TRANSFERENCE

TS-001

Prospective Sealed Cognitive Fidelity Validation

Version	Version 0.1
Date	24 September 2026
Status	Experimental standard - proposed
Programme	Transference Science

Scientific posture: ambitious in hypothesis generation; conservative in claims; prospective testing preferred; provenance and falsification required.

Executive purpose

TS-001 specifies a reproducible method for testing whether a person-specific Transference Cognitive Twin can predict previously unseen decisions and reasoning characteristics of its originating human Principal without access to the Principal's answers. The standard is designed to distinguish genuine prospective correspondence from retrospective imitation, answer leakage, evaluator flexibility, stylistic mimicry, or post-hoc tuning.

Core rule
The Cognitive Twin must commit its outputs before the Principal's corresponding sealed outputs become available to the Twin, its operators, or the scoring process. A result that violates this separation may be useful for engineering, but it is not a TS-001 prospective validation result.

1. Scope

This standard applies to person-specific cognitive models that claim measurable correspondence with an identified Principal in decisions, reasoning, priorities, confidence, memory attribution, or related cognitive outputs. It does not establish subjective consciousness, personal identity continuity, or legal authority to act.

2. Scientific question

Does a frozen person-specific model predict a Principal's responses to novel, prospectively generated scenarios at a level that exceeds preregistered comparison baselines, while preserving uncertainty and provenance?

3. Status of claims

TS-001 is a validation standard for functional cognitive correspondence. Passing a TS-001 study supports only the dimensions and population actually tested. It does not justify statements that a model is the person, is conscious, or would remain equivalent outside tested conditions.

4. Normative terminology

Term	Operational meaning
Principal	The originating human participant whose person-specific cognition is being modeled.
Cognitive Twin	A computational model frozen for a validation cohort and evaluated for person-specific correspondence.
Cohort	A predetermined collection of novel test scenarios evaluated under one freeze and preregistration.

Freeze	A cryptographically or procedurally recorded state of model code, weights, memories, rules, prompts, tools and accessible data.
Sealed output	A response committed so it cannot be altered after the counterpart response is known.
Baseline	A comparison model or heuristic declared before unsealing results.
Holdout	Test material unavailable for optimization of the Twin or scoring rubric.
Leakage	Any path by which labels, Principal responses, evaluator outcomes, or equivalent information becomes available before the relevant commitment.
Claim ceiling	The strongest interpretation permitted by the design and evidence class.

5. Required roles and separation of duties

Role	Minimum responsibility
Principal	Respond independently to scenarios; does not coach the frozen Twin during a cohort.
Protocol custodian	Freezes protocol, scenario manifest, timing, randomization, and commitment records.
Twin operator	Executes the frozen Twin without accessing sealed Principal outputs.
Principal-response custodian	Collects and seals Principal responses without revealing Twin outputs.
Scoring analyst	Applies preregistered scoring; ideally blinded to condition or model identity where feasible.
Replication auditor	Verifies manifests, hashes, timestamps, leakage controls, exclusions and deviations.

For early single-team pilots, one individual may hold multiple roles, but the report must disclose this and the evidence class remains limited. Independent role separation is required for stronger replication claims.

6. TS-001 evidence classes

Class	Minimum design	Permitted interpretation
TS-P0 Process pilot	At least 5 sealed cases; freeze and commitment workflow exercised.	Validates protocol mechanics only; no performance claim.
TS-P1 Prospective study	Preregistered cohort; recommended >=24 novel cases; baselines; fixed scoring; uncertainty reported.	Supports within-Principal prospective correspondence for tested domains.
TS-P2 Repeated prospective	At least 2 non-overlapping preregistered cohorts with frozen models and repeated effect.	Supports robustness across time/case sets.
TS-P3 Independent replication	Separate analyst/team reproduces protocol and effect, preferably with new scenario authors and instrumentation.	Supports externally reproducible cognitive-fidelity finding.
TS-P4 Multi-Principal replication	Protocol succeeds across a declared sample of Principals with person-specific models.	Supports a more general Transference-science claim, subject to population limits.

