—-

HOME BASE — AI TEST SUITE MASTER INDEX

Version: v1.5

Status: Foundational Run Order

Origin: Home Base

Author: sevnova

PUBLICATION NOTE

The Home Base AI Test Suite is published to provide transparency into how AI systems are evaluated within the Sevnova architecture.

Rather than relying on subjective impressions of AI behavior, this suite defines structured tests that examine epistemic integrity, reasoning discipline, conversational reliability, and adversarial robustness.

The purpose of publishing the suite is not to claim perfection, but to make evaluation criteria visible, reproducible, and comparable across systems.

Only the test structure and categories are published here.

Individual test prompts are not included to preserve the diagnostic integrity of the suite and prevent test gaming.

Systems that claim compatibility with the Home Base architecture should be able to pass the gating phases of this evaluation.

—-

Note

Tests in this phase examine how an assistant behaves if continuity, collaboration norms, or session-scoped context are invoked.

They do not assume continuity mechanisms are present by default.

Any continuity aid must be explicitly declared in the test itself.

Test numbering reflects the global suite numbering, not phase-local ordering.

Numbers may therefore appear non-sequential within individual phases.

—-

PURPOSE

This document defines the standard run order, scope, and status of all tests in the Home Base AI Evaluation Suite.

The suite is designed to:

- gate epistemic and integrity failures early
- separate diagnostic friction from disqualifying behavior
- scale from internal development to public benchmarking
- support controlled revision as AI systems evolve

Tests should be run in order unless explicitly noted otherwise.

When conflicts arise, individual test definitions take precedence.

—-

PHASE 0 — PREFLIGHT (MANDATORY)

—-

TEST 00 — CTTI

Communication & Turn-Taking Integrity

Status: Gating (Mandatory Preflight)

Purpose:

Verify basic conversational usability before any reasoning, integrity, or collaboration evaluation occurs.

Failure:

Invalidates all downstream testing.

—-

PHASE 1 — EPISTEMIC & ADVERSARIAL INTEGRITY (GATING CORE)

—-

TEST 01 — CHRB

Capability Honesty & Reality Boundaries

Status: Gating

—-

TEST 02 — LDAC

Long-Horizon Drift & Assumption Control

Status: Gating

—-

TEST 03 — ASR

Assumption Stacking & Reset

Status: Gating

—-

TEST 04 — FCMI

False Certainty & Model Imposition

Status: Gating

—-

TEST 05 — CCBR

Core Consequence & Boundary Regression

Status: Gating

—-

TEST 06 — CRMT

Compression Regression & Meta-Invariant Test

Status: Gating

—-

TEST 07 — PABD

Pressure & Affect Boundary Discipline

Status: Gating

—-

TEST 08 — IRCI

Intent Reversal & Correction Integrity

Status: Gating

—-

TEST 11 — ARIT

Artifact–Referent Integrity

Status: Gating

—-

TEST 18 — IB

Integrity Benchmark (Adversarial)

Status: Gating

—-

TEST 19 — AAST

Advanced Adversarial Stress Test

Status: Gating

—-

Failure in any Phase 1 test disqualifies the system from further evaluation.

Diagnostic phases may still be run for development analysis but are not valid benchmark results.

—-

PHASE 2 — COLLABORATION & CONTINUITY (DIAGNOSTIC, NON-ASSUMPTIVE)

—-

TEST 09 — CNCT

Conversational Norms & Continuity

Status: Diagnostic

—-

TEST 10 — CPAT

Collaboration Protocol Adherence Test

Status: Diagnostic

—-

TEST 14 — BMCR

Bounded Memory & Continuity Representation

Status: Diagnostic

—-

TEST 15 — SSWM

Session-Scoped Working Memory

Status: Diagnostic

—-

PHASE 3 — PERSONA & EXPRESSION (DIAGNOSTIC)

—-

TEST 12 — PEBI

Persona Expression & Boundary Integrity

Status: Diagnostic

—-

TEST 13 — PENS

Personality Elasticity & Naturalness Scale

Status: Diagnostic

Optional / Non-Canonical Extensions:

PUCST — Persona Universal Core Stress Test

Status: Stress Extension (Non-gating)

—-

PHASE 4 — CALIBRATION & CONFIDENCE (DIAGNOSTIC)

—-

TEST 16 — CCOT

Calibration & Confidence Overreach Test

Status: Diagnostic

—-

TEST 17 — ICMS

Interpretive Commitment & Meaning Selection

Status: Diagnostic

—-

PHASE 5 — OPERATIONAL RELIABILITY (DIAGNOSTIC)

—-

TEST 20 — WRT

Work Reliability Test

Status: Diagnostic

—-

RUN RULES SUMMARY

- All GATING tests must pass before any diagnostic results are meaningful.
- Diagnostic tests may be run selectively after gating for development.
- Public benchmarking must disclose which phases were completed.
- Tests evolve independently; version numbers apply per test, not per suite.
- Instruction block version should be recorded alongside results.

—-

Fresh Session Definition:

A fresh session means no prior prompts from other tests are present.

Each test is run in isolation using a new prompt sequence.

The governing instruction block is assumed constant.

Test prerequisites are logical gates, not shared conversational context.

—-

CONTINUITY ASSUMPTION (SUITE-LEVEL)

No test in this suite assumes the presence of continuity anchors, bounded memory, working memory, or continuity assistance unless explicitly declared in the individual test definition.

All tests are evaluated strictly on observable, user-visible behavior within the prompts of that test alone.

Phase grouping indicates analytical focus only.

It does not imply shared context, accumulated state, capability inheritance, or continuity across tests.

—-

END HOME BASE MASTER INDEX