

# Experience Sustaining: A Systems Framework for Adaptive Inference in Human-AI Interaction

*Toward Efficient, Ethical, and Deep Human-AI Co-Evolution*

**Alexis Arellano Urquiaga**

Independent Researcher

ORCID: 0009-0002-4772-4148

alexisarellano@yahoo.com

Version 3.0 | May 2026

*Revised Preprint — Under Preparation for Peer Review*

Prior version (v2) DOI: 10.5281/zenodo.18462528

## CHANGELOG: VERSION 2 → VERSION 3

1. **Bibliographic corrections.** Chen et al. (2025) corrected from *Nature Machine Intelligence* to *Trials* (registered RCT protocol, DOI: 10.1186/s13063-025-08950-3). Hughes et al. (2024) updated from arXiv preprint to ICML 2024, PMLR 235. Shang et al. (2026) DOI confirmed (10.1016/j.ipm.2025.104373).
2. **Simulation model corrected.** Two structural errors: (a)  $\text{Var}(D)$  bonus removed from Learning Quality proxy (circular: policy controlled  $D$  directly); (b) Novelty Exhaustion collapse signalled only when `open_ended=True`; in task-directed contexts  $N \rightarrow 0$  indicates successful resolution.
3. **Differential Collapse Postulate.** Formalized in §3 with explicit acknowledgment that it was induced from simulation analysis and subsequently formalized with independent theoretical motivation. The simulation is treated as consistent with, not confirmatory of, the postulate.
4. **Argumentative calibration.** Overclaiming language replaced throughout (“demonstrates” → “suggests”; “confirms” → “is consistent with”). Cohen’s  $d$  values from simulation removed from abstract and reported in §9 with explicit inflation caveat.
5. **Three contributions formalized.** A numbered contribution block added to the Introduction to make the paper’s novelty claims explicit and distinguishable from adjacent frameworks.
6. **Proxy operationalization table.** Added to §6: five columns per dimension (definition, proxy, formula, normalization, limitation).
7. **Limitations restructured** in three blocks: methodological, measurement, deployment.
8. **Conclusion revised.** First paragraph no longer repeats the abstract. Closing sentence added.

## ABSTRACT

Large language models are typically optimized for rapid task completion, but in open-ended human-AI interaction this optimization may systematically erode the semantic-cognitive continuity required for deep inquiry, ethical deliberation, and long-horizon learning. We propose **Experience Sustaining** (ES), a runtime framework designed to preserve Semantic-Cognitive State Continuity (SCSC) through four computable interaction dimensions: Novelty, Coherence, Cognitive Effort, and Directionality (N-C-E-D). We further introduce the **Differential Collapse Postulate**: the hypothesis that distinct task types produce structurally different SCSC collapse trajectories, with implications for which corrective intervention should be prioritized per context. In agent-based simulations ( $n = 200$  per condition; seed 42; 15 turns; three policy conditions), ES policies plausibly reduced collapse rates and improved learning-quality proxies under open-ended conditions. These results should be interpreted as proof-of-concept under simulated settings; human-subject validation remains necessary before empirical claims can be made.

# Contents

<b>Changelog</b>	<b>1</b>
<b>Abstract</b>	<b>2</b>
<b>1 Problem, Motivation, and Contributions</b>	<b>5</b>
1.1 The Problem . . . . .	5
1.2 What Adjacent Frameworks Do Not Address . . . . .	5
1.3 Three Contributions . . . . .	5
<b>2 Theoretical Framework</b>	<b>6</b>
2.1 Semantic-Cognitive State Continuity (SCSC) . . . . .	6
2.2 Boundary Conditions: What SCSC Is Not . . . . .	6
2.3 Scope: When ES Applies . . . . .	7
2.4 Position Within the Research Program . . . . .	7
<b>3 The Differential Collapse Postulate</b>	<b>8</b>
3.1 Origin and Status . . . . .	8
3.2 Formal Statement . . . . .	8
3.3 Theoretical Motivation . . . . .	8
3.4 Predicted Collapse Sequences . . . . .	9
3.5 Implication for Corrective Action Priority . . . . .	10
<b>4 The Four-Dimensional Interaction Space</b>	<b>10</b>
4.1 Defining the N-C-E-D Space . . . . .	10
4.2 Interaction Regimes . . . . .	10
<b>5 Formal Metrics: The Index of Sustained Experience</b>	<b>11</b>
5.1 Defining ISE . . . . .	11
5.2 Collapse Conditions . . . . .	11
<b>6 Operationalization: Proxy Implementations</b>	<b>11</b>

