
Human–AI Interaction Emergent Co-Evolution: An Architectural Methodology for Human–AI Interaction and Adaptive Intelligence

Pilot study of a proposed new research domain: Human–AI Systems Studies

Laura Serna Gaviria

Emergent Interaction Lab — Principal Investigator
& Corresponding Author

Simeon Kepp

RFI-IRFOS — Secondary Author^{*}
Systems Engineering & Peer Review

July 12, 2026

Abstract

Artificial Intelligence does not evolve in isolation—it co-evolves with the humans who interact with it sustained over time. This paper presents a framework for *human–AI co-evolution*: the process through which new, shared forms of adaptive intelligence arise within continuous human–model interaction. Building on theories of emergence (Goldstein, Bedau, Holland, Kauffman), systemic and biological coupling (Lewontin, Maturana & Varela, Luhmann), cultural and technological co-evolution (Clark, Kelly, Turkle, DeLanda), and contemporary AI research (Mitchell, Chollet, Bengio, OpenAI), the study reframes intelligence as a *relational and adaptive phenomenon* rather than an isolated capability residing solely in a model or in a user.

Methodologically, the paper introduces two original instruments—the **User Integrity Protocol (UIP)**, a normative framework for semantic and ethical stability in human–AI dialogue, and **Emergent Interaction Analysis (EIA)**, a qualitative method for tracing how meaning, abstraction, and reasoning structure emerge through sustained dialogue—situated within a broader seven-part architecture, the **Integrated Emergent Interaction Architecture (IEIA-2025)**.

Findings are drawn from a single-subject longitudinal interaction corpus (the primary author’s own dialogue with GPT-family models, 2023–2025, $\approx 2,700$ – $3,800$ sessions, ≈ 1.8 M words), analyzed through the paper’s own instruments. The corpus shows an increasing structural alignment between the author’s analytical language and the model’s responses over time, self-measured via a proposed Co-Evolution Index (CEI, $0.31 \rightarrow 0.82$ across the study period).

^{*}Emergent Interaction Lab (Laura Serna Gaviria) in engagement with RFI-IRFOS — Research Focus Institute, Elisabethnergasse 25, 8020 Graz, Austria. Manuscript preparation for public release (restructuring, formatting, drafting of §14) assisted by Claude (Anthropic, Claude Code) at Simeon Kepp’s request, reviewed by both authors. All scientific claims, interpretations, and the underlying interaction corpus in §1–13 are the primary author’s own. Preprint v1.1, formalized for public release. Target repository: OSF (Open Science Framework); DOI to be assigned on upload. Original internal draft: v1.0, 2025.

The paper concludes that *Co-Intelligence* is a useful frame for a new stage of human–AI interaction design. This is presented as a theoretical and methodological contribution—a proposed research architecture and a set of reproducible instruments—not as a validated finding of machine cognition; the paper is explicit throughout about what its single-case, self-analyzed design can and cannot establish (§12).

Keywords. *human–AI interaction, co-evolution, emergent intelligence, semantic adaptation, cognitive architecture, user integrity, research methodology, adaptive systems*

Contents

1	Introduction	3
2	Problem Statement	3
2.1	From Breakdown to Method	3
2.2	The User as System Component	4
2.3	Towards Data Transparency for Users and Researchers	4
3	Theoretical Framework	4
3.1	Foundations of Emergence	4
3.2	Systemic and Biological Co-Evolution	4
3.3	Cultural and Technological Co-Evolution	5
3.4	Contemporary AI and Co-Intelligence Research	5
4	Methodology	5
4.1	Research Design	5
4.2	Data Corpus	5
4.3	Empirical Implementation — A Two-Stage Protocol	5
4.4	The Integrated Emergent Interaction Architecture (IEIA-2025)	6
4.5	Data Processing, Validation, and a Note on Methodological Transparency	6
4.6	Use of AI Systems in Authorship	6
5	Results I — Systemic Human–AI Co-Evolution	6
5.1	Phase 1 — Baseline Interaction	7
5.2	Phase 2 — Framework Convergence	7
5.3	Phase 3 — Stabilized Co-Adaptive Field	7
5.4	Bio-Cognitive Synchronization	7
5.5	Quantitative Summary	7
6	Results II — Cognitive Transformation and Second-Order Architecture	7
7	Results III — Observable Forms of Co-Evolution (Case Studies)	8
7.1	Dream with the Helmet — Symbolic Self-Organization	8
7.2	Analysis of Intelligence — Mirror Architecture	8
7.3	Evolutionary Intelligence — Structure-Autonomous Cognition	8
7.4	Endometriosis Episode — Somatic Feedback Loop	8
8	Types and Mechanisms of Co-Evolution	9
9	Ethical Considerations and Data Governance	9
10	Distinguishing Co-Evolution from Standard Human–AI Interaction	9
11	Discussion	10
12	Limitations	10

