WHO HOLDS THE RECORD?

The Case for Independent Evidence in Physical AI

A Foundational Thesis by Roboblackbox

October 2026

ABSTRACT

Robots are becoming software-driven systems that continue to change after deployment. Their behavior depends on software, models, configurations, and physical environments that evolve over time.

Yet the evidence of how these systems operate remains distributed across manufacturers, facility controllers, cameras, and operator records. Each preserves part of the history, often for a different purpose and under different control.

We argue that autonomous physical systems require a distinct infrastructure responsibility: **independent preservation of operational evidence.**

The systems making operational decisions and the system preserving evidence of those decisions should not have to be controlled by the same party.

As physical AI scales, we believe the ability to independently examine how autonomous machines operate will become essential. This paper explains why, what an independent record requires, and what would challenge our thesis.

---

1. The world is changing

Industrial machinery has always required maintenance, modification, and reassessment. Established engineering practices help ensure that changes are understood and appropriate safeguards remain in place.

Autonomous robotic systems intensify this challenge.

A robot's behavior increasingly depends on software, perception models, navigation maps, configurations, and interactions with its environment. These dependencies can change throughout its operational life.

A manufacturer releases a navigation update. An integrator changes a fleet configuration. An operator modifies a restricted area. A warehouse introduces a new pedestrian crossing.

The machines continue operating, but the conditions defining their behavior have changed.

In robotic applications, software decisions can become physical actions. Failures can cause bodily injury, property damage, and interruption of facility operations.[1] Correcting the software afterward cannot undo an injury or erase a period of lost production. When an event leads to a dispute or claim, parties with different interests may need to examine the same operational history. We believe these stakes strengthen the case for evidence preserved before the outcome is known, with its sources, integrity, and limitations open to examination.

Testing, simulation, risk assessment, validation, and certification remain essential. They establish important properties of a system and its intended application. Changes may require renewed evaluation.

But evaluating whether a system is suitable for operation and preserving evidence of how it actually operated are different responsibilities.

Validation helps establish whether a robot is suitable for deployment. The operational record preserves what the recorder received from participating systems afterward.

Validation and reassessment continue throughout the equipment lifecycle. Operational evidence serves a complementary purpose: preserving information needed to examine what participating systems reported as the operation evolved.

Consider a warehouse six months after commissioning.

Several robots have received software updates. Their maps have changed. New equipment has been installed, and operating procedures have been revised.

An unexpected event occurs.

Someone asks what changed, what the machines reported, and what evidence remains of the sequence.

The original validation documents still matter. So do the latest configuration records.

But neither necessarily contains the operational history needed to answer those questions.

That history must come from the systems operating at the time.

And those systems may no longer retain it.

2. The evidence problem

A robotic operation does not have one source of evidence.

It has many.

A robot manufacturer maintains logs describing the states, positions, and faults its systems publish.

Facility controllers and PLCs expose information about equipment states, safeguarding devices, interlocks, and other physical systems.

Cameras may preserve visual evidence of the environment, while separate perception pipelines produce estimated observations derived from that video.

Operators maintain records of rules, configurations, operating conditions, and approved changes.

Each source describes a different part of the operation.

Each may use a different clock, follow a different retention policy, and operate under a different organization's control.

Some preserve extensive histories. Others expose information only briefly.

These are not necessarily poor records. They were created for different purposes.

A manufacturer's logs help engineers diagnose and improve a product. Facility controllers perform operational and protective functions. Cameras capture visual information. Operators document decisions and responsibilities.

But none of these sources necessarily describes the operation as a whole.

The unit of investigation is the operation, not the individual robot.

An event involving one robot may be explained by that robot's own records. Other events require evidence from the surrounding facility, particularly when multiple machines, controllers, and operating rules interact.

When something happens, the operator may need to reconstruct an event from several accounts.

A robot reports that it stopped. A facility controller publishes the state of a safeguarding device. A camera may show activity near the machine. An operator declaration describes the rule that was supposed to apply.

Understanding the event may require all four.

And those accounts may not agree.

The robot's clock may differ from the controller's. A camera may not cover the relevant area. An upstream system may omit a state that an investigator later considers important.

Without a durable cross-system record, reconstruction can become a process of requesting exports, recovering histories, comparing timestamps, and identifying missing evidence.

Sometimes the necessary information is no longer available.

The operator remains responsible for understanding the operation but may depend on systems controlled by several other parties to reconstruct its history.

We call the resulting burden **the cost of trust**: the effort required to establish a shared, defensible understanding of operational events when relevant evidence is distributed across systems and organizations.

That cost can appear as engineering investigation, evidence retrieval, reconciliation between sources, or disputes about what available records establish.

We have not yet measured that cost across operational sites. Whether independent evidence can reduce it is a question we intend to investigate.