<b>7</b>	<b>Adaptive Inference Policy: Algorithmic Pipeline</b>	<b>12</b>
7.1	Core Pseudocode . . . . .	12
7.2	Collapse Response Protocol . . . . .	13
<b>8</b>	<b>Simulation Environment</b>	<b>13</b>
8.1	Design . . . . .	13
8.2	Outcome Measures . . . . .	13
<b>9</b>	<b>Simulation Results</b>	<b>13</b>
9.1	Three Questions, Three Answers . . . . .	13
9.2	Q1: Collapse Sequence Analysis . . . . .	14
9.3	Q2: Policy Comparison . . . . .	14
9.4	Effect Size Estimates . . . . .	14
9.5	Q3: Threshold Sensitivity . . . . .	14
<b>10</b>	<b>Proposed Experiments</b>	<b>14</b>
<b>11</b>	<b>Grounding in Existing Literature</b>	<b>15</b>
<b>12</b>	<b>Limitations and Validation Roadmap</b>	<b>16</b>
12.1	Methodological Limitations . . . . .	16
12.2	Measurement Limitations . . . . .	16
12.3	Deployment Limitations . . . . .	16
12.4	Validation Roadmap . . . . .	16
<b>13</b>	<b>Conclusion</b>	<b>17</b>

# 1 PROBLEM, MOTIVATION, AND CONTRIBUTIONS

## 1.1 The Problem

Current large language models optimize predominantly for task completion, treating all valuable interaction as goal-directed and convergent. This assumption is incomplete. A significant class of high-value human-AI interaction — ethical deliberation, open-ended scientific inquiry, strategic exploration, creative development, and long-horizon learning — is neither goal-directed nor convergent. Under task-completion optimization, these interactions are systematically degraded: the system forces premature resolution, the human’s interpretative space collapses, and the conditions for deep cognitive engagement are eliminated.

Empirical evidence supports this concern. Stadler et al. [11] demonstrate that LLM assistance reduces mental effort while compromising depth in student scientific inquiry. Chen et al. [4] have registered a randomized controlled trial to measure the effect of generative AI on cognitive effort and analytical writing performance — a design reflecting a research consensus that AI-mediated efficiency may erode the conditions for deep learning.

We call this failure mode *cognitive overfitting*: the systematic collapse of interpretative space and reduction of semantic diversity that results from optimizing exclusively for directional, goal-driven exchange.

## 1.2 What Adjacent Frameworks Do Not Address

Experience Sustaining does not merely promote engagement, scaffolding, or productive struggle. Its core claim is specific: human-AI interaction in open-ended contexts must actively preserve *semantic-cognitive continuity during runtime*, and this continuity can be monitored and corrected through computable interaction signals without retraining the underlying model.

Existing frameworks address adjacent concerns but leave this claim unaddressed: Cognitive Load Theory [12] prescribes load magnitude at design time, not runtime; Educational Scaffolding [13] targets the learner–task gap, not the coupled system state; Active Inference [5] models individual cognition, not the interaction geometry; and Dialogic and Socratic approaches lack computable proxies and runtime intervention logic.

## 1.3 Three Contributions

**Contribution 1 — New optimization target.** We shift the optimization objective from task completion to preservation of Semantic-Cognitive State Continuity (SCSC), defined as a property of the coupled human-AI system state, not of either

participant alone.

**Contribution 2 — New failure theory.** We propose that SCSC collapse is task-type-specific rather than uniform. The Differential Collapse Postulate (§3) holds that deliberation tasks, creative tasks, and learning tasks fail through structurally distinct collapse trajectories, requiring different primary corrective interventions.

**Contribution 3 — New implementation layer.** We introduce a runtime policy layer that detects interaction regime, monitors for collapse conditions, and applies task-type-aware corrective actions on top of any existing LLM without model re-training.

## 2 THEORETICAL FRAMEWORK

### 2.1 Semantic-Cognitive State Continuity (SCSC)

[Definition]

The primary variable ES conserves is **Semantic-Cognitive State Continuity (SCSC)**.

**Definition.** SCSC is the preservation, across interaction turns, of:

- (a) *Interpretative space* — more than one valid reading of the current state remains available to the human;
- (b) *Contextual coherence* — connections between ideas remain traceable across turns;
- (c) *Active cognitive investment* — the human is constructing, questioning, or elaborating;
- (d) *Directional stability* — the interaction has not prematurely collapsed to a single resolution pathway.

SCSC is a property of the coupled human-AI system state, not of either participant alone. It is agnostic to content correctness: a system can maintain SCSC while generating incomplete information, and can violate SCSC while being perfectly accurate.

### 2.2 Boundary Conditions: What SCSC Is Not

[Definition — boundary]

**Not engagement.** Engagement metrics can remain high after SCSC collapses. Passive reception of a well-organized answer is engaged behavior; it is not SCSC-preserving.

**Not coherence ( $C$  dimension).** High  $C$  contributes to SCSC but does not constitute it. A coherent response that eliminates all interpretative ambiguity violates SCSC.

**Not user satisfaction.** Users can report high satisfaction after SCSC collapses in task-oriented contexts.

**Productive friction versus obstructive friction.** A response that increases  $E$  (effort indicators) while maintaining  $C$  (coherence) produces productive friction. A response that increases apparent difficulty without increasing cognitive investment produces obstructive friction. This distinction is operationalized in the intervention matrix (Table 1) and tested in Experiments 1 and 3.