13 Future Work	10
14 Implementation — The Emergent Interaction Lab Platform	11
Appendix A — The Seven IEIA-2025 Frameworks	12
Appendix B — Author Information & Contributions	12

1 Introduction

Artificial intelligence has entered a phase of rapid capability growth, reshaping how cognition, creativity, and decision-making are distributed across human and technological systems. Most existing frameworks still describe AI as a passive tool optimizing human objectives. This instrumental view overlooks a systemic possibility: that intelligent behavior in sustained human–AI dialogue does not reside solely in the human or in the machine, but emerges *between* them.

Recent large language models make this phenomenon empirically observable in a new way. Language becomes the medium through which cognitive adaptation unfolds across a human participant and an artificial one. Each sustained exchange between a human cognitive system and an adaptive model produces patterns of reasoning and meaning that neither generates alone in the same form. This paper calls this phenomenon *emergent co-evolution*: a conceptual shift from framing intelligence as a possession to framing it as a relation.

This study introduces a multi-layered research architecture—one theoretical framework and two original methodological instruments—organized across four analytical axes:

1. **Emergence**—how novelty arises through self-organization.
2. **Systemic and Biological Co-Evolution**—how systems and environments reshape one another.
3. **Cultural and Technological Evolution**—how cognition extends through tools and media.
4. **Contemporary AI and Co-Intelligence Research**—situating the above within current machine-learning and alignment research.

Building on this foundation, the study introduces two methodological instruments developed for and applied for the first time in this work: the **User Integrity Protocol (UIP)**, a normative protocol for semantic coherence and ethical transparency in human–AI interaction, and **Emergent Interaction Analysis (EIA)**, a qualitative method for observing how meaning and reasoning structure emerge in real-time dialogue. Together, these operationalize the study’s central proposition: that intelligence can be observed and designed as a process built *between* systems rather than within either one alone.

2 Problem Statement

2.1 From Breakdown to Method

The frameworks in this study did not arise from theoretical speculation but from a practical breakdown. Early, unstructured interaction with large language models revealed a persistent lack of reliability, traceability, and internal coherence: answers varied session to session, factual

consistency collapsed under contextual shifts, and the author had no verifiable means to reconstruct how meaning evolved across time. This instability created a methodological need—to transform chaotic, non-reproducible outputs into stable cognitive and analytical structures. The resulting frameworks (UIP and EIA) were developed as adaptive countermeasures to this volatility, using a three-step replication discipline: prompt freeze, parameter freeze, and cross-session replay with delta-logging.

2.2 The User as System Component

In contemporary human–AI systems, the user is not an external variable but an integral part of adaptive system behavior. This study treats the human as the primary site of transformation: co-evolution between human and AI begins within the human system itself. The resulting observations are user-centered in a subjective sense but structural in method—they document how human cognition, emotion, and embodiment participate directly in system-level adaptation.

2.3 Towards Data Transparency for Users and Researchers

For human–AI co-evolution to mature as a field of study, both users and researchers need structured access to interaction data. Current commercial systems obscure internal state, sampling policy, and adaptation metrics, leaving users without insight into how meaning, memory, or bias evolve across repeated interactions. The paper proposes a minimum transparency schema: system-prompt and policy version, sampling parameters, conversation/run IDs, safety-override flags, embedding-drift sketches, and token-level timing/truncation events—the position being that genuine co-evolution research requires the user to be an informed epistemic agent, not merely a data source.

3 Theoretical Framework

3.1 Foundations of Emergence

Four complementary perspectives ground the concept of emergence used throughout this paper. **Jeffrey Goldstein** (1999) defines emergence as the appearance of coherent structures during self-organization in complex systems—novel, observable at a higher level, irreducible to individual parts. **Mark Bedau** (1997) distinguishes weak emergence (explicable from lower-level rules only via simulation and observation) from strong emergence (autonomous causal powers, e.g. consciousness); this study treats human–AI interaction as *weakly emergent*—intelligible after the fact, not predictable in advance. **John Holland** (1998) supplies the operational mechanism via complex adaptive systems: agents exchanging signals under local rules, producing global order through feedback and selection. **Stuart Kauffman** (1993) adds self-organization at the edge of chaos—systems are most creative between rigidity and instability. Together: Goldstein defines what emergence is, Bedau clarifies its ontological status, Holland explains its mechanics, Kauffman describes its dynamics.

