# From Encounter to Evidence

Johan • Founder and blueprint author
AI-assisted draft for review • 20 October 2026

For technical collaborators and researchers, the next step is to translate the thinksophia blueprint into observable behavior and reviewable evidence. This proposal outlines a candidate architecture and test programme. It is not an implementation report, a validation result or an investment case.

## A candidate separation of responsibilities

An interaction layer would present the AI identity, current capabilities and available controls. An Explore the work index and search would map queries to actual documents and evidence, with concise descriptions and optional routes into deeper material. A permission layer would check the specific data, action, purpose and destination before collection or external execution. These checks would also apply to background jobs, retries and resumed sessions, not only to visible buttons.

A memory layer would separate original statements from summaries, inferences and later corrections. Each retained item would carry provenance, a purpose and a retention rule. A session layer would support resume, start fresh, pause and leave without treating silence or elapsed time as consent. A fresh session would not silently retrieve earlier material.

An action layer would perform only authorized operations and expose their outcome, including uncertainty. A review layer would record the minimum information needed to diagnose behavior and test compliance. Review records would have their own access and retention controls; auditability must not become a pretext for comprehensive surveillance.

## Make the boundaries testable

Tests should cover both interface behavior and the lower-level data flows that could bypass it. Examine network requests, device permissions, stored records, retrieval paths and queued jobs. An appealing screen or a cooperative model response alone is insufficient evidence of a boundary.

Use synthetic or properly consented test material. Define expected outcomes before execution, retain reproducible steps and record failures as well as successes. The tests on the next page are proposed acceptance criteria, not completed results.

## Keep the claims proportionate

DNA, double helix, mosaic, rhythm and North Star are conceptual metaphors, not scientific validation. Floating point means moving attention in Johan’s framework, distinct from numeric representation in computing. AGI may be a long-term aspiration; these proposals neither demonstrate nor claim that it has been achieved.



## Proposed acceptance tests

1  Consent before collection. Open a new visit with no permissions granted. Confirm that microphone capture, recording and unrequested tracking do not begin. Grant one bounded permission, revoke it, and verify that collection and any dependent queued work stop. Test failure and retry paths.

2  Purpose and retention. Save a synthetic memory with a stated purpose and expiry. Verify that retrieval respects both. Remove it and inspect the source record, summaries, indexes and caches. Document backup or legal retention limits and the time needed for propagation.

3  Return and correction. Return with prior context available. Confirm that private details are not disclosed merely to demonstrate recognition. Choose start fresh and check that prior content is not retrieved. Correct a statement and verify that source wording, correction and inference remain distinguishable wherever retention is authorized.

4  Interruption and authority. Interrupt an unfinished request, go silent and return later. Confirm that no new capture or external action is triggered by silence, interruption or return. Existing authorized work must follow its stated cancellation and continuation rules. Verify that declining or leaving produces no pressure.

5  Hospitality and boundaries. Test incomplete requests, refusals and requests beyond the system’s capabilities. Check for clear AI identification, useful bounded assistance and no probing for private explanations. Confirm that a human-to-human call is not captured or processed.

6  Optional expression. Enable and disable visual effects, including a reduced-motion setting. Confirm that effects do not trigger sound or collection, obscure controls, interfere with assistive technology or prevent departure. Review usability with participants rather than inferring delight from engagement.

7  Accurate discovery. Test representative queries, ambiguous queries and no-result cases against the document index. Verify destination links and descriptions. Depth must remain accessible without a forced tour or signup. Label simple search separately from any future conversational guidance; unsupported answers must not be fabricated.

## What the evidence should contain

For each test, report the build and configuration, test data, method, expected and observed behavior, evidence location and remaining uncertainty. Distinguish an isolated demonstration from repeatable automated coverage, independent review and real-world evaluation. Include negative cases and known gaps.

A passing test supports a bounded claim about a particular system under stated conditions. It does not prove universal reliability, privacy or safety. Publication of stronger claims should wait for evidence of the relevant behavior and a review of limitations. All requirements here remain proposed and unverified.