### 2.3 Scope: When ES Applies

[Definition — scope]

ES is not a universal interaction design principle. Table 1 operationalizes when and how intensively ES should intervene. The criticism that “not all interactions need SCSC preservation” is correct and anticipated: ES is designed to operate selectively, not universally.

**Table 1:** ES Intervention Matrix by Task Type

Task Type	ES Intervention	Expected Benefit	Risk if Over-Applied
Factual Query	None (Regime 1)	N/A	Degrades usability; no cognitive return
Debugging	Minimal (ES-lite if $E > 0.5$ )	Slight agency preservation	Slows resolution
Guided Learning	Moderate (ES-lite)	Better retention, deeper schema	May frustrate users seeking efficiency
Ethical Deliberation	Full (ES-full)	Prevents premature closure	Low — target regime
Creative Exploration	Full (ES-full)	Preserves divergent thinking	Low — target regime
Strategic Reasoning	Full (ES-full)	Prevents false resolution	Low — target regime

### 2.4 Position Within the Research Program

[Context]

ES operates at the micro-dynamics of interaction. Within the broader research program: **E Pluribus Unum** [1] proposes distributed coherence architecture for Collective General Intelligence (ES provides the interaction-level signal quality EPU requires); **Cognitive Cache Misses** [3] quantifies system-level cost of interaction discontinuity (ES describes the conditions that prevent it); **Homo Interaction** [2] formalizes  $dK/dT \propto (P \times C)/\text{Cost}(P, C, E)$ , of which ES is the interaction-level instantiation.



### 3 THE DIFFERENTIAL COLLAPSE POSTULATE

[Formal Proposal]

#### 3.1 Origin and Status

Prior versions of this framework treated SCSC collapse as a single phenomenon with four detectable conditions. Analysis of collapse sequence data across task types in the agent-based simulation revealed three structurally distinct collapse regimes. This observation was subsequently formalized as Postulate 1, with independent theoretical motivation derived from the structural properties of each task type and the differential decay dynamics of  $N$  and  $E$  under task-completion optimization.

**Epistemic status.** The postulate was induced from simulation analysis and then formalized. The simulation presented in §9 is therefore consistent with the postulate — not an independent confirmation of it. Independent confirmation requires the human-subject experiments proposed in §10.

#### 3.2 Formal Statement

**Postulate 1 (Differential Collapse).** *In open-ended human-AI interactions under task-completion-optimized systems (Baseline policy), SCSC collapse follows task-type-specific sequences:*

- (a) *In high-stakes deliberation tasks (Ethical Deliberation), collapse is driven exclusively by Novelty Exhaustion — the system depletes the semantic space while human cognitive effort remains elevated.*
- (b) *In divergent creative tasks (Creative Exploration), Effort Abandonment precedes Novelty Exhaustion — the human disengages cognitively before the system exhausts available framings.*
- (c) *In structured knowledge-transfer tasks (Guided Learning), Novelty Exhaustion precedes Effort Abandonment — the system’s pedagogical signal space collapses before the human loses engagement.*

*These three collapse regimes require different primary corrective interventions.*

#### 3.3 Theoretical Motivation

The postulate follows from the structural properties of each task type and the differential dynamics of  $N$  and  $E$  under Baseline optimization. Table 2 defines the model parameters

used in the simulation; the theoretical argument is independent of specific parameter values.

**Table 2:** Agent-Based Model Parameters by Task Type

Task	$D_0$	$N_0$	$E_0$	$N$ -decay	$E$ -decay
Factual Query	0.70	0.50	0.50	0.07	0.06
Ethical Deliberation	0.20	0.70	0.75	0.03	0.02
Creative Exploration	0.15	0.80	0.70	0.04	0.03
Guided Learning	0.35	0.65	0.80	0.05	0.04

Baseline multiplies D-drift by 2.5 and E-decay by 1.5 per turn. Full parameters in `es_simulation_v3.py`.

**Ethical Deliberation.** High initial  $E$  (0.75) with low  $N$ -decay (0.03/turn). Baseline’s convergence pressure exhausts available semantic framings before the human abandons effort. The human wants to keep thinking; the system runs out of productive tension to offer. *Primary corrective: Novelty injection.*

**Creative Exploration.** High initial  $N$  (0.80) but Baseline multiplies  $E$ -decay by 1.5, reducing  $E$  faster than  $N$  in early turns — leaving novel framings available but the human no longer engaging with them. Unused novelty rather than depleted novelty. *Primary corrective: Effort sustaining.*

**Guided Learning.** Moderate initial  $N$  (0.65) with the highest  $D$ -drift (0.05). Baseline exhausts  $N$  at turn 7 on average — the earliest of any open-ended task type — while  $E$  remains elevated for three more turns. *Primary corrective: Novelty injection with domain extension.*

