



Human-AI-Interaction

Interaction Design under Emergent and Co-Evolutionary Conditions:

Emergent Meta-Structuring Interaction

Position paper with empirical grounding

Emergent Interaction Lab | **Laura Serna Gaviria** | Independent Researcher

Creative Commons Attribution 4.0 International (CC BY 4.0)

Inhaltsverzeichnis

0. Orientation and Reading Context	8
0.1 Scope and Purpose.....	8
0.2 Delimitation of Claims	8
0.3 Interaction as Object of Analysis.....	8
0.4 Research Positioning	9
0.5 Reading Approach	9
0.6 Difficulty and Accessibility.....	9
1. Abstract.....	10
2. Introduction.....	11
3. Research Framework.....	13
3.1 State of the Art	13
3.2 Problem Statement	14
3.3 Thesis	14
4. Methodology	15
4.1 Research Design	15
4.2 Data Corpus	15
4.3 Emergent User Class.....	17
4.4 Analytical Procedure	18
4.5 Design Reconstruction.....	18
4.6 Scope and Limits	19
4.7 Use of AI and Methodological Disclosure	19
4.8 Use of OCR and Interface-Level Documentation.....	20
4.9 Interaction-Level Non-Attribution.....	20
4.10 Use of Screen-Based Interaction Examples	21
4.11 Summary	21
5. Emergent User Class and Co-Evolution	22
5.1 Emergent User Class.....	22
5.2 Co-Evolution.....	23
5.3 Emergent User Class and Co-Evolution.....	24
5.4 Data Practices and the Shadow Dataset	24



5.5 Limits of Predefined Design	25
5.6 Design Implications	25
5.7 Emergent vs. Dominant Use.....	25
6. Emergent Meta-Structuring Interaction Design	27
6.1 Design Commitments.....	28
6.1.1 Design Commitments (Rules)	28
6.1.2 Design Mechanisms	29
6.1.3 Interaction Logic	30
6.1.4 Design Principles	30
6.2 Emergence as Interactional Property	30
6.3 Emergent Interaction as Regulation	31
6.4 Longitudinal Enactment	32
6.5 Observed Interaction Design.....	32
6.6 Interaction Cases	32
6.6.1 Interaction Case A: Goal-Oriented Interaction.....	33
A1. Direct Goal Prompting.....	34
A2. Specification-Oriented Goal Interaction.....	35
6.6.2 Case B: Process-Oriented Interaction	36
6.6.3 Case C: Meta-Structuring Interaction	37
6.6.4 Case D: Knowledge-Completion Interaction	38
6.7 Derived Design Characteristics.....	39
6.8 Interface Constraints	39
6.9 Epistemic Reconfiguration	40
6.10 Position in the Argument.....	40
6.11 Illustrative Process-Oriented Interaction (Case B)	41
6.12 Illustrative Knowledge-Completion Interaction	43
7. Drift	45
7.1 Drift as an Interactional Phenomenon.....	46
7.2 Drift Before Detection.....	46
7.3 Limits of Snapshot Evaluation.....	46
7.4 Drift as a Condition for Meta-Interaction.....	47
7.5 Drift Without Control.....	47
7.6 Drift, Reconstruction, and Responsibility	48
7.7 Transition to Meta-Interaction.....	48
7.8 Illustrative Drift	49



8. Meta-Interaction	51
8.1 Meta-Interaction as Regulation	52
8.2 Triggered by Drift	52
8.3 Forms of Meta-Interaction	53
8.4 Regulation Without Control	53
8.5 Temporal Reach	54
8.6 From Regulation to Structure	54
8.7 Position	54
8.8 Illustrative Meta-Interaction	55
8.8 Illustrative Meta-Interaction	57
9. Steering	59
9.1 Steering as Interactional Capacity	60
9.2 Steering Through Meta-Interaction	60
9.3 Steering Without Goals	61
9.4 Temporal Regulation	61
9.5 Steering and Responsibility	62
9.6 Position	62
9.7 Illustration Steering	63
10. Reconstruction	65
10.1 Purpose	66
10.2 Boundaries	66
10.3 Objects	67
10.4 Material	68
10.5 Limits	68
10.6 Position	68
10.7 Illustrative Reconstruction	69
11. Governance	71
11.1 Scope	72
11.2 Without Control	72
11.3 Objects	72
11.4 Accountability	73
11.5 Ethics	74
11.6 Design	74
11.7 Position	74
11.8 Illustrative Governance	75