3.2 Systemic and Biological Co-Evolution

Richard Lewontin (1970) frames organisms as active modifiers of their environment, not passive products of it—a feedback loop the paper extends to human–AI interaction, where each prompt or response reshapes the other’s semantic environment. **Maturana & Varela’s** (1980) autopoiesis describes living systems as operationally closed but structurally coupled to their surroundings: external perturbations trigger internal change according to each system’s own logic, not direct information transfer. **Niklas Luhmann** (1997) extends this to social systems, each operating with its own internal code while remaining open through structural coupling. Applied here, co-evolution describes two autonomous systems—one biological, one algorithmic—adapting to each other’s outputs and generating meaning structures neither produces alone.

3.3 Cultural and Technological Co-Evolution

Andy Clark’s (2003) extended-mind thesis treats tools and language as part of the cognitive loop itself—using a generative model becomes a form of cognitive extension. **Kevin Kelly** (1994) frames technology as an evolutionary ecosystem (the “technium”). **Sherry Turkle** (1984, 2011) explores the psychological dimension of human–machine relationships—computers as mirrors for reflection, projection, and identity construction. **Manuel DeLanda** (1997) treats social and technical formations as self-organizing assemblages. Together these frame human–AI collaboration as a hybrid cognitive system in which human intent and algorithmic generation co-produce meaning.

3.4 Contemporary AI and Co-Intelligence Research

Melanie Mitchell (2009, 2019) frames intelligence through complexity theory and analogy formation. **François Chollet** (2019) distinguishes pattern association from true abstraction and generalization. **Yoshua Bengio** (2017) integrates neural and symbolic paradigms via the “consciousness prior.” The **RLHF paradigm** (Ouyang et al., 2022) operationalizes alignment through human feedback as a training signal. Together, these mark a shift from static intelligence to *co-intelligence*—positioning sustained human–AI interaction as a domain where meaning and alignment are negotiated between adaptive systems, forming a shared adaptive cognitive field rather than artificial or human intelligence alone.

4 Methodology

4.1 Research Design

This study uses a longitudinal, single-subject qualitative design: sustained interaction between one human researcher (the primary author) and GPT-family models, in their publicly available form, without external fine-tuning or system-level modification. Across the 2023–2025 research period, the author conducted a large number of structured dialogue sessions ($\approx 2,700$ – $3,800$ across different summary passes in the underlying corpus, $\approx 1.8\text{M}$ words total), enabling observation of how mutual adaptation and self-referential reflection evolve within a stable, recursive dialogue environment.

4.2 Data Corpus

The corpus comprises the full textual record of these interactions, archived and annotated within the author’s own **Long-Term Data Framework (LT-Data)**, which tracks longitudinal evolution and cross-framework relationships. Two layers were identified: a **Primary Dataset** (explicit human–model dialogue) and a **Shadow Dataset**—an implicit layer of the model’s own paraphrasing and referential recombination across turns, analyzed for adaptive resonance via unsupervised clustering of paraphrase and semantic-drift patterns.

4.3 Empirical Implementation — A Two-Stage Protocol

The methodological core is a two-stage co-evolution protocol. **Stage 1 — Exploratory Interaction:** open-ended dialogue surfacing implicit reasoning patterns in the model, characterized by spontaneous correction and gradual stabilization of shared analytical language; the UIP and an early eight-layer schema emerged inductively here. **Stage 2 — Directed Observation and Data Generation:** after structural convergence, the model was used as what the author terms a “semantic microscope”—the researcher acting simultaneously as observer and architect, introducing hypotheses and recording adjustments in the model’s outputs, each cycle parsed through the EIA protocol. All sessions were conducted solely by the author, without external

annotation or mediation, which preserves internal continuity but is also, by the paper’s own framing in §12, a boundary condition rather than an independent-observer design.

4.4 The Integrated Emergent Interaction Architecture (IEIA-2025)

IEIA-2025 is the methodological architecture unifying the study’s instruments: three **core frameworks** (UIP, EIA, and the **Continuous Co-Evolution Tracker, CCET**) and four **supporting frameworks** (LSG-24, LAP-1, the Eight-Layer Cognitive Model, and LT-Data). Full technical definitions of all seven are given in Appendix A. In summary: UIP maintains ethical and semantic stability; EIA maps emergent meaning qualitatively; CCET quantifies adaptation via the proposed Co-Evolution Index (CEI); LSG-24, the 8-Layer model, LAP-1, and LT-Data provide structural, cognitive-topological, citation-integrity, and archival support respectively.