The party responsible for the operation does not necessarily hold an independent history of it.

---

3. The missing infrastructure

Robotic operations depend on several kinds of infrastructure, each serving a different purpose.

Control infrastructure makes machines operate and implements required control and protective functions.

Observability infrastructure helps engineers understand performance, availability, faults, and operational conditions.

Evidence infrastructure preserves an attributable history of what participating systems published and the recorder received, so the operation can be examined later.

These responsibilities can overlap.

Existing technologies already preserve substantial operational histories. ROS bag recorders, industrial historians, and observability platforms can retain valuable evidence. Some also support strong access controls and integrity mechanisms.

Our argument is not that existing systems are incapable of recording data.

It is that independent preservation of evidence should be treated as a primary infrastructure responsibility, rather than merely a feature of systems built for another purpose.

The distinction is not what technology can store the data. It is whether independent custody, cross-system provenance, evidentiary integrity, and separation from operational control are requirements of the record itself.

Protected storage and independently verifiable evidence serve different purposes. Retention controls help preserve stored records. Cryptographic commitments held outside the custodian's administrative control can allow reviewers to check supplied records against a previously witnessed history, within a defined verification scope.[2][3]

Existing platforms can be combined to provide these properties; a purpose-built recorder makes their integration and continued verification its primary responsibility. Whether that specialization reduces the operator's integration and governance burden is a hypothesis to test.

Preserving operational evidence has established precedents. Maritime voyage data recorders retain sequential information from ship equipment and command-and-control systems for subsequent

investigation.[4] They preserve information for examination; they do not determine an investigator's conclusion.

Retention matters independently of collection. In February 2024, the NTSB reported that overwritten cockpit voice recordings had hampered at least 14 investigations since 2018.[5] Information had been generated, but relevant recordings were no longer available when investigators needed them.

These precedents use different custody and regulatory arrangements. They do not establish a requirement for robotic recorders. They demonstrate why preserving usable operational evidence deserves deliberate engineering attention.

That responsibility spans the whole operation.

A facility may contain machines from several manufacturers, different generations of software, independent controllers, and physical sensors observing the same environment from different perspectives.

The relevant history extends beyond any individual product.

An operational record should remain useful when the systems being examined have changed, failed, or become unavailable.

It should preserve source identities, timestamps, coverage limitations, and evidence integrity without participating in the systems responsible for controlling physical machinery.

Control determines what machines do. Observability helps engineers understand how systems are performing. Evidence preserves what was published and received so that others can examine it later.

We believe that third responsibility deserves its own architectural foundation.

4. Our thesis: independent operational evidence

The problem has a shape, and the shape suggests a principle.

The systems that make operational decisions and the system that preserves evidence of those decisions should not have to be controlled by the same party.

NIST's *AI Risk Management Framework* describes independent review as a way to improve testing and reduce internal biases and potential conflicts of interest. It recommends testing before deployment and regularly during operation.[6]

This is not a statement about distrust. Manufacturers have strong reasons to maintain accurate logs, and many already do. It is a statement about the separation of responsibilities.

We believe this principle has an operational evidence requirement. When a question concerns past operation, reviewers should be able to examine an attributable history without depending exclusively on the systems being examined. Preservation supports independent judgment; it does not replace it.

A record maintained exclusively by the system being examined remains dependent on that system's collection practices, retention policies, and accessibility.

When an operation includes several vendors and facility systems, no individual manufacturer necessarily has visibility across the entire environment.

The record of the operation should therefore be available under the custody of the party whose responsibility spans that environment.

In many industrial deployments, that party is the operator.

The operator should be able to retain and examine relevant evidence without depending exclusively on the continued availability or cooperation of every system's original provider.

This does not require excluding manufacturers from the record. Nor does it mean the operator must own every source of evidence.

It means independent custody should be possible.

Independence and accountability

Financial auditing offers a useful precedent.

A company maintains its accounting records, prepares financial statements, and remains responsible for its financial reporting. Independent auditors examine evidence and provide an assessment that does not depend solely on management's assertions.

The importance of this separation is reflected in established auditing standards. The Public Company Accounting Oversight Board requires auditors to remain independent of the companies they audit and to evaluate the reliability of evidence supporting their conclusions.[7][8]

We believe autonomous physical operations need a related separation of responsibilities.

Manufacturers and operational systems should not be the exclusive custodians of evidence used to examine their behavior. The operator should be able to preserve an independent, cross-system history, while authorized third parties should be able to evaluate the integrity and provenance of that history.

In this analogy, the recorder corresponds to a tamper-evident audit trail that an authorized reviewer can examine. **It is not the auditor and does not provide an assurance opinion.** Judgment remains with the people examining the evidence.