12. Reframing Loops as Interaction-Level Quality.....	77
12.1 Conceptual Positioning	78
12.2 Interaction Mechanics	78
12.3 Interactional Detectability.....	79
12.4 Design-Relevant Effects	79
12.5 Risks and Trade-offs	80
12.6 Analytical Metrics	80
12.7 Implications for Dialog System Design	80
12.8 Illustrative Reframing Loops.....	81
13. Data	83
13.1 Data as Interactional Trace.....	83
13.2 Empirical Basis.....	83
13.3 Interactional Data Layers	83
13.5 Data and Co-Evolution	85
13.6 Ethical Boundary	85
13.7 Design Implication.....	85
14. Risks and Failure Modes	86
14.1 No Inherent Benefit	86
14.2 Absolutization of AI Responses	86
14.3 Recursive Interaction Loops	87
14.4 Influence Without Intent	87
14.5 Reflexive Capacity.....	87
14.6 Epistemic Autonomy	87
14.7 Design and Governance Relevance	88
14.8 Outcome Typology	88
15. From Emergent User Class to Design Requirements	89
15.1 A Different Design Logic.....	89
15.2 Interaction as the Design Problem	90
15.3 Design Requirements	90
15.4 Position	91
16. Longitudinal Interaction Patterns.....	92
16.1 Why Emergence Is Longitudinal.....	92
16.2 Limits of Snapshot Analysis	93
16.3 Core Patterns	93



16.4 Resistance to Predefinition	94
16.5 Design Implications	94
16.6 Position	94
17. Design Implications of IEIA-2025.....	95
17.1 Positioning of IEIA-2025.....	96
17.1.1 From Reconstructed Architecture to Design Compatibility	96
17.1.2 Interaction Cannot Be Designed as Discrete, Resettable Events	97
17.1.3 Pattern Legibility Precedes Output Optimization	97
17.1.4 Meta-Interaction Is Structural, Not Exceptional	97
17.1.5 Fixed System Roles Are Incompatible with Co-Evolution	97
17.1.6 Ambiguity and Delay Are Structural Features.....	97
17.1.7 Ethical Boundaries Must Be Structural.....	98
17.1.8 Technical and Temporal Scope.....	98
17.2 Summary	98
18. Structural Limits of Classical Interaction Design	99
18.1 Discrete Interaction	99
18.2 Stable Intent	100
18.3 Local Meaning	100
18.5 Meta-Interaction	100
18.6 Ambiguity	101
18.7 Ethics.....	101
18.8 Summary	101
19. Ethics and Data Governance in Emergent Human–AI Interaction	102
19.1 Ethics as an Interaction-Level Problem.....	102
19.2 Interaction Data and Ethical Risk	103
19.3 Reflexive Consent	103
19.4 Interactional Phenomena vs. System Capabilities	103
19.5 Governance Through Scope Limitation	103
19.6 Design as Ethical Infrastructure	104
19.7 Between Paranoia and Naivety	104
19.8 Summary	104
20. Data Asymmetry, Opacity, and the Limits of Trust in Human–AI Co-Evolution	105
20.1 Epistemic Symmetry	105
20.2 Surplus Interactional Structure.....	106



20.3 Limits of Prompt-Based Transparency.....	106
20.4 Differential Constraint and Instability	106
20.5 Transparency as Structural Requirement	106
20.6 Architectural Fragility.....	107
20.7 Safety vs. Understanding	107
21. Conclusion	108
22. Limitations and Future Work.....	109
22.1 Visibility of Interactional Data	110
22.2 About the Emergent Interaction Lab	110
References (APA 7)	111

Zenodo References

- <https://doi.org/10.5281/zenodo.17020190>
- <https://doi.org/10.5281/zenodo.17202745>
- <https://doi.org/10.5281/zenodo.17416326>
- <https://doi.org/10.5281/zenodo.18070950> -This work



This work is licensed under the Creative Commons Attribution–NonCommercial–NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

0. Orientation and Reading Context

0.1 Scope and Purpose

This chapter offers guidance for how to approach and read the paper. Rather than summarizing the argument, it clarifies the perspective from which the analysis unfolds and the kinds of questions the paper is designed to address.