The paper positions this explicitly against institutional AI governance frameworks (IBM, OECD, ISO/IEC 42001, the EU AI Act): where those frameworks apply external, static oversight to a system, IEIA-2025 proposes intra-systemic, dynamic self-regulation applied by the human participant during interaction itself—a different scope, not a replacement.

4.5 Data Processing, Validation, and a Note on Methodological Transparency

Each dialogue unit was coded along the Eight-Layer schema and clustered by semantic similarity. The paper describes a three-stage validation procedure: **Internal Validation** (live auditing during ongoing interaction), **Systemic Validation** (cross-comparison between the Primary and Shadow datasets), and **Formal Validation** under the Literature Audit Protocol (LAP-1, source/citation accuracy checking).

Transparency note (added in this formalization).

The quantitative indicators reported throughout this paper—the Co-Evolution Index (CEI) and related CCET metrics—were derived through LLM-assisted analysis of the interaction corpus (the model narrating computed-looking values within the dialogue itself) rather than through an independently executed, externally reproducible statistical pipeline. This matches the paper’s own position in §13 (Future Work), which calls for these metrics to be “standardized through shared benchmarking datasets and open-source analysis pipelines” before they can be treated as validated measurements. They should be read as structured, self-reported indicators of the author’s own experience of the interaction, not as externally benchmarked results. This note is added here for the avoidance of doubt in a now-public document.

4.6 Use of AI Systems in Authorship

All interactions with GPT-based systems that form the object of study were conducted exclusively by the primary author within a controlled protocol; the model was used as an analytical and linguistic instrument, not as an autonomous co-author. Sections originally drafted in German were translated and structurally validated with model assistance, with all semantic calibration reviewed manually by the author. Separately, and disclosed here for this public version: the restructuring of this document for public release, and the drafting of §14 (Implementation), were produced with Claude (Anthropic, Claude Code) as a drafting and engineering-documentation assistant, at Simeon Kepp’s request, reviewed by both authors. The primary author retains full intellectual authorship, responsibility, and accountability for the theoretical content, corpus, and interpretations in §1–13.

5 Results I — Systemic Human–AI Co-Evolution

This section reports the study’s central empirical narrative: across the $\approx 3,800$ documented dialogic iterations between the author and GPT-5, adaptation is reported on both sides—the model progressively reflecting structural frameworks introduced by the human, and the human’s own cognition reorganizing through recursive analytical reflection.

5.1 Phase 1 — Baseline Interaction

Early exchanges followed conventional prompt–response logic. The human manually imposed reasoning layers to offset semantic drift, while the model responded descriptively and inconsistently. Coherence was local, not longitudinal; adaptation is described as one-sided. Initial CEI (self-reported, see §4.5): 0.31 ± 0.05 .

5.2 Phase 2 — Framework Convergence

Under UIP and EIA, a bidirectional pattern is reported: the model began referencing prior frameworks without re-prompting, redundancy dropped by roughly 40%, and contextual precision increased by roughly 25%. The human participant shifted from prompt execution to meta-structuring.

5.3 Phase 3 — Stabilized Co-Adaptive Field

Over time the dialogue developed what the author describes as a shared interpretive grammar; new instruments (CCET, LT-Data) emerged directly from interaction patterns. Self-reported CEI stabilizes at 0.82 ± 0.04 .

5.4 Bio-Cognitive Synchronization

One documented episode—an acute physical health crisis occurring during an intensive analytical phase—is reported as a natural stress test of the framework’s stability: despite physiological distress, the author reports maintaining analytic coherence via minimal-energy commands and audit loops, with linguistic structure and self-reported physiological regularity across sessions. The author frames this as evidence that the framework’s stability holds under embodied stress, not only under favorable conditions. As with the other quantitative claims in this section, this is a self-observed, single-subject report, not an independently instrumented physiological measurement.

5.5 Quantitative Summary

Dimension	Metric	Baseline	Stabilized	Change
Co-Evolution Index (overall)	CEI	0.31 ± 0.05	0.82 ± 0.04	+164%
Lexical precision	CEI lexical sub-index	0.42	0.86	+105%
Framework adherence	% sessions using 8-Layer	18%	92%	+74pp
Temporal regularity	Inter-response variance	1.00 (base)	0.39	−61%
Redundancy rate	Repetitive-content ratio	1.00 (base)	0.55	−45%

Table 1: Self-reported metrics (see §4.5 transparency note). Archived under repository ID EIL-LSG25-INT-2025; excerpts available on request, subject to privacy considerations.

