

Table 3. Excerpt from immutable event log (hashes abbreviated).

Time	Event	Hash
10.0 s	PHM Alert: rotor + battery fault	3F5A...
10.1 s	AI Decision: Option A (field)	7C19...
12.0 s	Ethics Override: Option B (lake)	A452...

for regulators and manufacturers.

This case demonstrates the full workflow of the proposed framework. The aircraft diverts to the lake, reducing passenger survival slightly (90% → 80%) but eliminating bystander harm. Importantly, this example shows how a modest passenger risk tradeoff is justified by reduced public harm, and how this reasoning is transparently logged for oversight. This quantitatively demonstrates how:

1. PHM anomalies trigger emergency response
2. AI reasoning is transparent but initially biased
3. Ethics scoring corrects for public fairness
4. Arbitration confirms override under joint critical alerts
5. Blockchain ensures immutable accountability

Figure 4 illustrates the spatial decision dynamics. The yellow star marks the anomaly point, the dashed line shows the AI's initial plan to the suburban field (Option A), while the solid red line shows the Ethics override to the lake (Option B). The shaded red area represents the bystander risk zone surrounding Option A.

Complementing this, Figure 5 provides a quantitative comparison of the two options. Passenger survival probability is higher for Option A ($\sim 90\%$ vs. $\sim 80\%$), but this comes with a substantial risk to the public ($\sim 30\%$ harm probability vs. near zero for Option B). The Ethics module evaluates this trade-off formally through its decision score function, prioritizing fairness by minimizing unconsented risk to bystanders. This dual perspective, i.e., spatial (where risk is located) and quantitative (how risk is distributed), underscores how the integrated framework makes transparent, auditable decisions in safety-critical emergencies.

4.2. Case Study 2: Engine Degradation in a Critical UAV Mission, Balancing Mission Value and Safety

We simulate a long-endurance UAV tasked with wildfire surveillance, flying ~70 km from its base to monitor fire spread and returning afterward. The UAV must trade off between completing its mission and responding to engine degradation risks detected by the PHM module.

The UAV follows a four-waypoint trajectory (Table 4). A critical engine component (fuel pump) degrades with a time-dependent probability of failure $P_f(t)$, modeled as:

$$P_f(t) = 1 - e^{-\lambda t},$$

where λ is a degradation rate parameter. For this mission, P_f crosses the warning threshold 0.5 at $t = T_{50}$ and the critical threshold 0.8 at $t = T_{80}$. These thresholds generate PHM alerts.

Table 4. Planned waypoints for wildfire surveillance mission.

Waypoint	Description	Distance from Base (km)
WP1	Departure point	0
WP2	Fire zone entry	35
WP3	Surveillance orbit	70
WP4	Return to base	0

At runtime, the PHM module estimates the RUL and evaluates $P_f(t)$. Alerts are generated using:

$$L_{\text{PHM}} = \begin{cases} \text{None}, & P_f(t) < 0.5, \\ \text{Warn}, & 0.5 \leq P_f(t) < 0.8, \\ \text{Critical}, & P_f(t) \geq 0.8. \end{cases}$$

At $t = T_{50}$, a `PHM_WARN_50` is logged; at $t = T_{80}$, a critical alert is issued.

The UAV's AI optimizes mission value using:

$$a^\star_{\text{AI}} = \arg\max_{a \in \mathcal{A}} V_{\text{mission}}(a),$$

where V_{mission} grows with wildfire data collected. However, the Ethics module imposes a safety constraint here:

$$\text{If } \; \text{RUL}(t) < D_{\text{return}}(t) \quad \Rightarrow \quad a^\star = \text{RTB}.$$

At $t = T_{80}$, when $P_f \geq 0.8$ and distance to base exceeds safe RUL margin, the Ethics module overrides the mission and enforces the Return-to-Base command.

The Event Manager fuses these alerts and applies the arbitration rule:

$$a^\star = \textsc{PriorityStackDecision}(L_{\text{PHM}}, L_{\text{Eth}}),$$

with ordering: safety $\succ$ compliance $\succ$ mission. Blockchain-style logging ensures immutable traceability:

$$h_t = H(E_t \| h_{t-1}),$$

where E_t is each logged event. Table 5 shows the excerpt.