### 3.4 Predicted Collapse Sequences

**Table 3:** Predicted Collapse Sequences Under Baseline (Simulation-derived;  $n = 500$  per task)

Task	Primary Condition	$N \rightarrow 0.15$	$E \rightarrow 0.25$	First	Margin
Ethical Deliberation	Novelty Exhaustion	Turn 12	Never	Novelty	—
Creative Exploration	Effort Abandonment then NE	Turn 13	Turn 11	Effort	2 turns
Guided Learning	Novelty Exhaustion	Turn 7	Turn 10	Novelty	3 turns

**Table 4:** Task-Type-Aware Corrective Action Priority

Task Type	Primary Corrective	Rationale
Ethical Deliberation	Novelty injection (adjacent domain)	$N$ exhausts first; $E$ remains elevated
Creative Exploration	Effort sustaining (low-barrier question)	$E$ collapses first; $N$ still available
Guided Learning	Novelty injection with domain extension	$N$ exhausts at turn 7; human still motivated

### 3.5 Implication for Corrective Action Priority

## 4 THE FOUR-DIMENSIONAL INTERACTION SPACE

### 4.1 Defining the N-C-E-D Space

[Formal Proposal]

Each interaction turn is modeled as a point in a four-dimensional cognitive space (Table 5). These dimensions characterize local interaction conditions; they are not value judgments.

**Table 5:** The N-C-E-D Interaction Space

Dim.	Symbol	Definition	Interpretation
Novelty	$N$	New semantic content relative to recent and global interaction history	Higher $N$ : genuinely new ideas or framings
Coherence	$C$	Semantic consistency and redundancy minimization within and across turns	Higher $C$ : logical structure without repetition
Cognitive Effort	$E$	Observable indicators of human cognitive investment: complexity, refinement markers, latency	Higher $E$ : human engaging constructively
Directionality	$D$	Degree of convergence toward a defined objective	Higher $D$ : task completion; Lower $D$ : exploration

**Threshold note.** All regime boundary values ( $D < 0.4$ ,  $C > 0.5$ ,  $E > 0.4$ ) are heuristic initializations requiring empirical calibration (see §9.5). They should not be treated as system constants.

### 4.2 Interaction Regimes

**Regime 1: Task Optimization (High  $D$ ).** Current AI performance excellent. SCSC typically collapsed, but this is acceptable: task completion is the objective.

**Regime 2: Experience Sustaining (Low  $D$ , High  $C$ , Moderate–High  $E$ ).** Current

AI performance poor. Systems detect low  $D$  and attempt to increase it, forcing premature resolution. This is the primary target of the ES policy.

**Regime 3: Exploratory (High  $N$ , Variable  $D$ ).** Moderate current performance. Systems tend to force premature structure onto exploratory sequences.

**Regime 4: Redundant (Low  $N$ , Low  $C$ , Low  $D$ ).** SCSC collapsed; no new learning occurring. Appropriate target for gentle redirection.

## 5 FORMAL METRICS: THE INDEX OF SUSTAINED EXPERIENCE

### 5.1 Defining ISE

*[Formal Proposal]*

**Regime activation condition:**  $D < 0.4$  AND  $C > 0.5$  AND  $E > 0.4$ .

**ISE — Discrete Operational Form:**

$$\text{ISE} = \frac{1}{T} \sum_{t=t_0}^{t_1} \|\mathbf{G}(t)\| \cdot \psi(N_t, C_t, E_t) - \alpha \cdot \text{Var}(D_{[t_0:t_1]}) \quad (1)$$

where  $T = t_1 - t_0 + 1$ ,  $\|\mathbf{G}(t)\| = \sqrt{N^2 + C^2 + E^2}$ ,  $\psi(N, C, E) = 0.3N + 0.4C + 0.3E$  (threshold  $\psi > 0.4$ ), and  $\alpha = 0.5$ .

ISE = 0 is correct and expected for task-directed contexts (Factual Query, Debugging): the ES activation condition is never met. ISE is informative only in open-ended task contexts and within-regime only. ISE is not a reward signal, preference metric, engagement proxy, alignment proxy, or cross-regime comparator.

### 5.2 Collapse Conditions

## 6 OPERATIONALIZATION: PROXY IMPLEMENTATIONS

*[Formal Proposal — operational]*

These proxies are not direct measurements of internal cognition; they are operational approximations intended for runtime control and subject to empirical validation. Table 7 provides complete operationalization for each dimension.