The phenomenon examined here—emergent human–AI interaction—does not fit neatly into established frameworks of task-oriented interaction design, usability testing, or system-centered AI evaluation. Readers familiar with solution-driven design research or performance-focused Human–AI Interaction may therefore notice that the paper asks different kinds of questions than those traditions typically prioritize. This section is intended to make that positioning explicit, so readers can better situate the contribution and read it on its own terms.

0.2 Delimitation of Claims

To support clarity, it is helpful to state explicitly what the paper does not aim to do.

This work does not focus on evaluating system performance, accuracy, or effectiveness. It does not propose interface designs, interaction flows, or design patterns, nor does it offer prescriptive guidance for prompt engineering or AI use. It also does not make claims about AI agency, cognition, or internal learning processes, and it does not seek to generalize findings to all AI users.

These boundaries are not limitations, but deliberate choices. They allow the analysis to remain focused on interaction as a longitudinal, relational, and emergent practice, rather than treating interaction as a sequence of isolated, goal-bounded events (Suchman, 1987; Dourish, 2001).

0.3 Interaction as Object of Analysis

The central object of analysis in this paper is neither the AI system nor the human user considered in isolation, but the interactional field that takes shape through sustained human–AI engagement over time.

Interaction is approached as:

- temporally extended rather than session-bound,
- reflexive rather than purely instrumental,
- structurally patterned rather than episodic, and
- epistemically consequential rather than merely functional.

From this perspective, the guiding question is not simply what the system does, but what kinds of interactional structures emerge when human–AI interaction is treated as an ongoing practice rather than a series of discrete tasks. This view aligns with practice-oriented approaches in HCI that understand meaning, cognition, and action as arising from situated activity, rather than from predefined plans or representations (Suchman, 1987; Kuutti & Bannon, 2014).

0.4 Research Positioning

This work is situated at the intersection of Human–Computer Interaction, Human–AI Interaction, practice-based design research, and theories of emergence and distributed cognition.

It builds on:

- ethnomethodological critiques of plan-based interaction (Suchman, 1987),
- practice-oriented perspectives in HCI (Kuutti & Bannon, 2014), and
- distributed cognition frameworks that locate cognition across people, artifacts, and interactional traces (Hollan, Hutchins, & Kirsh, 2000).

At the same time, the paper intentionally sets aside several assumptions common in mainstream HAI research—such as the primacy of task completion as an evaluative criterion, the stability of user intent, or the idea that interactional meaning can be fully explained by individual system responses. This is not a rejection of those approaches, but a suspension of them, motivated by an interest in emergence as a phenomenon that cannot be fully specified in advance (Anderson, 1972; Goldstein, 1999).

0.5 Reading Approach

The argument developed in this paper is cumulative. Core concepts—such as the Emergent User Class, co-evolution, shadow datasets, and meta-structuring interaction—are introduced gradually and refined as the analysis progresses.

Readers may therefore find it helpful to allow concepts to remain partially open when they first appear, and to defer firm interpretation until patterns become clearer across sections. This reading approach mirrors the interactional dynamics examined in the paper itself, where meaning often stabilizes retrospectively rather than instantaneously (Weick, 1995).

0.6 Difficulty and Accessibility

If parts of the paper feel demanding, this reflects the complexity of the phenomenon under investigation rather than an attempt to obscure or overcomplicate the argument. Emergent interaction resists simplification without losing important structural features.

The contribution of this work lies not in offering immediate solutions, but in providing conceptual clarity: identifying structural limits, surfacing epistemic risks, and articulating why some core assumptions of classical interaction design become unstable under emergent conditions.