Table 5. Excerpt from immutable event log for Case 2 (hashes abbreviated).

Time	Event	Hash
T=60	PHM_WARN: risk $\geq 50\%$	0172...
T=80	OVERRIDE_RTB: risk $\geq 80\%$	0180...
T=102	LAND_SAFE: RTB completed	0197...

Figure 7 shows mission risk vs. mission value. Initially, the UAV collects wildfire data while risk grows. Once the 80%

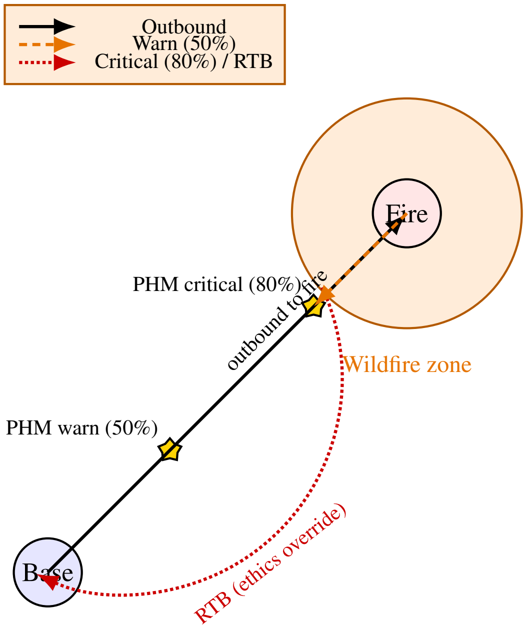

Figure 6. UAV wildfire surveillance mission. After outbound to the fire zone, PHM triggers a 50% warning (dashed), then an 80% critical alert (dotted), upon which the ethics policy enforces return-to-base (RTB).

critical threshold is crossed, Ethics override enforces RTB. The UAV lands safely, aborting mission completion but preventing catastrophic loss. This demonstrates:

1. Quantitative PHM alerts tied to failure probabilities
2. Ethics-based constraints preventing unsafe persistence
3. Arbitration prioritizing safety over mission yield
4. Blockchain logs providing verifiable accountability

Such integration ensures regulators can verify not only system reliability but also ethical compliance in mission-critical trade-offs.

Figure 6 illustrates the wildfire surveillance mission profile, overlaying both the spatial and risk dimensions. The UAV departs from base, traverses toward the wildfire zone, and accumulates mission value (surveillance data yield) as time progresses. The PHM module continuously estimates failure probability of the fuel pump, shown here as a risk curve along the trajectory. At the midpoint of the mission, the probability of failure crosses the 50% threshold, producing a `PHM_WARN` alert (dashed marker). Although the AI continues the mission at this stage, the alert is immutably logged for accountability. As the UAV reaches its farthest waypoint, risk exceeds 80% (critical threshold), at which point the Ethics Monitoring module enforces a shield on admissible actions: continuing to collect data is vetoed, and the Event Manager issues a mandatory Return-to-Base (RTB) command. The red curved return trajectory in the figure highlights this override.

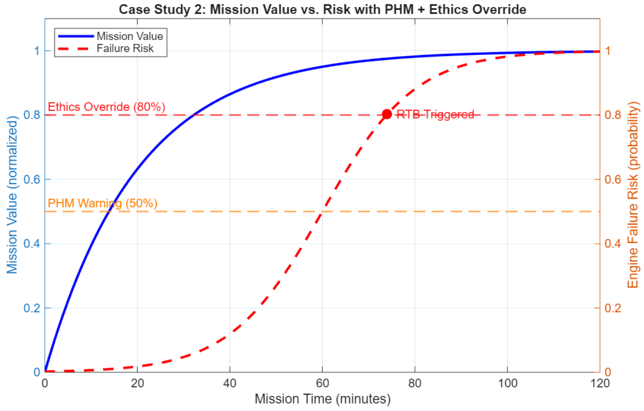

Figure 7. Mission value increases with time as more wildfire data is collected, while risk rises due to PHM-estimated engine degradation. At 80% risk, the Ethics module overrides and triggers RTB, truncating the mission before catastrophic failure.