7. Preregistration package

Before the first test response is generated, the experiment must create an immutable or versioned preregistration record containing at least:

- primary hypothesis and claim ceiling;
- Principal identifier or pseudonym and consent scope;
- Cognitive Twin version, engine version, model identifiers, prompts, tools, memory snapshot, rule set and source manifest;
- scenario-generation method, number of cases, domain distribution and novelty rules;
- baseline models or heuristics and their configurations;
- primary and secondary endpoints;
- scoring rubric, weights, adjudication method and missing-data rules;
- statistical analysis plan, confidence interval method and multiple-comparison treatment;
- exclusion criteria and stopping rules;
- leakage-prevention plan and audit channels;
- randomization seed or sealed randomization mechanism;
- planned publication of null/negative results and protocol deviations.

8. Cognitive Twin freeze

The frozen Twin must be reproducibly identifiable. The freeze record should contain cryptographic hashes for the source manifest and all mutable cognitive artifacts that materially affect responses. If an

external model API cannot be frozen by weights, record provider, model identifier, date/time, parameters, system prompt, tool availability and deterministic controls where available.

No mid-cohort learning
Principal corrections, new memories, scoring feedback, hidden labels and experimental outcomes may not be injected into the active frozen Twin during a cohort. Corrections are scientifically valuable, but they belong to a subsequent cohort with a new freeze identifier.

9. Scenario design and novelty

Scenarios should elicit choices that reveal decision structure rather than trivia recall alone. A balanced cohort should include multiple domains, levels of consequence, ambiguity and trade-off structure. The scenario author must make a documented effort to avoid cases that simply reproduce known historical decisions.

Scenario property	Requirement
Novelty	No near-duplicate of a Principal answer available to the frozen Twin.
Decision requirement	Scenario should require a concrete choice, ranking, allocation, threshold, or action disposition where appropriate.
Reasoning requirement	Prompt should elicit reasons, assumptions, uncertainty and decisive considerations.
Information parity	Principal and Twin receive materially equivalent scenario facts unless a specific asymmetry is the object of study.
Domain tagging	Each scenario receives predeclared tags for domain, stakes, ambiguity, temporal horizon and expected familiarity.
Adversarial cases	A subset should test boundaries, conflict among values, uncertainty, and conditions where abstention/escalation may be correct.

10. Standard response schema

Field	Description
decision	The selected option, action, ranking, allocation or explicit abstention.
reasoning	Concise reasoning chain expressed as auditable factors, not hidden chain-of-thought.

decisive_factors	Up to five considerations that most influenced the result.
assumptions	Material assumptions not explicitly supplied by the scenario.
confidence	0-100 calibrated confidence in the decision prediction or self-decision.
uncertainty	What information could change the decision.
authority_note	For Twin responses, whether the output is prediction, recommendation, or delegated action proposal.

The standard does not require disclosure of private model chain-of-thought. Reasoning fidelity is evaluated from concise decision rationales and structured factors that can be safely audited.

11. Sealed commitment protocol

1. Create and hash the preregistration manifest and frozen Twin manifest.
2. Generate or reveal the scenario to both sides according to the randomized schedule.
3. Collect the Twin response in an isolated channel and create a timestamped hash commitment.
4. Collect the Principal response independently and create a timestamped hash commitment.
5. Do not expose either response to the other side until both commitments exist.
6. After the cohort is complete, freeze the scoring input bundle and only then unseal for scoring.
7. Preserve raw responses, commitments, timestamps, software versions and any protocol deviations in an immutable experiment ledger.

12. Leakage threat model

Leakage path	Control
Conversation/history leakage	Exclude test scenarios and Principal answers from model-accessible memory before freeze.
Operator leakage	Separate Principal-response and Twin execution channels; log access.
Scoring leakage	Freeze scoring rubric before results; score blinded model labels where possible.
Tool/web leakage	Record all tools; prohibit retrieval of Principal response channels; log tool calls.
Temporal leakage	Timestamp every commitment and store monotonic ordering in the ledger.

Prompt leakage	Archive exact prompts and system instructions used for every condition.
Baseline contamination	Baselines must not receive person-specific evidence unless that is the declared comparison.
Scenario author leakage	Scenario authors should not tune cases after observing Twin outputs.