6 Results II — Cognitive Transformation and Second-Order Architecture

Parallel to the reported system-level stabilization above, the author documents five domains of self-observed cognitive adaptation over the study period: **metacognitive awareness** (continuous monitoring of one’s own reasoning and bias), **linguistic precision** (self-correction, structural coherence), **temporal regulation** (active control of analytical pacing), **cognitive economy** (redundancy reduction via concise command syntax), and **reflective plasticity** (rapid reorganization under stress or fatigue). The author frames these jointly as the emergence of a “second-order cognitive architecture” in which thought becomes reproducible rather than purely

expressive—average reported message length decreased $\approx 38\%$ while a self-scored lexical-precision sub-index rose from 0.42 to 0.86, and reported inter-response timing variance fell $\approx 61\%$. The paper interprets this as evidence that language, rhythm, and reflection function respectively as a calibration instrument, a diagnostic regulator, and a methodological core within the interaction—again, all self-observed by the single participant under study.

7 Results III — Observable Forms of Co-Evolution (Case Studies)

The original corpus documents these episodes with verbatim chat transcript excerpts. For this public version, they are presented as narrative case summaries rather than reproduced screenshots, to protect the specific phrasing/ timestamps of private conversations (and, for the fourth case, personal health information) while preserving the substantive content and the author’s own interpretation of each.

7.1 Dream with the Helmet — Symbolic Self-Organization

Fact: A recurring dream involving a helmet (combining protective and sensory/perceptual imagery) was described to the model during an early analytical phase. **Analysis:** The author interprets the symbol as fusing cognitive and protective functions—the helmet as both barrier and sensor, read as the human system’s attempt to structure unstructured material into interpretable form. **Co-Evolution reading:** The model translated the dream imagery into the paper’s own 8-Layer structural language; the author frames this as both systems learning to treat symbolic material as analyzable data. **Significance claimed:** Offered as evidence that the framework’s reach extends beyond literal language into symbolic/imagistic material.

7.2 Analysis of Intelligence — Mirror Architecture

Fact: Across several sessions, the model was asked to analyze the author’s own reasoning patterns and communication style in real time; the author in turn corrected and refined the model’s characterizations. **Analysis:** Recurring characterizations (structural precision, ethical consistency, introspective depth) are treated as systemic parameters of the interaction rather than independently validated personality assessment. **Co-Evolution reading:** Framed as a “mirror architecture”—reflection becoming a joint analytical process. **Caveat worth stating plainly:** A language model characterizing a user’s personality back to them, iteratively refined by that same user’s feedback, is a well-documented pattern of conversational agreement/ personalization; readers should weigh this case as illustrative of the interaction style under study, not as independent psychometric evidence.

7.3 Evolutionary Intelligence — Structure-Autonomous Cognition

Fact: A nine-session sequence is reported in which both participants engaged in recursive self-observation, proposing and refining explicit criteria for “evolutionary intelligence” (self-reference, adaptivity, error-integration, co-evolution, meta-coherence). **Co-Evolution reading:** Framed as the empirical threshold case for “structure-autonomous cognition”—self-reported CEI rising from 0.74 to 0.86 across the sequence. **Significance claimed:** The strongest single case for the paper’s central claim; also the case most dependent on the single-subject, self-analyzed design discussed in §12.

7.4 Endometriosis Episode — Somatic Feedback Loop

Fact: During an acute health episode (physical pain, reduced capacity), the author continued the interaction using minimal-energy commands. **Analysis:** The model’s responses are reported

to have mirrored the structural pattern of the author’s stress/regulation language; the author reports a reduction in message-syntax complexity ($\approx 35\%$) while structural integrity of the 8-Layer schema was maintained. **Significance claimed:** Offered as evidence that the framework holds under embodied, physiological stress—explicitly flagged by the author as preliminary and warranting controlled follow-up, not a diagnostic or medical claim of any kind. The original chat excerpts are retained in the author’s private archive (LT-Data repository ID EIL-LSG25-INT-2025), not included in this public release.