Nor does operator-controlled custody make the operator a disinterested party. An operator may have interests in the outcome of an investigation, just as a manufacturer may.

Custody of the underlying evidence and independent verification of its integrity need not reside with the same party. An operator can retain the underlying records while commitments are held by multiple independently administered services. The commitment format and associated metadata must be designed to protect confidentiality.

Where implemented, this separation can allow reviewers to check disclosed records against previously witnessed commitments without relying solely on the operator's, manufacturer's, or recorder provider's assurances.[2][3] Verification must state the recording identity, record range, checkpoints, and witness assumptions it relies on. Multiple storage locations alone are not a guarantee of consistent histories.

Such mechanisms cannot, by themselves, establish complete collection or disclosure, recover deleted evidence, or establish the physical time of recorded events. The record should support examination regardless of whose account it supports or challenges.

Independent recording does not replace independent judgment. It makes evidence available for that judgment.

What makes a record independent?

We believe an independent operational record requires five properties.

Operator-controlled custody. The responsible operator governs retention and access to the preserved evidence without depending exclusively on the systems being examined.

Separation from operational control. The recorder consumes authorized publications without directing machine behavior or participating in protective functions.

Source attribution and provenance. Robot publications, facility states, camera-derived observations, and operator declarations retain their distinct origins.

Evidentiary integrity. Changes to retained records can be identified within the guarantees of the recording and verification mechanisms. Independent verification should be possible without relying solely on the custodian's assertions.

Coverage transparency. Missing publications, interrupted feeds, conflicting timestamps, and known uncertainties remain visible. Coverage must be assessed against declared expectations about available sources and their expected publications. A passive recorder cannot establish the existence of information it never received.

Receipt order is not necessarily event order. Where the available evidence cannot establish cross-source ordering or the cause of a collection gap, the record should leave those questions unresolved.

These properties must be implemented and verified. They are not automatic guarantees of installing a recorder.

There is also an important limit.

A robot can publish an inaccurate position. A perception system can misclassify an object. A controller can expose incomplete information.

An independent recorder may faithfully preserve those publications without establishing that they accurately represent physical reality.

Record integrity is not source accuracy. Source accuracy is not causal proof.

An independent record does not certify safety, determine fault, or replace engineering judgment.

It preserves evidence that qualified people can investigate.

When two sources disagree, the record should preserve both. When evidence is missing, it should identify the gap rather than substitute an assumption.

The purpose of independent evidence is not to eliminate uncertainty. It is to preserve what is known and make what remains unknown visible.

5. What this makes possible

An independently preserved history can support comparisons across changes and reconstruction of operational events. The following synthetic examples illustrate both uses.

In each, a finding should remain traceable to received publications, relevant operator declarations, the comparison method, and source coverage. Where evidence is missing, the unanswered question should remain explicit.

These examples make the proposed evidence model inspectable. They do not establish successful field collection, independently verified integrity, or customer benefit.

A warehouse changes its navigation map

A synthetic warehouse operates 23 autonomous mobile robots from two models.

The operator has declared a restricted aisle and documented the applicable operating rules.

Twelve robots receive a map update.

Before the update, the retained sample record contains no reported movement matching the defined restricted-aisle condition during the selected comparison period.

After the update, publications from several updated robots match that condition.

A later map correction is followed by a period with no matching publications among those received.

Operating exposure and available source coverage must be stated alongside these comparisons.

What does the record establish?

It shows that a reported configuration change preceded a change in published behavior. A subsequent correction preceded another change.

It preserves the entries behind those observations.

It does not establish that the first update caused the movement or that the reported positions perfectly represent physical reality.

Those conclusions require further investigation.

The operator can examine a change across time without relying exclusively on a manufacturer's retrospective account.

When evidence is missing

Consider another event in the same synthetic facility.

A facility controller publishes a safeguarding interruption state. The robot reports a stop and later a return to motion. However, the recorder has no available publication establishing the reset sequence.

The record should not claim that a reset occurred.

It should not claim that no reset occurred.

It should preserve the published sequence and identify the missing evidence.

An independent record does not replace the judgment of a safety specialist. It provides an attributable history to examine alongside other evidence and engineering assessments.

The same principle applies to other investigations.

Operators can examine behavioral changes. Manufacturers can investigate evidence shared by customers. Integrators can review commissioning and configuration changes. Safety specialists can examine operational publications alongside existing validation processes.

These parties do not need to share every interpretation.

They need evidence whose origins and limitations they can examine.

The value we propose is not automatic agreement. It is a more durable foundation for reaching informed conclusions.

Whether that foundation reduces investigation effort, improves reconstruction, or lowers the cost of trust remains an empirical question.

It is one we intend to test.

6. The bigger picture

The need for operational evidence is not new.

What is changing is the nature of the systems producing it.

