AI Decision Auditability Benchmark

PoC Assessment Guide

Aligned with VCP (VeritasChain Protocol)

Version 1.0 — How to Measure (The Minimum Viable Assessment)

Document ID:	VSO-SCORE-002
Related:	VSO-SCORE-001 (Scorecard)
Total Time:	~3 hours (all criteria)
Publisher:	VeritasChain Standards Organization

veritaschain.org

Overview

This guide provides the **minimum viable test procedure** for each criterion. The goal: give you a concrete way to measure each criterion in the shortest time possible.

Assessment Procedures

1. Third-Party Verifiability

第三者検証可能性

Time: 30 min | **Check:** Can someone outside your organization independently verify your audit trail?

Steps:
1. Export a sample audit log (10-100 records)
2. Give it to someone unfamiliar with your system
3. Ask them to verify 3 random records using only the exported data
4. No phone calls, no vendor support, no internal tools allowed

0	Verifier cannot proceed without proprietary tools or insider help
1	Verifier can partially verify but needs assistance for some steps
2	Verifier completes full verification using only exported data + public docs

Evidence: Verification attempt log, verifier feedback

2. Tamper Evidence

改ざん検知

Time: 20 min | **Check:** If someone modifies a historical record, will you know?

Steps:
1. Take a backup of your audit log
2. Modify one field in one historical record
3. Run your integrity check process
4. Observe: Does the system detect and alert?

0	Modification goes undetected
1	Detected only if you know where to look
2	Automatic detection with alert; modification location identified

Evidence: Screenshot of tamper test, alert log

3. Sequence Fixation

順序の固定

Time: 15 min | **Check:** Can the order of events be disputed or rewritten?

Steps:
1. Pick any trade: find the Decision → Order → Execution chain
2. Check: Are timestamps monotonically increasing?
3. Check: Is there cryptographic binding between events?
4. Try to insert a backdated event

0	Events can be reordered or backdated without detection
1	Timestamps exist but no cryptographic proof of sequence
2	Hash chain or similar; backdating is cryptographically impossible

Evidence: Sample event chain with timestamps and hashes

4. Decision Provenance

判断由来

Time: 20 min | **Check:** Can you trace WHY a decision was made?

Steps:
1. Pick a random algorithmic decision from last week
2. Can you find: (a) Input market data, (b) Model parameters, (c) Decision logic?
3. Time yourself: How long to reconstruct the full context?

0	Only the outcome is logged; inputs/logic unavailable
1	Partial reconstruction possible (>30 min, incomplete)
2	Full context retrievable in <10 min with complete audit trail

Evidence: Reconstructed decision context for one sample trade

5. Responsibility Boundaries

責任境界

Time: 15 min | **Check:** Can you identify WHO approved or modified each action?

Steps:
1. Find a manual override or parameter change from the past month
2. Can you identify: (a) Who made it, (b) When, (c) What was changed?
3. Is this attribution cryptographically signed or just a username field?

0	No attribution; "admin" or generic accounts used
1	Username logged but no signature; could be spoofed
2	Digital signature on all overrides; non-repudiable

Evidence: Sample override log with attribution details

6. Audit Submission Readiness

監査提出性

Time: 30 min | **Check:** Can you produce a complete evidence package on demand?

Steps:
1. Simulate: "Regulator requests all trading activity for Account X, Date Y"
2. Time yourself: How long to produce a complete, structured export?
3. Is the format documented? Can a third party parse it without help?

0	Manual gathering required; takes hours/days
1	Export exists but incomplete or requires manual assembly
2	One-click export; complete package in <5 minutes

Evidence: Sample export package, time-to-produce measurement

7. Retention & Durability

保持期間・耐久運用

Time: 15 min | **Check:** Will your records survive for required retention periods?

Steps:
1. Document: What is your stated retention policy?
2. Check: Can you retrieve records from 1 year ago? 3 years ago?
3. Verify: Is there automated backup? Redundancy? Integrity monitoring?

0	No policy; records may be deleted or lost
1	Policy exists but retrieval is manual or untested
2	Automated retention with verified retrieval and integrity checks

Evidence: Retention policy document, sample retrieval

8. Timestamp Reliability

時刻の信頼性

Time: 15 min | **Check:** Are your timestamps trustworthy?

Steps:
1. Check: What time source do your systems use?
2. Verify: Is there drift monitoring? What is typical drift?
3. Document: Can you prove timestamps to an external party?

0	Local system clocks only; no sync verification
1	NTP sync but no monitoring or documented accuracy
2	PTP or RFC 3161 TSA; documented accuracy; monitoring in place

Evidence: Time sync configuration, drift monitoring logs

9. Cryptographic Strength

暗号強度

Time: 10 min | **Check:** Are your cryptographic choices defensible?

Steps:
1. List: What hash algorithm(s) do you use?
2. List: What signature algorithm(s)?
3. Check: Is there documented key management (rotation, storage)?

0	Deprecated algorithms (MD5, SHA-1, RSA <2048)
1	Adequate algorithms but no key management procedures
2	Strong algorithms (SHA-256+, Ed25519) with documented key lifecycle

Evidence: Algorithm inventory, key management policy

10. Cryptographic Agility

暗号移行性

Time: 10 min | **Check:** Can you upgrade algorithms without breaking verification?

Steps:
1. Check: Are algorithm identifiers stored with signatures?
2. Ask: If you switched to a new algorithm tomorrow, could old records still verify?
3. Document: Is there a migration plan for post-quantum (PQC)?

0	Hard-coded algorithms; migration would break old signatures
1	Algorithm identifiers exist but migration untested
2	Documented migration path; backward compatibility verified

Evidence: Sample record with algorithm identifier, migration plan