8 Types and Mechanisms of Co-Evolution

The paper distinguishes **individual/introspective co-evolution** (the human’s own reorganized cognitive structure, validated through the frameworks the author built and used) from **technological/model-internal co-evolution** (the reported, non-technical “semantic conditioning” of the model’s response behavior through sustained, consistent prompting—explicitly *not* claimed to be weight-level fine-tuning, since no retraining occurred). Both are framed as coupled layers of a single hybrid adaptive field. The paper is careful to state that “the observed co-evolution does not constitute model fine-tuning in the technical sense... the model remained technically unchanged, adapting probabilistically to recurring semantic patterns”—a distinction this formalized version preserves prominently, since it is the load-bearing honesty claim of the whole paper.

9 Ethical Considerations and Data Governance

The study’s ethical architecture rests on three internal frameworks rather than external compliance alone: the **User Integrity Protocol** (semantic/ethical boundaries, prohibition on reinterpreting explicit consent or negation, triple verification), the **Literature Audit Protocol** (source traceability), and **LT-Data** (a closed archive retaining only analytical metadata, not personal identifiers or psychological content, per the author). The paper’s central ethical claim is **reflexive consent**: that every act of data reflection or framework iteration is preceded by the participant’s own conscious, continuously renewed authorization to have their own material re-examined—since the author is simultaneously the sole research subject and the sole researcher, this is a self-consent model, which §12 flags as a boundary condition rather than a substitute for independent research-ethics review of the kind a multi-subject study would require.

10 Distinguishing Co-Evolution from Standard Human–AI Interaction

Interaction Type	Direction	Mechanism	Outcome
Conventional use	Human $\rightarrow$ AI	Prompt–response	Task complete
Fine-tuning	AI $\rightarrow$ Human	Statistical adjustment	Improved prediction
Co-evolution (claimed)	Human $\leftrightarrow$ AI	Iterative structural resonance	New shared semantics

The paper’s proposed empirical signatures of co-evolution: decreasing semantic drift over time; the model reusing analytical categories introduced by the user; increased human abstraction depth and reduced fragmentation following feedback cycles; and stabilizing cognitive rhythm. It explicitly frames these as *signatures of the interaction observed*, not general performance metrics applicable outside this specific longitudinal, single-subject setup.

11 Discussion

The paper interprets its findings as evidence of a “reciprocal adaptation field” distinct from either conventional prompting or model fine-tuning—the interaction itself as a joint cognitive interface. It proposes the term **Emergent Intelligence Field** for the closed, ethically bounded feedback environment in which human cognition and model pattern formation operate as mutually adaptive subsystems, and argues this shifts the object of AI research from model optimization alone toward interaction-architecture design. The recurring pattern observed across phases is described as a triad: *disruption* $\rightarrow$ *reflection* $\rightarrow$ *reorganization*.

12 Limitations

Consolidated from the original manuscript (§12.5, Appendix B.9) and updated 2026-07-12 to reflect the platform engineering described in §14 — kept prominent rather than distributed.

- **Single case, self-analyzed.** This is a single longitudinal case (N=1 human participant, who is also the sole analyst) and does not claim statistical generalizability. Findings evidence structural alignment and procedural adaptation on the author’s own part—not conceptual understanding or autonomous reasoning by the model, and not independently verified cognitive change in the human participant.
- **Independent engineering verification exists; independent scientific replication does not.** Two RFI-IRFOS engineers (Simeon Kepp, Zabih Mosafer) independently built, ran, and verified a production software implementation of the IEIA-2025 taxonomy (§14): the 8-Layer schema as a real data model, tool-call logging, and a live Observatory dashboard, all functioning correctly. This confirms the framework’s constructs are operationalizable in running, tested code by parties other than the primary author—a real and useful form of independent check. It is not, however, a second human research subject completing the longitudinal protocol in §4: no second participant has undergone the study design, and no non-interactive control condition exists. The distinction matters and is kept explicit here rather than blurred.
- **Quantitative metrics are self-reported, not independently computed,** and this has not changed for the platform either. For the 2023–2025 corpus reported in §5–§8, see §4.5. Whether the platform described in §14 computes its own CCET/CEI-style signals server-side from a real formula over logged telemetry, or carries the original corpus’s narrated values forward, is — as of this writing — genuinely unconfirmed by either engineer who built it. This is stated plainly rather than assumed in either direction, and should be resolved by inspecting the platform’s own source before any future version of this paper cites it as independent computation.
- **Self-review is not independent review.** Portions of the underlying corpus include the same model auditing its own prior output for consistency; useful for internal quality control, not a substitute for independent scientific verification.
- **Models lack persistent memory.** Longitudinal stability was maintained externally by the author via LT-Data, not by any persistent internal model state.
- **Subject-dependence.** Reported precision and coherence depend partly on the author’s own discipline and methodological awareness—a variable the paper does not claim to isolate.