Autonomous machines increasingly combine software, models, sensing, and physical action. Their behavior depends on interactions extending beyond individual components, and their operating conditions evolve after deployment.

In March 2026, NIST published *Challenges to the Monitoring of Deployed AI Systems*, identifying obstacles to understanding AI behavior after deployment, including fragmented logging and limitations in existing assessment methods.[9]

That research does not establish that independent evidence infrastructure is the necessary solution.

Our thesis is more specific.

Before an autonomous operation can be examined, relevant evidence must be available. And for that evidence to remain useful across systems, vendors, and changes, its custody, provenance, integrity, and limitations must be preserved.

We believe this requirement will become more important as physical AI scales.

The practical case must be tested beyond synthetic examples. Implementation tests should examine collection, attribution, gap reporting, and integrity verification against documented requirements. Field studies should compare investigation effort, supported conclusions, and recognition of evidence limits against existing workflows, alongside the cost of maintaining the record.

Unacceptable operational burdens or no meaningful investigative improvement would weaken the practical case. Existing platforms meeting the requirements would support the principle while challenging the need for a dedicated recorder.

Until these evaluations are completed, reductions in investigation time, downtime, or dispute costs remain hypotheses.

What we are building

Roboblackbox was founded around the belief that the people responsible for robotic operations should have an independent record of what those systems publish.

We are building a passive, read-only recorder under operator control.

Our evidence model is designed to bring together separately attributed robot publications, facility states, references to camera video retained on site, video-derived observations with their provenance and uncertainty, and operator declarations.

The recorder preserves evidence without directing machines or participating in protective functions.

Its purpose is to make operational history available for examination across vendors, configurations, and changes.

The bet behind Roboblackbox is not that there will be more robots, although there will be. It is that physical AI changes what a machine is. Machines are becoming continuously changing software systems that perceive the world, decide and act on it, and their makers' own records, however good, will not add up to the history of an operation that contains several of them.

As physical AI scales, the ability to independently examine how autonomous machines operate will become essential. We already value independent evaluation before deployment. We believe physical AI will need independent evidence after deployment too, and that the party holding it should be the one that owns the consequences.

Roboblackbox is building that record.

References

[1] Stouffer, K., Pease, M., Tang, C. Y., Zimmerman, T., Pillitteri, V., Lightman, S., Hahn, A., Saravia, S., Sherule, A., and Thompson, M. (2023). *Guide to Operational Technology (OT) Security.* NIST SP 800-82 Rev. 3, Section 3.2.1. National Institute of Standards and Technology.
https://doi.org/10.6028/NIST.SP.800-82r3

[2] Adams, C., Cain, P., Pinkas, D., and Zuccherato, R. (2001). *Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP).* RFC 3161, Sections 1, 2.1, 2.2, and 2.4.2. Internet Engineering Task Force.
https://www.rfc-editor.org/rfc/rfc3161.html

[3] Laurie, B., Messeri, E., and Stradling, R. (2021). *Certificate Transparency Version 2.0.* RFC 9162 (Experimental), Sections 2.1.3, 2.1.4, 8.3, and 11.3. Internet Engineering Task Force.
https://www.rfc-editor.org/rfc/rfc9162.html

[4] International Maritime Organization. *Voyage Data Recorders.*
https://www.imo.org/en/ourwork/safety/pages/vdr.aspx

[5] National Transportation Safety Board. (February 13, 2024). *NTSB: Retrofit all Cockpit Voice Recorder-Equipped Airplanes With 25-Hour CVRs.*
https://www.ntsb.gov/news/press-releases/Pages/NR20240213.aspx

[6] National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* NIST AI 100-1, Section 5.3, including MEASURE 1.3.
https://doi.org/10.6028/NIST.AI.100-1

[7] Public Company Accounting Oversight Board. *AS 1000: General Responsibilities of the Auditor in Conducting an Audit,* paragraphs .04-.05.
https://pcaobus.org/oversight/standards/auditing-standards/details/as-1000--general-responsibilities-of-the-auditor-in-conducting-an-audit

[8] Public Company Accounting Oversight Board. *AS 1105: Audit Evidence,* paragraphs .06-.10.
https://pcaobus.org/oversight/standards/auditing-standards/details/AS1105

[9] Rao, A. K., Keller, A. J., Kalra, N., Steed, R., Kwegyir-Aggrey, K., Klyman, K., Staheli, D., and Bergman, A. S. (2026). *Challenges to the Monitoring of Deployed AI Systems.* NIST AI 800-4, Sections 3.1.1 and 3.2.2. National Institute of Standards and Technology.
https://doi.org/10.6028/NIST.AI.800-4

ROBOBLACKBOX
The independent record for robotic operations.

October 2026 · Foundational Thesis