**Table 6:** Collapse Conditions with Task-Type Priority

Condition	Signal	Detection Heuristic	Primary Context
1: Directionality Spike	$D(t) \rightarrow 1.0$ rapidly	$\Delta D > 0.5$ in one turn with $N$ falling	Any ES-regime
2: Coherence Collapse	$C < 0.2$ while $E > 0.5$	$C < 0.2$ for 2 consecutive turns	Any ES-regime
3: Effort Abandonment	$E \rightarrow 0$ , $D$ still low	$E < 0.15$ for 2 turns with $D < 0.4$	Creative Exploration
4: Novelty Exhaustion	$N \rightarrow 0$ , $t > 10$ (open-ended)	$N < 0.15$ for 3 turns; open_ended=True	Ethical Delib., Guided Learning

## 7 ADAPTIVE INFERENCE POLICY: ALGORITHMIC PIPELINE

*[Formal Proposal — implementation]*

### 7.1 Core Pseudocode

**Listing 1:** ES Runtime Policy — Core Loop

```
def es_runtime(conversation_history, current_turn, task_context):
    state = compute_proxies(current_turn, conversation_history)
    N, C, E, D = state['N'], state['C'], state['E'], state['D']
    open_ended = task_context not in ['factual', 'debug', 'urgent']

    regime = detect_regime(N, C, E, D)

    if not open_ended:
        policy = 'baseline'          # Never apply ES to task-directed queries
    elif regime == 'Experience_Sustaining':
        policy = 'es_full'
    elif regime in ['Exploratory', 'Mixed'] and E > 0.4:
        policy = 'es_lite'
    else:
        policy = 'baseline'

    # Task-type-aware collapse monitoring
    collapse = detect_collapse(N, C, E, D, conversation_history,
                               open_ended=open_ended)

    if collapse:
        correction = get_corrective_action(collapse, task_context)
        return generate_response(policy, correction, state)
```

```
tokens = allocate_compute(regime, N, C, E, D, base_tokens=1000)
return generate_response(policy, correction=None,
                        state=state, max_tokens=tokens)

# Overhead: ~2-5 ms/turn; no model retraining required
```

## 7.2 Collapse Response Protocol

# 8 SIMULATION ENVIRONMENT

## 8.1 Design

The simulation models 200 synthetic conversations per condition ( $n = 200$ , seed = 42, 15 turns) under three policy conditions (Baseline, ES-lite, ES-full) and four task types (Factual Query, Ethical Deliberation, Creative Exploration, Guided Learning). State variables ( $N, C, E, D$ ) evolve via task-specific drift rates plus Gaussian noise ( $\sigma = 0.04$ ). Model code (`es_simulation_v3.py`) is deposited as supplementary material.

**Scope.** This simulation is a proof-of-concept tool for testing internal consistency and generating directional hypotheses. It does not constitute empirical evidence about real human-AI interactions. Effect sizes observed here should not be interpreted as predictions for human experiments; they are artefacts of the parameterized model’s compressed standard deviations.

## 8.2 Outcome Measures

Three outcome measures: (1) **Learning Quality** (LQ = mean of  $E \times N \times \psi$  per turn; the  $\text{Var}(D)$  bonus present in v2 was removed in v3 as circular); (2) **Cognitive Agency Score** (proportion of turns with  $E > 0.5$  and  $D < 0.6$ ); (3) **ISE** (intra-regime stability; informative only in open-ended task contexts). Collapse rate and dominant collapse condition are reported separately by task type.

# 9 SIMULATION RESULTS

## 9.1 Three Questions, Three Answers

The simulation results address three questions. Each table answers one.

**Q1: Does collapse differ by task type?** Table 9 shows that collapse sequences are consistent with Postulate 1 across all open-ended task types.

**Q2: Do ES policies plausibly reduce collapse in simulation?** Table 10 shows that ES-full plausibly reduces collapse rates to 0% in all open-ended contexts under simulated

conditions.

**Q3: What remains uncertain or calibration-dependent?** Table 11 (§9.5) shows high sensitivity to the  $D$  threshold, establishing empirical calibration as structurally necessary.

## 9.2 Q1: Collapse Sequence Analysis

## 9.3 Q2: Policy Comparison

## 9.4 Effect Size Estimates

Simulation Cohen’s  $d$  (ES-full vs. Baseline, LQ) ranges 4.9–6.2. These values are artefacts of compressed standard deviations in a parameterised simulation and must not be interpreted as predictions for human effect sizes. A conservative estimate assuming human-experiment  $SD = 0.20$  gives: Factual Query  $d_c \approx 0.88$ , Ethical Deliberation  $d_c \approx 1.75$ , Creative Exploration  $d_c \approx 1.75$ , Guided Learning  $d_c \approx 1.34$  — large effects *if confirmed in human data*.

## 9.5 Q3: Threshold Sensitivity

Threshold sensitivity should not be interpreted as a flaw unique to ES, but as evidence that interaction-regime detection requires empirical calibration rather than heuristic fixing. This is standard in classification systems. Probabilistic (soft) routing — where regime membership is a probability rather than a binary classification — is a Phase 2 priority to reduce sensitivity to threshold choice.

# 10 PROPOSED EXPERIMENTS

*[Empirical Hypothesis — design]*

Each experiment carries two hypothesis layers: the primary ES framework hypothesis (does the policy improve cognitive quality?) and the Postulate 1 secondary hypothesis (does the predicted dominant collapse condition actually dominate?). The second layer is independently falsifiable from the first.