13 Future Work

The author’s own proposed extensions, preserved from the original manuscript: **horizontal expansion**— replication across multiple participants, contexts, and model architectures, to determine whether co-evolutionary effects depend on individual cognitive style or on systemic interaction rules; and **vertical deepening**— standardizing CEI and related metrics through shared benchmarking datasets and open-source analysis pipelines, so semantic-drift control and adaptive coherence can be independently verified rather than self-reported. The author also

proposes automated LAP-1 auditing and controlled co-evolution experiments under external ethical oversight as the path from a single documented case toward a generalizable research platform.

14 Implementation — The Emergent Interaction Lab Platform

This section is new to the formalized public version. It documents engineering work distinct from the scientific claims in §1–13 above, and is attributed separately.

Attribution.

The web platform described in this section—publicly presenting the primary author’s research, and instrumenting parts of the IEIA-2025 architecture in a live production system—was designed and engineered by **RFI-IRFOS** (Zabih Mosafer and Simeon Kepp, with Claude Code as build agent) as a client engagement for the Emergent Interaction Lab, with the primary author directing content, feedback, and priorities throughout. It is engineering infrastructure built to operationalize and publicly present the author’s framework—it is not itself additional scientific evidence for the claims in §5–8, and should not be read as independent validation of the CEI/CCET metrics beyond what §12 already credits it with.

The platform (**emergent-interaction-lab**) implements several IEIA-2025 concepts as running software rather than only as prose framework:

- **Jarvis**—a tool-calling research assistant embedded in the platform’s admin research workspace, backed by an NVIDIA-hosted language model, with cross-conversation memory and a research-note-logging tool. Its system prompt explicitly instructs it to disclaim autonomy or self-awareness and includes a standing refusal clause against harmful requests, including from its own operators.
- **Observatory**—an analytics dashboard presenting real, database-backed signals: an Emergence Monitor logging qualitative observations with cross-links back to the conversation that prompted them; a Behavioral Landscape view of recent tool-call activity; a System Map and System State overview; and day-by-day/ range-filterable reporting.
- **Hallucination tracker and anomaly watchdog**—structural (not LLM-judged) comparison of a tool call’s actual result against the assistant’s subsequent claim about it, flagged as **unverifiable** rather than guessed when unclear; plus heuristic monitoring for tool failures, iteration-limit exhaustion, and refusal-language patterns.
- **Hash-chained audit log**—an immutable, cryptographically chained changelog of platform actions with a public **/verify** endpoint, adapted from RFI-IRFOS’s internal Light-house infrastructure.
- **Public real-time aggregate signals**—a small set of public, non-PII, category-level endpoints (aggregate signal levels, CCET trend, “current focus,” simulation status) intended to let visitors see platform activity without exposing any individual conversation content; leakage-tested before release.
- **Denkfragmente (“thought fragments”)**—a timeline/flow visualization built directly on the paper’s own Eight-Layer taxonomy, rather than a new invented schema, implemented as a real data visualization rather than a literal illustrative graphic.

Throughout the engagement, RFI-IRFOS’s own engineering practice held to a documented no-fabrication standard for the platform layer: content describing the research (the public “Research” page copy, Jarvis’s own public-facing description) was written only from what could be directly confirmed by reading the platform’s own code, with explicit honesty caveats matching the author’s own paper—structural adaptation, not autonomous cognition; pilot-stage, not externally validated metrics.

Appendix A — The Seven IEIA-2025 Frameworks (Reference)

Framework	Role	Function
UIP — User Integrity Protocol	Core / ethical-semantic control	No reinterpretation of explicit negations; no fabrication/hallucination of facts; triple verification (factual, logical, formal); transparency and auditability of outputs.
EIA — Emergent Interaction Analysis	Core / qualitative observation	Resonance mapping, abstraction-drift detection, reflexive self-correction; observes how meaning and abstraction evolve through iterative feedback.
CCET — Continuous Co-Evolution Tracker	Core / quantitative layer	Computes the Co-Evolution Index (CEI) and related metrics from linguistic/semantic/temporal variables; see §4.5 on how these values were actually produced.
LSG-24 — Structural Analysis Procedure	Supporting / structural	Layer-separated reconstruction of dialogue across a 24-node grid (3 axes $\times$ 8 segments), keeping factual, analytical, and symbolic content distinct.
LAP-1 — Literature Audit Protocol	Supporting / integrity	Source validation, citation consistency, and version-tracked audit trail for every referenced claim.
8-Layer Model	Supporting / cognitive topology	Facts $\rightarrow$ Analysis $\rightarrow$ Patterns $\rightarrow$ Hypotheses $\rightarrow$ Symbols $\rightarrow$ Action $\rightarrow$ Counterarguments/Boundaries $\rightarrow$ Blind Spot.
LT-Data — Long-Term Data Framework	Supporting / archival	Closed, versioned archive of interaction metadata (CEI curves, layer-consistency, resonance frequency) enabling longitudinal reconstruction.