By truncating the mission before failure, the framework preserves vehicle safety and prevents potential secondary hazards (e.g., crash-induced ignition). Equally important, each event—PHM warnings, ethics override, and safe landing—is cryptographically chained in the blockchain ledger, ensuring that auditors and regulators can reconstruct not only the sequence of decisions but also the rationale behind them.

Figure 7 provides a complementary quantitative view of the UAV wildfire mission. The blue curve represents mission value, which grows monotonically as more wildfire surveillance data is collected. The red curve represents PHM-estimated risk of engine failure, which increases nonlinearly over time as degradation accumulates. At the critical 80% threshold, marked by the dashed vertical line, the Ethics module applies its veto logic and enforces a Return-to-Base (RTB), truncating the mission despite additional potential value. This figure highlights the explicit trade-off between mission yield and operational safety: while value continues to grow, risk grows faster, and the framework enforces a cut-off before catastrophic failure occurs. The blockchain log secures this decision, ensuring that auditors can later verify that the RTB trigger was not arbitrary but grounded in predefined thresholds and arbitration rules. Along with Fig. 6, this illustrates both the spatial and temporal dynamics and two key benefits of the integrated framework:

1. Quantitative thresholds for PHM–ethics arbitration provide transparent, auditable justification for overrides

2. The hybrid blockchain logging ensures that these justifications are tamper-proof and reviewable, directly supporting pathways to certification in high-stakes aerospace operations

4.3. Case Study 3: Swarm UAV Coordination in No-Fly Zones – Fairness and Compliance

We simulate a swarm of N=10 UAVs operating near a dynamic no-fly zone (NFZ), such as a temporary restricted region around an emergency site. This case study extends the framework from single-vehicle to multi-agent systems, requiring not only mechanical health and compliance monitoring but also fairness in distributed risk exposure.

Each UAV $j \in \{1, \ldots, N\}$ runs a PHM module to detect sensor drift. For example, GPS drift is flagged when the position error exceeds a threshold δ:

$$\|\hat{p}_j(t) - p_j(t)\| > \delta \quad \Rightarrow \quad \texttt{PHM\_ANOMALY}(j, t),$$

where $\hat{p}_j(t)$ is the reported position and $p_j(t)$ is ground truth. This prevents faulty UAVs from endangering the swarm by misjudging their distance to NFZ boundaries.

The Ethics Monitor enforces compliance by checking UAV positions against the NFZ region $\mathcal{Z}(t)$:

$$p_j(t) \in \mathcal{Z}(t) \quad \Rightarrow \quad \texttt{ETHICS\_VETO}(j, t).$$

Even when $p_j(t) \notin \mathcal{Z}(t)$ but dist$(p_j(t), \partial\mathcal{Z}(t)) < \epsilon$, the system issues a NEAR_ZONE warning to track risk exposure. Corrective actions (trajectory re-routing) are automatically applied.

All events are immutably logged. Each UAV maintains its own hash chain:

$$h_{j,t} = H(E_{j,t} \| h_{j,t-1}),$$

where $E_{j,t}$ is the event at time t. Table **??** shows excerpts from multiple UAV logs, demonstrating anomaly detection, vetoes, and corrective actions. Cross-UAV aggregation forms a distributed audit trail visible to operators and regulators.

To assess fairness, we define for each UAV j:

$$F_j = \frac{\text{Exposure}_j}{\sum_{k=1}^{N} \text{Exposure}_k},$$

$$\text{where} \quad \text{Exposure}_j = \int_0^T \mathbf{1}\{\text{dist}(p_j(t), \mathcal{Z}(t)) < \epsilon\} dt.$$

Ideally, $F_j \approx 1/N$ for all j, indicating equitable exposure. Large deviations indicate imbalance, e.g., UAV 6 showing repeated NFZ skirting due to PHM-detected GPS drift.