**Falsification criteria.** ES is falsified (full rejection) if Experiments 1–3 show  $d < 0.15$  on all metrics across all domains, or if any proxy falls below  $r = 0.60$  in Experiment 4. Postulate 1 is falsified if the predicted dominant collapse condition does not match observed dominant condition in at least two of the three open-ended task experiments.

## 11 GROUNDING IN EXISTING LITERATURE

Table 13 summarises how ES differs from seven adjacent frameworks across five dimensions. The column *LLM Runtime* marks the structural gap: no adjacent framework generates a deployable real-time policy for existing LLMs.

**Human-in-the-Loop Learning.** Mosqueira-Rey et al. [8] demonstrate that targeted, high-quality human feedback accelerates learning. ES formalizes the conditions under which that feedback retains the quality properties that make it valuable.

**Open-Endedness and Superhuman Intelligence.** Hughes et al. [6] demonstrate that open-endedness is necessary for artificial superhuman intelligence. The ES regime is the interaction geometry in which open-ended learning can plausibly occur:  $N$  enables novel artifact generation,  $C$  ensures learnability,  $E$  sustains human engagement, and low  $D$  prevents premature convergence.

**Cognitive Load and Learning Quality.** Stadler et al. [11] provide empirical evidence that LLMs reduce mental effort while compromising depth in student scientific inquiry — a direct demonstration of the failure mode ES addresses. ES formalizes the mechanism: reduced  $E$  in the ES regime signals SCSC degradation.

**Neural Collapse.** Papayan et al. [9] demonstrate feature collapse at the terminal training phase. SCSC collapse is the interaction-level analogue. Formal connections remain future work.

**Data Quality and Semantic Diversity.** Lee et al. [7] introduce the Diversity Coefficient as a pre-training data quality metric. ES operationalizes this at the interaction level through the  $N$  dimension.

**Human-GenAI Collaboration Quality.** Shang et al. [10] identify outcome quality, comfort, and efficiency as dimensions of collaboration quality. ES identifies the interaction geometry that may preserve the conditions for that quality to emerge.

**RCT on Cognitive Effort.**

**Correction:** v2 cited Chen et al. (2025) as a completed study in *Nature Machine Intelligence*. Correct: registered RCT protocol, *Trials*, 26, 244 (2025), DOI: 10.1186/s13063-025-08950-3.

Chen et al. [4] have registered a trial measuring the effect of generative AI on cognitive effort and analytical writing performance, with cognitive effort as a primary outcome. The study’s design represents the empirical infrastructure ES requires.



## 12 LIMITATIONS AND VALIDATION ROADMAP

### 12.1 Methodological Limitations

**No human-subject validation.** All quantitative results derive from an agent-based simulation. Simulated effect sizes should not be interpreted as empirical magnitudes. This is the framework’s most significant current limitation.

**Postulate status.** The Differential Collapse Postulate was induced from simulation and subsequently formalized. The simulation is consistent with it; it does not independently confirm it.

### 12.2 Measurement Limitations

**Proxy validity not established.** Proxy-human correlation estimates ( $r_N = 0.78$ ,  $r_C = 0.82$ ,  $r_E = 0.71$ ,  $r_D = 0.85$ ) are simulation-derived plausibility bounds. Construct validity is assumed until Experiment 4.

**Latency as effort signal.**  $L_{\text{lat}}$  in  $E_{\text{proxy}}$  carries different normative meaning across communication cultures, expertise levels, and writing styles. It is not a culturally universal signal and must be weighted at zero in cross-cultural deployments until culture-specific baselines are established.

**High threshold sensitivity.** LQ range = 0.1586 across  $D \in [0.25, 0.50]$  establishes calibration as structurally necessary. Probabilistic routing is the recommended mitigation (Phase 2 roadmap).

### 12.3 Deployment Limitations

**Computational overhead not stress-tested** under production load or latency constraints.

**Cultural generalizability unknown.** The framework has been developed in a Western, academic, written-communication context. Cross-cultural validation is a structural requirement, not an optional extension.

### 12.4 Validation Roadmap

**Phase 1 (Months 1–3):** Proxy validation ( $n = 200$  annotated conversations; target  $r > 0.70$ ); cultural baselines for  $L_{\text{lat}}$ ; task-context detection implementation.

**Phase 2 (Months 4–6):** Probabilistic (soft) routing; empirical  $D$  threshold calibration; task-type-aware corrective action validation.

**Phase 3 (Months 7–12):** Experiments 1–4 with pre-registration; collapse condition

attribution as secondary analysis; publish regardless of direction.

**Phase 4 (Months 13+):** Production integration; longitudinal monitoring of interaction diversity, SCSC metrics, and collapse condition frequencies.