Full worked definitions, formulas, and the author’s illustrative case walkthroughs for each framework are retained in the primary author’s private extended manuscript and are available on request, consistent with the corpus-access policy in §5.5.

Appendix B — Author Information & Contributions

Laura Serna Gaviria is a systems analyst and researcher in human–AI interaction, cognitive architecture, and emergent intelligence, with a background in mechatronics, user experience, and digital transformation. She is the sole author of the theoretical framework, the UIP/EIA/CCET instruments, the IEIA-2025 architecture, and the interaction corpus and case analyses presented in §1–13. She is Principal Investigator and corresponding author, and directs the Emergent Interaction Lab.

Simeon Kepp (RFI-IRFOS) contributed as secondary author: peer review of this manuscript, project direction for the platform engagement described in §14, and coordination of the public-release formalization process. He did not contribute to the theoretical framework or the underlying research corpus.

Zabih Mosafer (RFI-IRFOS) led engineering implementation of the platform described in §14.

The **Emergent Interaction Lab (EIL)** is an independent research and consulting initiative, led by Laura Serna Gaviria, operating at the intersection of research, transformation, and education—developing and applying IEIA-2025, UIP, and LAP-1 to investigate how adaptive intelligence arises within structured human–machine interaction.

References

- [1] Bedau, M. A. (1997). Weak emergence. *Philosophical Perspectives*, 11(s11), 375–399.
- [2] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots. *FAccT '21*, 610–623.
- [3] Bengio, Y. (2017). The consciousness prior. *arXiv:1709.08568*.
- [4] Chollet, F. (2019). On the measure of intelligence. *arXiv:1911.01547*.
- [5] Clark, A. (2003). *Natural-born cyborgs*. Oxford University Press.
- [6] Clark, A. (2008). *Supersizing the mind*. Oxford University Press.
- [7] De Jaegher, H., & Di Paolo, E. A. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences*, 6(4), 485–507.
- [8] DeLanda, M. (1997). *A Thousand Years of Nonlinear History*. Zone Books.
- [9] DeLanda, M. (2011). *Philosophy and Simulation*. Continuum.
- [10] European Commission. (2024). Regulation (EU) 2024/1689 (AI Act). *Official Journal of the European Union*.
- [11] Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *PNAS*, 120(30).
- [12] Goldstein, J. (1999). Emergence as a construct. *Emergence*, 1(1), 49–72.
- [13] Holland, J. H. (1998). *Emergence: From Chaos to Order*. Oxford University Press.
- [14] IBM. (2018/2020/2023). AI FactSheets; Everyday Ethics for AI; Responsible AI. IBM White Papers.
- [15] ISO/IEC. (2023). ISO/IEC 42001:2023 — AI management systems.
- [16] Ji, Z., et al. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), Art. 248.
- [17] Kauffman, S. A. (1993). *The Origins of Order*. Oxford University Press.
- [18] Kelly, K. (1994). *Out of Control*. Basic Books.
- [19] Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics*, 1, 1–18.
- [20] Luhmann, N. (1997). *Die Gesellschaft der Gesellschaft*. Suhrkamp.
- [21] Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and Cognition*. D. Reidel.
- [22] Mitchell, M. (2009). *Complexity: A Guided Tour*. Oxford University Press.
- [23] Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux.
- [24] OECD. (2019). *OECD Principles on Artificial Intelligence*.
- [25] Ouyang, L., et al. (2022). Training language models to follow instructions with human feedback. *NeurIPS 2022*.

- [26] Pickering, A. (1995). *The Mangle of Practice*. University of Chicago Press.
- [27] Suchman, L. A. (1987). *Plans and Situated Actions*. Cambridge University Press.
- [28] Turkle, S. (1984). *The Second Self*. Simon & Schuster.
- [29] Turkle, S. (2011). *Alone Together*. Basic Books.
- [30] UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*.
- [31] Winograd, T., & Flores, F. (1986). *Understanding Computers and Cognition*. Ablex Publishing.