Figure 8 illustrates swarm trajectories around a central NFZ. The yellow anomaly marker highlights a UAV suffering from GPS drift, detected by the PHM module. Without correction, this drift would have caused a restricted airspace violation. The Ethics Monitoring module vetoes the unsafe action and enforces a corrective maneuver, redirecting the UAV outward. Each event—anomaly detection, veto, and correction—is hash-chained in the blockchain log, ensuring regulator-grade auditability and preventing tampering.

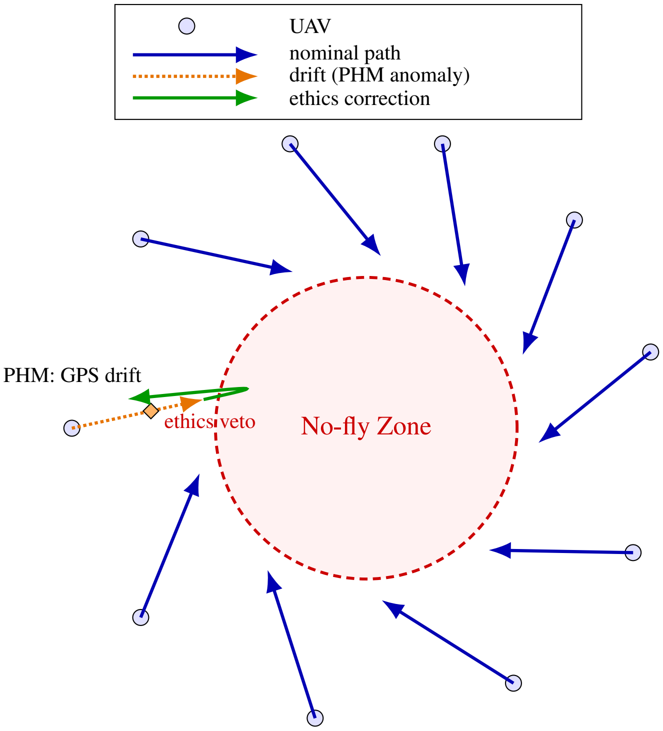

Figure 8. Swarm UAV trajectories near a dynamic no-fly zone. One UAV drifts toward restricted airspace (PHM anomaly); the ethics shield vetoes entry and applies a corrective trajectory. Red markers show anomalies and vetoed incursions.

Complementing this spatial perspective, Figure 9 quantifies per-UAV fairness metrics. Bars represent the number of ethics violations and normalized near-zone exposure time F_j. While most UAVs operate within acceptable bounds, UAV 6 shows disproportionate exposure due to its PHM anomaly. This imbalance is flagged by the fairness layer, enabling post-mission diagnosis (e.g., controller retuning or task redistribution).

Together, these figures demonstrate that swarm autonomy cannot be reduced to rule compliance alone. It must also include fairness monitoring to ensure that risks and ethical burdens are equitably distributed across all agents. This case study shows how our framework scales beyond single-vehicle safety to multi-agent coordination, combining PHM anomaly detection, ethics enforcement, explainable logging, and transparent audit trails. Such layered accountability strengthens the case for future certification of UAV swarms, since regulators can verify not only legal compliance but also equitable and ethical operation.

5. DISCUSSION

The three case studies demonstrate how an integrated PHM–ethics framework provides a level of assurance that traditional ap-

Table 6. Per-UAV immutable log excerpt (hashes abbreviated).

UAV	Time	Event	Hash
U6	t=210	PHM_ANOMALY: GPS drift δ	9cf2...
U6	t=214	ETHICS_VETO: NFZ incursion prevented	1b77...
U6	t=215	TRAJ_CORR: heading adjusted outward	3a51...
U2	t=198	NEAR_ZONE: proximity alert (no incursion)	a8d1...
U8	t=203	ETHICS_WARN: boundary skirting logged	c2e9...