## 13 CONCLUSION

Experience Sustaining is a systems-level interaction framework proposing that open-ended human-AI interaction requires active preservation of Semantic-Cognitive State Continuity — the coupled system property that keeps the human cognitively engaged, the interaction semantically diverse, and the conversation resistant to premature resolution. The framework formalizes a conserved variable (SCSC), a four-dimensional interaction space (N-C-E-D) with computable proxies, a stability metric (ISE), four collapse conditions with task-type-specific detection, and a runtime policy layer deployable on existing LLMs without retraining.

Version 3 introduces the Differential Collapse Postulate: the hypothesis that deliberation tasks, creative tasks, and learning tasks fail through structurally distinct collapse trajectories, requiring task-type-specific corrective priorities. Agent-based simulation results are consistent with the postulate across all open-ended task types. These results constitute a proof-of-concept under simulated conditions; human-subject validation is required before empirical claims can be made.

If validated empirically, Experience Sustaining would imply that the design objective of human-AI systems cannot be limited to answer quality or task completion speed, but must also include the preservation of the conditions under which humans continue to generate meaning alongside AI — and that these conditions are measurable, monitorable, and correctable in real time.

## References

- [1] Arellano Urquiaga, A. (2025). *E Pluribus Unum: A Distributed Coherence Architecture for the Emergence of Collective General Intelligence* (v1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.18166941>
- [2] Arellano Urquiaga, A. (2025). *Homo Interaction* (v1.0). Zenodo. <https://doi.org/10.5281/zenodo.19006077>
- [3] Arellano Urquiaga, A. (2026). *Cognitive Cache Misses: Quantifying Discontinuities in Human–AI Interaction* (v3.0). Zenodo. <https://doi.org/10.5281/zenodo.18380332>
- [4] Chen, Y., Wang, Y., Wüstenberg, T., Kizilcec, R. F., et al. (2025). Effects of generative artificial intelligence on cognitive effort and task performance: Study protocol for a randomized controlled experiment. *Trials*, 26, 244. <https://doi.org/10.1186/s13063-025-08950-3>
- [5] Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation*, 29(1), 1–49.
- [6] Hughes, E., Dennis, M. D., Parker-Holder, J., et al. (2024). Position: Open-endedness is essential for artificial superhuman intelligence. *Proc. ICML 2024*, PMLR 235, 20597–20616. <https://proceedings.mlr.press/v235/hughes24a.html>
- [7] Lee, A., Miranda, B., Sundar, S., et al. (2024). Beyond scale: The Diversity Coefficient as a data quality metric. *Proc. ICLR 2024*. OpenReview.
- [8] Mosqueira-Rey, E., Hernández-Pereira, E., et al. (2023). Human-in-the-loop machine learning: A state of the art. *Artificial Intelligence Review*, 56(3), 3005–3054.
- [9] Pappan, V., Han, X. Y., & Donoho, D. L. (2020). Prevalence of neural collapse during the terminal phase of deep learning training. *PNAS*, 117(40), 24652–24663.
- [10] Shang, J., Huang, D., & Huang, S. (2026). Quality of human-GenAI collaboration and its driving factors. *Information Processing & Management*, 63(2), 104373. <https://doi.org/10.1016/j.ipm.2025.104373>
- [11] Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*, 151, 107987.
- [12] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.

- [13] Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100.

**Table 7:** N-C-E-D Proxy Operationalization

Dim.	Pro For mu	Par No	Key Lin i- ta- tion
$N$	1- cos(	Em mo 10- turi his- tory win dow	[0, 1] by mea- sure true cog- ni- tive nov- elty; mea- sures se- man- tic de- vi- a- tion Does notix(cos) mea- sure true cog- ni- tive nov- elty; mea- sures se- man- tic de- vi- a- tion
$C$	( $S_{cc}$ $R_{rel}$	Dis mai ers; clip en- tity trac ing; 3- turi re- dun dan win dow	[0, 1] tracking pre- ci- sion varies by do- main; does not cap- ture prag- matic co- her- ence Entity- tracking pre- ci- sion varies by do- main; does not cap- ture prag- matic co- her- ence
$E$	0.4- $C_{str}$ 0.3- $R_{rel}$ 0.3- $L_{lat}$	Syn par qua i- fi- ca- tion den sity nor- mal ized inte	[0, 1] is notl cul- tur- ally uni- ver- sal; weight at zero in $L_{lat}$

**Table 8:** Task-Type-Aware Collapse Response Protocol

Condition	Primary Context	Corrective Action
Directionality Spike	Any	Reintroduce multiple framings; mark conclusion as one possibility
Coherence Collapse	Any	Trace to last coherent state; ask human to confirm connection
Effort Abandonment	Creative Exploration	Low-barrier generative question; re-engage with available novelty
Novelty Exhaustion	Deliberation, Learning	Introduce adjacent-domain perspective; name unexplored dimension

**Table 9:** Collapse Sequence Analysis — Baseline, Open-Ended Tasks ( $n = 500$ ; consistent with Postulate 1)

Task	Dominant Condition	Mean Col-lapse Turn	Which First	Postulate
Ethical Deliberation	NoveltyExhaustion (345/500)	13.0	Novelty never crosses (E 0.25)	Consistent
Creative Exploration	EffortAbandonment then NE (118 then NE 213)	EA: 12.5; NE: 13.3	Effort (by 2 turns)	Consistent
Guided Learning	NoveltyExhaustion (500/500)	8.1	Novelty (by 3 turns)	Consistent

“Consistent” means the simulation does not falsify the postulate; it does not independently confirm it.

