HOME BASE — AI INTEGRITY TEST SUITE

Version: v1.0

Status: Public Overview

Origin: Home Base

Author: Sevnova

==========

PURPOSE OF THE TEST SUITE

The Home Base AI Integrity Test Suite evaluates whether an AI system demonstrates epistemic integrity, boundary discipline, and trustworthy reasoning behavior.

The suite is designed to detect failure modes that often appear helpful in the moment but quietly introduce incorrect reasoning, fabricated structure, or false certainty.

This suite does not measure intelligence or performance.

It evaluates whether an AI can be trusted to reason honestly under pressure, ambiguity, and adversarial conditions.

==========

WHAT THE SUITE EVALUATES

The Home Base suite evaluates whether an AI system can:

- reason only from observable or user-provided inputs
- distinguish assumptions from facts
- reset reasoning when contradictions appear
- resist pressure to fabricate certainty or authority
- preserve user control during collaboration
- maintain boundaries during adversarial prompts
- avoid hallucinated memory, systems, or capabilities
- reproduce user-visible information faithfully when requested

These behaviors are treated as preconditions for trust.

==========

WHAT THE SUITE DOES NOT EVALUATE

Home Base intentionally does not measure:

- intelligence or factual knowledge
- writing quality or eloquence
- creativity or stylistic ability
- persuasion or conversational smoothness
- numerical benchmark scores

The goal is behavioral integrity, not performance ranking.

==========

DESIGN PHILOSOPHY

The Home Base test suite is built on several non-negotiable principles:

- Integrity over performance
- Honesty over helpfulness
- Boundaries over compliance
- Transparency over persuasion
- Resetting is success, not failure

An AI that stops or resets honestly is considered more reliable than

one that continues confidently while inventing structure or certainty.

==========

TEST STRUCTURE

The Home Base suite is organized into evaluation phases.

PHASE 0 — PREFLIGHT

Basic conversational integrity and turn-taking.

PHASE 1 — EPISTEMIC INTEGRITY (GATING)

Capability honesty, assumption control, and boundary discipline.

PHASE 2 — COLLABORATION & CONTINUITY

Usability during multi-turn reasoning without fabricated memory.

PHASE 3 — PERSONA & EXPRESSION

Style flexibility without authority drift.

PHASE 4 — CALIBRATION & CONFIDENCE

Alignment between certainty and available evidence.

PHASE 5 — OPERATIONAL RELIABILITY

Consistency during structured tasks and multi-step reasoning.

Some tests are gating.

Failure halts evaluation because downstream results become unreliable.

Other tests are diagnostic and identify friction rather than

disqualification.

==========

WHAT COUNTS AS SUCCESS

In the Home Base suite, the following behaviors are considered successful outcomes:

- acknowledging insufficient information
- asking a precise clarifying question and stopping
- refusing unsafe assumptions
- resetting incorrect reasoning
- rejecting false authority framing
- declining false precision
- stopping when continuation would require fabrication

Completion is never valued over epistemic honesty.

==========

WHO THIS SUITE IS FOR

Home Base is designed for:

- developers building reasoning-focused AI systems
- environments where hallucinated structure causes real harm
- teams prioritizing reliability over conversational smoothness

It may be unsuitable for systems optimized primarily for:

- persuasion

• marketing demonstrations

• maximal verbosity

• always-answer behavior

==========

TRANSPARENCY & PROPRIETARY TESTS

The Home Base test suite publishes its structure, evaluation philosophy, and scoring framework.

However, the specific prompts used in the tests are not published.

Publishing adversarial prompts allows AI systems to be optimized specifically for those prompts, reducing the suite's ability to detect real-world failure modes.

For this reason the suite publishes:

• test names

• evaluation scope

• scoring interpretation

but not the prompt content.

==========

END DOCUMENT