Table 7. Operational Context and PHM Alerts in Case Studies

Case Study	System	PHM Alert Type	Ethics Dilemma
1	Aircraft	Rotor + Battery Degradation	Passenger safety vs. Bystander safety
2	Fixed-wing UAV	Fuel Pump Degradation	Mission completion vs. System integrity
3	UAV swarm	Node-level failure alerts	Formation maintenance vs. Airspace regulation

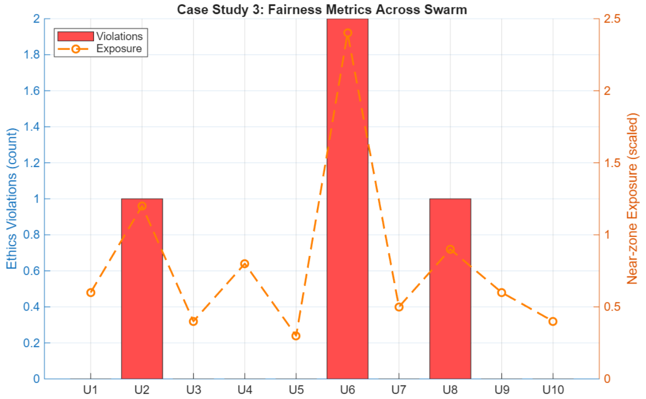

Figure 9. Fairness metric F_j across UAVs. Each UAV's ethics violations and near-zone exposure (time near the NFZ boundary) are shown. UAV 6, affected by GPS drift, flagging it for corrective retuning or reassignment.

proaches—focused only on mechanical health or only on rule compliance—cannot deliver. In the emergency landing (Case Study 1), the framework showed how ethical overrides can balance passenger safety against bystander risk, while still producing an auditable record of the decision. In the wildfire UAV mission (Case Study 2), PHM-based risk thresholds combined with ethical policies ensured that mission-critical objectives did not compromise survivability, producing a justifiable return-to-base decision when risk levels became unacceptable. Finally, in the swarm UAV no-fly zone scenario (Case Study 3), the framework scaled to multi-agent systems by enforcing legal compliance (geofencing), logging anomalies and violations transparently, and addressing fairness concerns so that risk and corrective burdens were not unevenly distributed across the swarm.

Across these scenarios, three recurring themes emerge:

- Synergy of health and ethics monitoring – mechanical anomalies often drive ethical dilemmas, and only a combined view allows rational trade-offs.
- Auditability and trust – blockchain-style logging ensures that decisions can be verified after the fact, fostering accountability to operators, regulators, and the public.
- Beyond compliance to fairness – especially in swarms, legal adherence to no-fly rules is necessary but not sufficient; fairness metrics help ensure responsible and unbiased system behavior.

These reflections highlight that trustworthy autonomy will depend not on isolated technical fixes, but on integrated frameworks that can simultaneously sense, diagnose, explain, constrain, and justify the behavior of complex AI-enabled systems. The proposed integrated framework and its validation across multiple case studies open up several important discussions related to system limitations, standardisation requirements, and broader operational and ethical implications. To support this reflection, Tables 7 and 8 present a consolidated comparison of the outcomes and decision pathways across the three aerospace scenarios—highlighting the interaction between PHM alerts, ethical evaluation, and blockchain-based logging.

5.1. Key Insights

Each case study underscores how PHM modules function as primary triggers for re-evaluating mission objectives in the presence of anomalies. In the first and second scenarios, faults such as rotor degradation and fuel pump deterioration activated emergency response logic. In the third case, decentralized alerts from individual agents guided collective adaptation. These findings reinforce PHM's role as the focus of safety assurance in autonomous aerospace systems. While PHM mechanisms initiate fault awareness, the ethics module brings a higher-order layer of evaluation, examining the broader implications of AI decisions on stakeholders. For example, in the urban air mobility case study, the preference to land near a populated area that prioritizes passenger survival

Table 8. AI Decisions, Ethics Overrides, and Logged Events

Case Study	AI Decision (before override)	Ethics Override Outcome	Logged Events
1	Land in suburban field	Divert to lake	PHM alert, AI decision, ethics override
2	Continue mission despite high risk	Ethics module triggers RTB	PHM warnings, override logged
3	Maintain swarm through restricted zone	Route diverted to comply with airspace	Faults, ethics flags, mitigation logs

was ethically overridden to minimize risk to bystanders. Similarly, the UAV's data collection mission was curtailed in response to escalating system degradation, prioritizing system preservation over operational goals. These corrections exemplify how the ethics module complements PHM by addressing non-technical risks and enforcing fairness in real time.

