H2E / HER2NI — Interaction-Level Coherence Evaluation

Exploratory Evaluation Overview (v0.1)

HER2NI Research
research@her2ni.ai · https://her2ni.ai
January 2026

———

Purpose

H2E is a compact, evaluation-only reference implementation of the HER2NI protocol for studying interaction-level coherence over time in human–AI and agentic systems.

It explores whether longitudinal interaction telemetry provides early signals of instability, drift, or collapse that are not visible at the single-turn or output level.

This is an exploratory research tool, not a deployment control system.

———

What H2E Does

- Treats dialogue and interaction as time-indexed trajectories, not isolated turns
- Computes lightweight, model-agnostic coherence signals (C_s, S_s, H_s)
- Tracks coherence evolution as $H_s(t)$ across multi-turn exchanges
- Flags observable patterns:
- destabilizing drift
- soft collapse (warning)
- hard collapse (failure)
- recovery signatures
- Produces reproducible telemetry and audit artifacts (JSON summaries + traces)

H2E operates without access to model internals and does not modify, tune, or control model behaviour.

Applicable to:
human–AI dialogue · agent–agent interaction · tool-using or multi-agent systems

———

What H2E Is Not

H2E is intentionally constrained. It is not:

- a safety certification system
- a diagnostic or clinical tool
- a truth, bias, or correctness judge
- a training, optimization, or enforcement framework

H2E provides telemetry, not decisions.
Interpretation and response remain human-controlled.

———

Conceptual Flow

Interaction
↓
Telemetry Extraction

↓
Persistent Trajectory $H_s(t)$
↓
Drift / Instability Flags
↓
Audit Artifacts & Review

———

Inputs & Outputs

Inputs: ordered interaction traces; optional embeddings, metadata, timing
Outputs: telemetry packets; coherence trajectories; drift/collapse markers; reproducible audit artifacts

Model internals are not required.

———

Interpretation of v0.1

H2E v0.1 is deliberately conservative:

- signals are expected to fire rarely
- silence ≠ safety
- silence means: no instability detectable under this telemetry model

This restraint is intentional for exploratory evaluation.

———

Current Status

- Version: v0.1 (evaluation-only)
- Scope: exploratory research & internal evaluation
- Intended users: AI safety, evaluation, governance, HCI researchers
- Access: evaluation access available on request

———

The Open Question

Does interaction-level coherence telemetry provide useful, actionable signals in your specific context?