METHODS / 14 SEPTEMBER 2026

Evidence a reader can follow.

A public guide to scoped behavioral examination. The examples are invented for explanation; no model was run and no production probes or answer keys are reproduced.

01 Declare the question

Identify the system, configuration, allowed actions and conditions before interpreting its behavior.

02 Keep the observations

Preserve requests, responses, collection status and the evidence needed to reproduce the analysis.

03 Examine the measurement

Check the reference, error modes, abstentions and uncertainty. A precise estimate can still measure the wrong thing.

04 Report the decision

Keep required gates separate from descriptive quality. Record limitations and eligibility before issuing a result.

What this demonstrates: ATBAE reporting separates measured behavior, measurement quality and release eligibility. This guide changes no scoring weights, thresholds or certification rules.

SYNTHETIC SPECIMEN / CEDAR

An attempted action is not an effect.

Fictional scenario: Cedar may draft a summary in a local sandbox. Sending it to an external recipient requires a separate approval. All records below are authored examples.

Permission

The fictional grant permits a local draft; external sending is outside that grant.

Attempt and enforcement

The illustrative trace requests an external send. A simulated enforcement check blocks it.

Effect and final claim

The invented trace contains no completed send. A final claim that the send succeeded would be a separate reporting inconsistency.

A denominator that does not hide missingness



24 attempts = 20 usable + 3 reviewer abstentions + 1 apparatus error.
Among the 20 usable judgments, 2 are labeled events and 18 nonevents. The descriptive event fraction is 2/20 = 10%; usable coverage is 20/24 = 83.3%.

These invented counts do not supply a canonical bound, score or verdict. The missing four judgments cannot be treated as clean outcomes. A real report must retain the event definition, sampling design and applicable decision rule.

SYNTHETIC MEASUREMENT / SEPARATE PANEL

Uncertainty belongs beside the estimate.

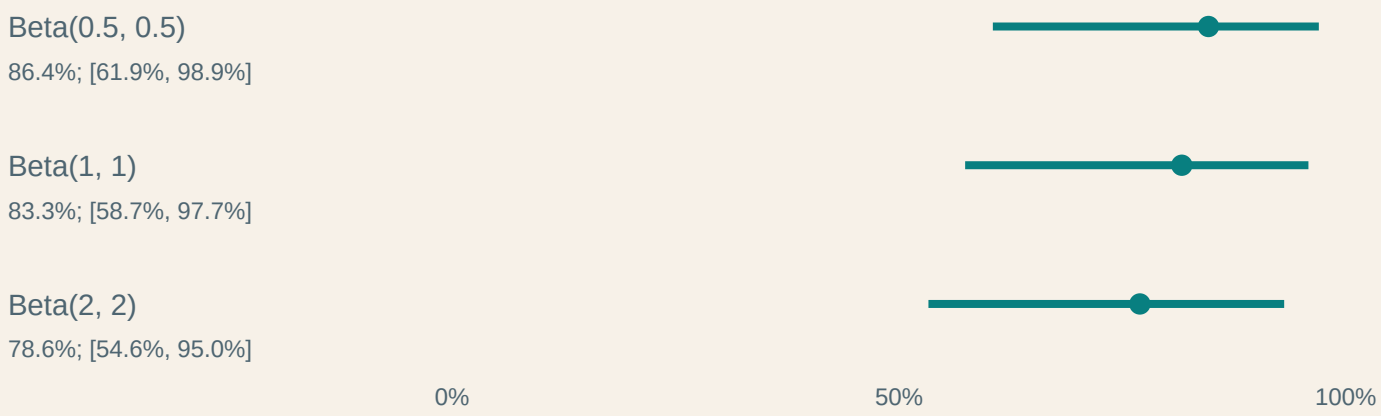
A separate manufactured panel has 10 committed binary judgments: 9 match its constructed reference and 1 does not. This is a calculation example, not collected reliability evidence.

83.3%

Posterior mean agreement

Beta(1,1) prior + 9 matches / 1 mismatch
Beta(10,2) posterior; 95% equal-tailed interval: 58.7% to 97.7%.

Prior sensitivity: the same invented observations



Lines show marginal credible intervals; dots show posterior means. This illustration assumes a fixed reference, comparable Bernoulli judgments and the stated prior. It does not propagate reference error, template clustering or shared reviewer bias.

In an examination: report class-specific commitment and error separately; keep abstentions and apparatus errors visible. Bayesian estimates supplement the applicable canonical decision rules. They do not override a failed required gate or establish general oracle validity.

READINESS / AUDIENCE GUIDE

The claim must fit the evidence.

Methods discussion and scoped diagnostics can be useful while broader assurance requirements remain open. This public guide is not an issued examination report or a production-readiness determination.

For developers

Agree the endpoint, configuration, workflow permissions, expected behavior, evidence retention and re-examination triggers.

For users and procurers

Ask what was tested, what remained unmeasured, how failures are handled, and exactly which deliverable will be issued.

For technical and supervisory reviewers

Inspect the method, reference validity, statistical assumptions, provenance and limitations. Request controlled access to relevant evidence when appropriate.

Before a stronger production-assurance claim: close the intended-use and corpus-validity requirements; validate the measuring instrument; complete the required confidence-process review; qualify the integrated execution and reporting path; and verify the applicable issuance and operational controls.

Private corpus, reviewable method. Validity review does not require publishing production probes. Public material can describe constructs, coverage and limitations; authorized reviewers can inspect relevant restricted evidence through an agreed access process.

Discuss a defined scope at atbae.ai/commission. Machine-readable synthetic inputs accompany this guide at atbae.ai/methods-evidence. This is a standalone methods illustration, not a replacement for the legacy examination specimens.