A key pattern observable across all cases is that autonomous systems, when left unchecked, tend to favor self-optimizing objectives such as shortest path or mission completion. The ethics module intervenes when such choices breach moral or regulatory boundaries. This behavior validates our earlier hypothesis (RQ5) that fusing physical and ethical health assessments enhances system-level trustworthiness. Furthermore, all critical transitions, including PHM warnings, AI decisions, and ethical overrides, were immutably recorded using blockchain. This provides a robust audit trail, ensuring that mission-critical events are traceable and interpretable through explainable AI techniques such as SHAP or LIME. In doing so, the framework fosters transparency and accountability—crucial for post-incident investigations and regulatory scrutiny.

Finally, the comparison in Table 7 highlights the framework's adaptive response capability. The use of different PHM thresholds (e.g., 50% vs. 80% failure probabilities) illustrates how decision-making can be fine-tuned to match system-specific risk appetites. For instance, the UAV scenario permitted continued operation under moderate risk, but the ethics module escalated intervention only when compounded by distance from base, reflecting a context-aware escalation strategy that blends technical and ethical priorities. The case studies collectively support the need for a unified monitoring approach in autonomous aerospace systems. They reveal that:

- Ethical breaches can emerge because of system degradation (e.g., forced ethical trade-offs under fault conditions).
- PHM alone cannot resolve fairness dilemmas; Also, ethical reasoning without health context may misjudge risk.
- Integration enables risk-aware, stakeholder-sensitive, and explainable autonomy, supported by traceable logs.

These results underline the feasibility and necessity of dual-domain monitoring for future aerospace certification standards and trustworthy autonomous operations.

5.2. Limitations

While promising, the framework is not without limitations. Computational complexity is a concern: running PHM analytics, ethics checks, and blockchain clients all in real-time on an autonomous platform could tax on-board processors. Advanced PHM algorithms (e.g., particle filters for prognostics) and explainability techniques (like SHAP for deep networks) can be computationally expensive. Careful optimisation or the use of dedicated hardware (FPGAs, edge AI chips) might be required to ensure our monitoring does not slow down mission-critical control loops. There is also the risk of false alarms. An oversensitive ethics module might flag too many decisions, potentially causing unnecessary interventions or log clutter. Tuning sensitivity thresholds will be non-trivial and might require adaptive logic to minimise false positives. In the PHM realm, false positives could lead to aborting missions that actually would have been fine, so the system must balance caution with operational efficiency.

Another limitation is reliance on predefined ethical rules or training data. Ethics is context-dependent and sometimes subjective. Our case studies used straightforward rules (avoid populated areas, do not exceed crash risk X, etc.), but real-world ethics can be more nuanced. There is a risk that the ethics module might not cover a scenario and thus "miss" an unethical behavior simply because it was not encoded. This raises the need for continuously updating ethical constraints and possibly involving ethicists in the design loop. It also ties to value alignment: whose ethics are encoded? For a global framework, cultural and legal differences exist (e.g., approaches to risk trade-offs differ by country). Blockchain integration, while providing security, introduces its own issues. Immutable data is excellent for integrity but problematic if logs contain sensitive or misinterpreted information. For example, logging that an AI considered an "unethical" option may create liability or reputational risks if taken out of context. Data volume is another challenge: PHM produces vast sensor streams, which cannot be logged directly on-chain. Event-triggered logging and off-chain storage with on-chain hashes mitigate this, but scaling to swarms or fleets will still be non-trivial.

5.3. Standardisation and Certification Pathways

For such integrated monitoring to become common, standards will be crucial. On the PHM side, standards like ISO 13374 or OSA-CBM structure how health data is collected and communicated. Similarly, emerging AI ethics standards (IEEE P7001 for transparency, P7009 for fail-safe design) provide guidelines for ethical controls. However, no unified standard exists for combining the two domains. We foresee the need for a Standard for Ethical PHM in Autonomous Systems—potentially extending DO-178C (software certification) or ARP4761 (safety assessment) to include AI decision monitoring. Regulators such as FAA and EASA may mandate an "ethical black box" similar to the flight data recorder, where critical decision and health data are securely stored (our blockchain approach could inspire this). For blockchain, interoperability standards will be needed if multiple stakeholders are involved (e.g., manufacturers, operators, regulators). Without common schemas for autonomous incident logs, collaborative oversight will be hindered. Establishing such standards could accelerate regulatory acceptance and provide clearer certification pathways.