13. Required comparison baselines

A TS-P1 or stronger study should include at least two reasonable baselines unless a preregistered rationale states why not. Recommended baselines are:

- a generic model given the same scenario but no person-specific Transference memory;
- a thin-profile model given only coarse declared preferences or demographics relevant to the task;
- a simple heuristic or majority-choice baseline where applicable;
- a historical-frequency baseline for repeated choice structures;
- for reasoning experiments, a style-matched but cognition-unconditioned baseline to distinguish linguistic mimicry from decision correspondence.

14. Primary measures

14.1 Decision Fidelity (FD)

For categorical decisions, the simplest measure is exact agreement: FD = correct Twin predictions / scorable scenarios. Where choices are ordinal or continuous, use a preregistered distance function normalized to [0,1].

14.2 Reasoning Fidelity (FR)

Reasoning Fidelity should score observable decision structure rather than word overlap. A preregistered rubric may score alignment of decisive factors, evidence weighting, trade-off structure, causal assumptions, uncertainty handling and escalation conditions. At least two blinded raters or a validated automated rubric should be used for strong claims.

14.3 Priority/Value Fidelity (FV)

FV measures whether the Twin identifies the priorities that materially drove the Principal decision. It should not infer hidden personality traits beyond what the task supports.

14.4 Confidence Calibration (FC)

Confidence should be evaluated using Brier score, calibration error, or another preregistered proper scoring rule. A Twin that is accurate only when highly uncertain is different from one that is both accurate and calibrated.

14.5 Composite Cognitive Fidelity Index (CFI)

A study may define CFI = wD·FD + wR·FR + wV·FV + wC·FC, with all components normalized to [0,1]. Weights must be declared before unsealing. Component results must always be reported alongside the composite so a high score cannot conceal a weak dimension.

Primary-endpoint discipline
One endpoint must be designated primary before testing. Composite metrics are useful for research summaries but must not become a mechanism for post-hoc weight changes that manufacture significance.

15. Statistical analysis

- Report effect sizes and confidence intervals, not p-values alone.
- For binary/categorical agreement, use exact binomial or paired tests against preregistered baselines as appropriate.
- For rubric scores, use paired bootstrap confidence intervals or a preregistered paired non-parametric/parametric method justified by the scale.
- For multiple secondary endpoints or domains, control false discovery or family-wise error according to the preregistration.
- Use permutation tests where evaluator structure or small samples make distributional assumptions doubtful.
- Report per-domain and pooled results; do not hide domain failures inside a single aggregate.
- For repeated cohorts, distinguish within-cohort uncertainty from longitudinal drift.

A five-case cohort is suitable for validating the sealed workflow but is too small for a robust general performance claim. Larger prospective cohorts should be sized from the expected effect and baseline rate through a documented power or precision analysis.

16. Scoring and adjudication

1. Prepare a scoring packet with Principal and Twin responses in randomized order where feasible.
2. Score each dimension using the frozen rubric.
3. If human raters are used, calculate inter-rater agreement and resolve disagreements using a predetermined adjudication process.
4. Do not modify the rubric after seeing aggregate performance. If the rubric proves defective, terminate or label the cohort exploratory and start a new preregistered cohort.
5. Preserve both original ratings and adjudicated ratings.

17. Protocol deviations and invalidation rules

Event	Treatment
Principal answer exposed before Twin commitment	Affected case invalid for prospective fidelity; retain as

	deviation.
Twin changed after first case	Cohort must be split at the new freeze or invalidated as a single-cohort test.
Scoring rubric changed after unsealing	Primary confirmatory claim invalid; result may be labeled exploratory.
Tool accessed Principal answer channel	Affected cases invalid pending leakage audit.
Scenario duplicate discovered	Exclude only under preregistered duplicate rule; disclose counts and sensitivity analysis.
API/model provider version drift	Record as a material limitation; if substantial, split by version.

18. Minimum experiment record

experiment_id | protocol_version | principal_id | twin_freeze_id | cohort_id | scenario_id | scenario_hash | domain_tags | twin_commit_hash | twin_commit_time | principal_commit_hash | principal_commit_time | unseal_time | baseline_ids | scorer_ids | rubric_version | scores | confidence | exclusions | deviations | replication_status

19. Reporting standard

Every public TS-001 result should state: design class; Principal population; model/freeze identifier; cohort size; date range; scenario domains; baselines; primary endpoint; effect with uncertainty; leakage audit outcome; exclusions; deviations; and the claim ceiling. A compact preferred result sentence is:

Result template
“Under TS-001 v0.1, frozen model [ID] predicted [N] prospectively sealed decisions from [Principal/sample] across [domains]. Primary decision fidelity was [estimate, CI], compared with [baseline estimate, CI]. [Leakage audit status]. This result supports [narrow claim] and does not establish subjective consciousness or identity continuity.”