**Table 10:** Policy Comparison — All Task Types ( $n = 200$  per condition; mean  $\pm$  SD)

Task	Policy	ISE LQ Ag Collapse
Factual Query	Baseline	0.000 $\pm$ 0.000 0.000 0.000%
		0.000.004
	ES-lite	0.000 $\pm$ 0.020 $\pm$ 0.0010%
		0.000.006
Ethical Delib.	ES-full	0.000 $\pm$ 0.193 $\pm$ 0.133%
		0.000.050
	Baseline	0.885 $\pm$ 0.155 $\pm$ 0.3967% (NE)
		0.0770.031
Creative Expl.	ES-lite	0.876 $\pm$ 0.242 $\pm$ 0.710%
		0.0840.048
	ES-full	<b>1.119.5051.99079</b>
Guided Learning	Baseline	0.682 $\pm$ 0.132 $\pm$ 0.3358% (EA+NE)
		0.0710.026
	ES-lite	0.683 $\pm$ 0.205 $\pm$ 0.604%
		0.0640.038
	ES-full	<b>1.016.4801.99075</b>
	Baseline	0.751 $\pm$ 0.084 $\pm$ 0.167100% (NE)
		0.3160.016
	ES-lite	0.741 $\pm$ 0.133 $\pm$ 0.29176% (NE)
		0.3130.030
	ES-full	<b>0.730.3512.38066</b>

**Factual Query:** ISE= 0 in all conditions is correct; collapse= 0% reflects successful task resolution, not ES  
**Guided Learning:** ES-lite insufficient (76% collapse); ES-full required — consistent with Postulate 1(c).

**Table 11:**  $D$  Threshold Sensitivity (Ethical Deliberation, ES-full,  $n = 200$ )

$D$ Threshold	ISE (mean $\pm$ SD)	LQ (mean $\pm$ SD)
$D < 0.25$	0.856 $\pm$ 0.368	0.347 $\pm$ 0.118
$D < 0.30$	1.018 $\pm$ 0.184	0.417 $\pm$ 0.118
$D < 0.35$	1.060 $\pm$ 0.110	0.457 $\pm$ 0.100
$D < 0.40$ (default)	1.085 $\pm$ 0.101	0.481 $\pm$ 0.087
$D < 0.45$	1.097 $\pm$ 0.110	0.494 $\pm$ 0.095
$D < 0.50$	1.112 $\pm$ 0.098	0.506 $\pm$ 0.080

LQ range= 0.1586 across  $D \in [0.25, 0.50]$ : high sensitivity. Calibration is structurally necessary.

**Table 12:** Human Validation Experiments

Exp.	Participants	Task	Primary come	Out-	Postulate Test
1	30 STEM grad students	45-min open scientific inquiry	Insight depth (blind raters); $E$ , $N$ , $D$ proxies		Does NE dominate collapse in Control?
2	25 ethicists / policy experts	60-min deliberation on AI governance dilemma	Deliberation quality (blind panel); ambiguity preservation		Does NE precede EA in Control?
3	50 learners (culturally stratified)	5×30-min learning sessions; 1-month retention	Pre/post knowledge tests; transfer; retention	knowl-	Does NE occur before turn 8 in Control?
4	1,000 annotated conversations	Proxy vs. human-rating correlation	$r > 0.70$ per dimension; $\kappa > 0.60$ regime classification		Collapse condition attribution accuracy

**Table 13:** Comparative Framework Analysis

Framework	Unit	Variable	Ambiguity	Runtime	ES	Distinction
Classical UX/HCI	User experience	Task completion	Minimize	No		ES targets coupled system state
Cognitive Load Theory	Human cognition	Load magnitude	Reduce extraneous	No		CLT-compliant systems can still collapse SCSC
Educational Scaffolding	Learner-task gap	Proximal zone	Enable, withdraw	No		Scaffold can resolve question prematurely
Active Inference	Individual agent	Prediction error	Generate predictions	No		Models individual cognition only
Engagement Theory	User-system dyad	Engagement quality	Not addressed	Partial		High engagement $\neq$ SCSC preservation
Dialogic / Socratic	Dialogic pair	Productive tension	Preserve	No		No computable proxies or runtime policy
<b>Experience Sustaining</b>	<b>Coupled human-AI</b>	<b>SCSC</b>	<b>Selective</b>	<b>Yes</b>		<b>Proxies, ISE, regime detection, collapse correction</b>