Two problems with a shared evidence lifecycle

Research question

How can physical observations and digital authenticity signals support better decisions while preserving validity, privacy and accountable authority? The project joins these topics at the evidence-lifecycle level while keeping their objects of inference separate. A calibrated measurement and a transaction credential answer different questions. [H1]

Evidence	What it can support	What needs separate evidence
Sensor provenance	Which instrument produced a record	Accuracy of the physical measurement
Possession of a cryptographic key	Control of that credential in context	Named identity, intent or freedom from coercion
Liveness signal	A bounded tested interaction property	Universal personhood or absence of fraud
Model prediction	Forecast under stated inputs	Causal explanation or permission to act

Work produced

The package contains a research monograph, custom research standards, an operationalization paper and review records. It proposes two separately scored pilots, rather than treating physical measurement quality and proof-of-human signals as one universal trust score.

Completed contribution

The substantive output is a more testable research program: specified comparisons, evaluation units, evidence requirements, falsifying outcomes and unresolved implementation decisions. It is not an empirical demonstration that additional sensing improves decisions or that an authenticity method defeats attacks.

The sensing comparison before and after review

The defect

A dense-data pilot can change several things at once: observation policy, sensor hardware, model training, decision threshold and operator workflow. If outcomes improve, that design cannot identify which change caused the improvement. The companion BPL record classified this as a high-severity issue and repaired the proposed comparison. [H2]

Design choice	Revised primary comparison
Comparators	Current practice; optimized conventional layout; candidate dense/autonomous policy
Downstream model	Hold model family and training data fixed
Decision policy	Hold thresholds and operator workflow fixed
Broader adaptation	Evaluate retraining and other system changes separately
Reference	Use fit-for-purpose measurements independent of candidate sensors
Evaluation	Protect confirmation data and define precision before opening it

Why the conventional baseline matters

Comparing a candidate only with a weak current layout can overstate its value. An optimized conventional sensing policy asks whether the extra density or autonomy earns its lifecycle cost. The proposed study allows the candidate to fail this comparison.

Unit and outcome

A site-by-nonoverlapping decision window is a candidate unit, subject to domain justification. The primary outcome is a prespecified difference in decision-weighted loss. Measurement error, calibration, missingness, false alarms, energy, maintenance and lifecycle cost remain visible secondary outcomes. No sample size or expected effect is invented.

A bounded authenticity experiment

Decision and threat model

Select one transaction or communication channel and specify the attacker capabilities before testing. Candidate signals could include transaction-bound confirmation or credential-based methods. Each combination needs a statement of exactly what it can establish and the use for which that evidence is sufficient. [H1, section 7.1]

Protocol element	Concrete requirement
Independent unit	Transaction or communication attempt in a declared scenario
Source clustering	Variants from one source remain one source cluster
Comparator	Existing controls on the same eligible attempts
Primary outcome	Decision-weighted loss with declared error and recovery costs
Secondary outcomes	Attack completion, calibration, abstention, latency and friction
Human outcomes	Accessibility, false rejection, appeal, recovery and privacy/linkability

Protected evaluation

Separate development, calibration and confirmation partitions. Cluster by person, device, source or attack family where relevant to prevent leakage. Use consented and lawfully sourced material, define invalid runs and preserve no-decision outcomes. A downstream service must actually honor abstention for it to serve as a useful safeguard.

What would defeat the hypothesis

A candidate that fails to improve decision quality at acceptable cost, or shifts errors onto excluded users, may be unsuitable. If uncertainty includes both meaningful benefit and harm, classify the result as inconclusive. More persuasive media or a stronger-looking badge cannot substitute for measured transaction outcomes.

Standards applied to a research-to-operation path

Scope of HSDLG methods

TUP structured the intended work, sources and completion checks. Document-guided Backpass examined the argument and claim boundaries. BPL linked eligible findings to repair and whole-target rereview. The producing assistant performed these roles with exposure to its own work; this is not independent assurance. [H2]

Gate	Evidence that must be resolved
Domain and owner	One actual decision and accountable domain authority
Rights and reference	Permitted access, reference method and protected data handling
Comparison	Frozen policies, controls, outcomes and analysis plan
Empirical qualification	Field or transaction evidence with uncertainty and invalid runs
Operational admission	Action limits, escalation, rollback and responsible acceptance
Monitoring	Changes, new attacks, drift and revisited conclusions

Roles of the broader research components

TAIE/EMSM inform separation of physical realization, observation and inference. Their geological origin does not validate transfer to every physical domain. AAAM is discussed as a candidate way to relate claims and evidence; association alone is not proof. ValidityBench is proposed to evaluate the evaluator as well as outputs, including polished errors and valid abstentions. No execution of these components is claimed for this paper. [H1]

Why standards are useful here

The standards convert general aspirations into questions a reviewer can answer before allowing consequential use. They leave critical missing inputs visible: domain, site, data rights, reference, adjudicator, sample/precision plan and independent evaluation.

The review record and next decisive evidence

Recorded finding	Repair
Gate language was procedural	Added decision-specific planning questions and review holds
Sensing comparison was confounded	Fixed the primary downstream model/workflow and separated adaptation
Predictability and control sounded established	Added conditional claims and falsifying outcomes
Candidate architecture sounded deployed	Separated framing, component maturity and operational qualification

Review disposition

The companion record reports two document-guided cycles and acceptance as an analytical, prospective paper. Native execution, independent domain/legal review and empirical field study remained NOT_RUN. The title Proving Human therefore describes an explored problem, not an implemented identity product. [H2]

A realistic next project

Choose either a narrow physical measurement decision or one authenticity transaction. Secure an owner, rights and reference judgments, then freeze the protocol. Test whether the revised design produces an interpretable result at acceptable lifecycle cost. Evaluate both correct decisions and burdens imposed on users and reviewers.

Sources

[H1] Data for the Real World and Proving You Are Human: An Evidence-Governed Research Monograph, revised supplied version, especially sections 4 and 7.1.
[H2] BPL Review Record - Making Data for the Real World Real, findings F01-F04 and disposition.
[H3] Making Data for the Real World Real - HSDLG_v2 Research and Operationalization Paper, deployment gates and proposed comparison.
These documents are in the supplied Proving Human archive. This expansion examines their reasoning and review changes; it introduces no empirical result.

Diligence

Additional standards, protocol detail and supporting review records are available under an appropriate NDA or contract. The public case exposes the research design and its meaningful repairs while protecting implementation details.

Additional technical documentation, standards, test records and implementation details are available for scoped diligence under an appropriate NDA or contract.