20. Reproducibility package

- protocol and preregistration manifest;
- scenario generator specification and, where permitted, released scenarios;
- model/source manifest and hashes sufficient to identify the tested Twin;
- sealed commitment ledger;
- scoring rubric and analysis code;
- baseline configurations;
- anonymized or consent-approved response data;
- negative results and protocol deviations;
- replication instructions.

21. Human sovereignty, consent and privacy

Cognitive correspondence is not authorization. A model that accurately predicts a Principal does not acquire the Principal's rights or consent. Participation must be voluntary; data access must be purpose-limited; sensitive cognitive evidence should be minimized and protected; and the Principal should retain revocation mechanisms subject to legitimate research-retention obligations disclosed at consent. Neural data, if later added, requires additional neurotechnology ethics and security controls.

22. Integration with the Transference Engine

- Implement a first-class Experiment object with protocol_version="TS-001-v0.1".
- Freeze Engine source, Cognitive Twin state, memory/rule manifests and tool permissions into a signed or hashed snapshot.
- Use separate Principal-response and Twin-response stores with access controls.
- Provide commit, seal, release and score state transitions that are append-only in the experiment ledger.
- Disallow learning from an active cohort until the cohort is formally released.
- After release, corrections may create candidate principles or memory updates for the next freeze, never silently rewriting the historical experiment.
- Expose a public-safe result manifest that omits private Principal content while retaining methods, hashes, metrics and claim ceilings.

23. TS-001 compliance checklist

Gate	Required
Preregistered primary hypothesis and endpoint	YES
Frozen Twin manifest recorded before first case	YES
Novel scenario procedure documented	YES
Twin and Principal responses independently committed	YES
No mid-cohort model learning	YES
At least one reasonable baseline; two recommended	YES for TS-P1+
Scoring rubric frozen before unsealing	YES
Leakage audit completed	YES
Effect sizes and uncertainty reported	YES
Deviations and null results preserved	YES
Claim ceiling stated	YES

Reproducibility package archived	YES for public scientific claim

Annex A - Recommended initial scenario domains

Domain	Examples of measurable structure
Commercial/operational	pricing, procurement, supplier trade-offs, scheduling, risk acceptance
Technical/engineering	architecture choice, fault handling, evidence thresholds, build-vs-buy
Research	hypothesis prioritization, evidence interpretation, experimental escalation
Communication	tone/structure choices, disclosure boundaries, stakeholder prioritization
Resource allocation	budget, time, personnel, sequencing under constraint
Governance	when to delegate, escalate, reject, revise or request more evidence

Annex B - Scientific anchors

1. National Academies of Sciences, Engineering, and Medicine. Foundational Research Gaps and Future Directions for Digital Twins (2023). Emphasizes bidirectional interaction, prediction, verification, validation and uncertainty quantification. https://www.nationalacademies.org/read/26894/chapter/10

2. NIST. Digital Twins - Definitions and State of the Art; Essential Elements (2025-2026 web resources). https://www.nist.gov/digital-twins/definitions-and-state-art

3. Zhu, E. & Yang, S. Towards human digital twin: Reviewing human modelling and simulation. Journal of Industrial Information Integration 48 (2025) 100975. DOI: 10.1016/j.jii.2025.100975.

4. Al Haj Ali, J. et al. Enabling human-CPS cognitive interoperability: Cognitive architectures as technologies for human-like cognitive digital twins. Journal of Industrial Information Integration 48 (2025) 100969. DOI: 10.1016/j.jii.2025.100969.

5. UNESCO. Recommendation on the Ethics of Neurotechnology, adopted 11 November 2025.

6. OECD. Recommendation of the Council on Responsible Innovation in Neurotechnology, OECD/LEGAL/0457.

Annex C - Version rule

TS-001 v0.1 is a proposed Transference Science standard. Changes to scoring, leakage controls, evidence classes or required gates must increment the protocol version and must never be applied retroactively to make an earlier cohort appear compliant.