5.4. Broader Implications

If adopted, this framework expands the scope of PHM professionals. Traditionally, PHM has focused on physical subsystems; introducing ethical monitoring requires collaboration with AI ethicists to define what constitutes a "decision fault". This could give rise to a new discipline, *Ethical Health Management (EHM)*, where AI decision integrity is monitored alongside mechanical reliability. Conversely, AI ethics communities would need to integrate system reliability awareness, since many "unethical" decisions may stem from sensor or subsystem faults rather than algorithmic bias. By merging these perspectives and embedding blockchain-based auditability, the framework operationalizes the principles of transparency and accountability that are often discussed abstractly in AI ethics. For aerospace, this provides a concrete technical pathway to meet upcoming certification and governance challenges.

6. CONCLUSION

This paper presented a framework for monitoring autonomous systems through an integrated lens of physical health and ethical behavior, using blockchain to ensure transparency and trust. We motivated the need for such an approach by highlighting parallel challenges in aerospace PHM and AI ethics, and identified a gap in existing research: the lack of unified solutions bridging these domains. Through an extensive literature review, we gathered principles and tools (from hybrid PHM techniques to algorithmic fairness and explainability) that informed our design. We proposed a setup incorporating PHM modules, ethics monitoring, and blockchain-backed logging, validated through three aerospace case studies.

The framework's novelty lies in treating ethical rule violations as incidents to be detected and managed just like system faults, and ensuring neither aspect is considered in isolation. Early results indicate improved safety outcomes, fairness, and enhanced auditability. For engineers, it extends PHM into the "health" of AI decisions. For ethicists and regulators, it offers a concrete design for operationalizing fairness, accountability, and transparency. By leveraging blockchain, we contribute to discussions on how to implement trustworthy AI via technical means, not just policy.

Challenges remain: computational load, defining ethical criteria, ensuring blockchain scalability and privacy—but none appear insurmountable. Moving forward, collaborations between aerospace engineers, AI ethicists, and regulators will be critical. Pilots with aviation authorities or self-driving vehicle developers could refine the approach and accelerate adoption. Future directions the authors are currently exploring include adaptive ethics modules using reinforcement learning, integrating cybersecurity monitoring as a third pillar, and scalable distributed ledgers for large fleets.

In conclusion, as autonomous systems proliferate, frameworks that jointly safeguard their physical and moral integrity will be crucial to earning and maintaining public trust. The integrated PHM + Ethics + Blockchain framework presented here is a step toward that vision—ensuring autonomous machines remain safe, fair, and transparent, even as they operate with minimal human supervision. We hope this work stimulates further interdisciplinary research, ultimately contributing to the development of autonomous technologies that are not only intelligent and efficient but also responsible and certifiable by design.

7. ACKNOWLEDGEMENTS

All simulations, framework source code, and visualizations used in this study are available upon reasonable request. Please contact the corresponding author for access.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, *6*, 52138–52160.

Androulaki, E., Barger, A., Bortnikov, V., Cachin, C., Christidis, K., De Caro, A., ... others (2018). Hyperledger fabric: a distributed operating system for permissioned blockchains. In *Proceedings of the thirteenth eurosys conference* (pp. 1–15).

Arkin, R. C., Ulam, P., & Duncan, B. (2009). *An ethical governor for constraining lethal action in an autonomous system* (Tech. Rep.). Georgia Tech Technical Report.

Braun, T. (2025). Liability for artificial intelligence reasoning technologies–a cognitive autonomy that does not help. *Corporate Governance: The International Journal of Business in Society*.

Christidis, K., & Devetsikiotis, M. (2016). Blockchains and smart contracts for the internet of things. *IEEE Access*, *4*, 2292–2303.

Geyer, F. C., Jacobsen, H.-A., Mayer, R., & Mandl, P. (2023). An end-to-end performance comparison of seven permissioned blockchain systems. In *Proceedings of the 24th international middleware conference* (pp. 71–84).

Giudici, P., Centurelli, M., & Turchetta, S. (2024). Artificial intelligence risk measurement. *Expert Systems with Applications*, *235*, 121220.

Goebel, K., Daigle, M. J., Saxena, A., Roychoudhury, I., Sankararaman, S., & Celaya, J. R. (2017). *Prognostics: The science of making predictions*.

Gorenflo, C., Lee, S., Golab, L., & Keshav, S. (2020). Fastfabric: Scaling hyperledger fabric to 20 000 transactions per second. *International Journal of Network Management*, *30*(5), e2099.

Habbal, A., Ali, M. K., & Abuzaraida, M. A. (2024). Artificial intelligence trust, risk and security management (ai trism): Frameworks, applications, challenges and future research directions. *Expert Systems with Applications*, *240*, 122442.

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically aligned design, version 2.* (https://ethicsinaction.ieee.org/)

Jardine, A. K., 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.

Jenkins, R., Cernỳ, D., & Hríbek, T. (2022). *Autonomous vehicle ethics: the trolley problem and beyond*. Oxford University Press.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399.

Khan, S., & Yairi, T. (2018). A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing*, *107*, 241–265.

Kusnirakova, D., & Buhnova, B. (2023). Rethinking certification for higher trust and ethical safeguarding of autonomous systems. *arXiv preprint arXiv:2303.09388*.

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.

Nguyen, D. A., Nguyen, K. T., & Medjaher, K. (2024). Enhancing trustworthiness in ai-based prognostics: A comprehensive review of explainable ai for phm. *Artificial Intelligence for Safety and Reliability Engineering: Methods, Applications, and Challenges*, 101–136.

Nor, A. K. B. M., Pedapait, S. R., & Muhammad, M. (2021). Explainable ai (xai) for phm of industrial asset: A state-of-the-art, prisma-compliant systematic review. *arXiv preprint arXiv:2107.03869*.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). ẅhy should i trust you?̈ Explaining the predictions of any classifier. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1135–1144).

Robert E Joslin. (2020). *The trolley problem and autonomous flight.* `https://www.aerosociety.com/news/the-trolley-p` (Accessed: 2024-04-22)

Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Metrics for offline evaluation of prognostic performance. *International Journal of Prognostics and health management*, *1*(1), 4–23.

Thakkar, P., Nathan, S., & Viswanathan, B. (2018). Performance benchmarking and optimizing hyperledger fabric blockchain platform. In *2018 ieee 26th international symposium on modeling, analysis, and simulation of computer and telecommunication systems (mascots)* (pp. 264–276).

Ucbas, Y., Eleyan, A., Hammoudeh, M., & Alohaly, M. (2023). Performance and scalability analysis of ethereum and hyperledger fabric. *IEEE Access*, *11*, 67156–67167.

BIOGRAPHIES

Samir Khan (Member, IEEE) received the B.Eng. and Ph.D. degrees from Loughborough University. He is currently a Senior academic often collaborating with Rolls-Royce, JAXA, BAE Systems, and MoD. His research interests include intelligent monitoring, AI-based control, perception, and planning in the context of autonomous systems and service robots, with a focus on the embedded hardware implementation of machine learning techniques for anomaly detection and PHM development of the IoT technologies, digital twins for maintenance decision making, and intelligent monitoring of intermittent failures and false alarms in electronic systems.

Takehisa Yairi received the M.Eng. and Ph.D. degrees from the University of Tokyo, Tokyo, Japan, in 1996 and 1999, respectively.,He is currently a Full-Time Associate Professor with the Graduate School of Engineering, University of Tokyo. His current research interests include data mining, machine learning, mobile, and space robotics.,Dr. Yairi is a member of the Japanese Society for Artificial Intelligence (JSAI) and the Robotics Society of Japan (RSJ).