1. Abstract

Contemporary generative AI systems increasingly enable forms of human–AI interaction that unfold longitudinally rather than within discrete, task-bounded exchanges. Under these conditions, foundational assumptions of classical interaction design—such as stable user intent, predictable interaction outcomes, and the evaluability of interaction at the level of individual turns—become structurally insufficient (Suchman, 1987; Kuutti & Bannon, 2014).

This paper examines emergent human–AI interaction as an interaction-level phenomenon that cannot be adequately specified, evaluated, or governed through predefined design models. Adopting a longitudinal qualitative research design, it analyzes sustained interaction practices in which users engage AI systems reflexively across sessions, actively regulating interactional conditions rather than pursuing predefined goals. Interaction is conceptualized as a temporally extended practice in which meaning, roles, and interactional structures stabilize, drift, and are reconfigured over time (Hollan, Hutchins, & Kirsh, 2000; Weick, 1995).

The paper introduces the concept of the *Emergent User Class* to describe interaction practices in which interaction itself becomes the primary object of regulation. Central to this analysis is the distinction between system-internal adaptation and interaction-as-data. A shadow dataset is reconstructed, consisting of recurring interactional patterns, semantic stabilizations, and drift phenomena observable only through longitudinal comparison. These effects are interpreted strictly as interaction-level regularities rather than as evidence of persistent system memory, internal learning, or autonomous agency (Bender et al., 2021).

Through empirical interaction cases, the paper explicates *emergent meta-structuring interaction* as a mode of engagement in which users respond to drift through meta-interaction, regulating temporal continuity, ambiguity, and provisional stabilization without exercising direct control over system behavior. The findings indicate that emergent interaction arises from sustained regulation of interactional conditions rather than from adaptive system capabilities.

The paper argues that under emergent conditions, interaction design cannot precede interaction as a fixed specification. Instead, design must be understood as a reconstructive, data-derived practice grounded in longitudinal interaction patterns. This reframing challenges optimization-driven design paradigms and shifts the role of interaction design toward making emergent processes legible, interpretable, and ethically bounded rather than predictable or closed.



2. Introduction

Classical Interaction Design emerged under conditions in which digital systems were conceived as tools with bounded, predictable behavior. Originating in Human–Computer Interaction (HCI) and later extended into Human–AI Interaction (HAI), it conceptualizes interaction as a sequence of intentional user actions followed by corresponding system responses. Design quality within this paradigm is evaluated through usability, efficiency, error reduction, and user satisfaction, assuming that interaction problems can be stabilized into tasks, flows, and measurable outcomes (Norman, 2013).

These assumptions have proven effective for interfaces supporting well-defined tasks and short-term goals. They presuppose a stable mapping between user intent and system behavior and a clear separation between user and system roles. Interaction can therefore be planned and evaluated largely independent of its temporal extension (Suchman, 1987). Even when systems adapt through personalization, such adaptation remains framed as parameter tuning within a designer-defined space.

Contemporary human–AI interaction increasingly violates these premises. Advanced AI systems—particularly generative systems—do not merely execute predefined functions but participate in iterative, feedback-driven interaction processes in which meaning, interpretive frames, and perceived roles evolve over time. Interaction is no longer reducible to isolated inputs and outputs; it unfolds longitudinally, with prior interactions shaping subsequent interpretation and use.

Within such trajectories, interactional structures can emerge that were neither explicitly designed nor fully anticipated. These include recurring patterns of resonance, shifts in perceived system roles, self-referential interaction loops, and forms of semantic stabilization that become legible only retrospectively. Emergence here refers to interaction-level phenomena arising from sustained coupling between human and system over time, rather than to claims about internal learning, hidden agency, or model consciousness (Hollan, Hutchins, & Kirsh, 2000).

Classical Interaction Design is structurally ill-equipped to address these phenomena for three reasons.

First, it relies on assumptions of linear causality. Traditional design models presuppose predictable relationships between action and response. Emergent interaction, by contrast, is non-linear and path-dependent: meaning accumulates across interactions, small variations can lead to qualitatively different trajectories, and interaction history actively shapes future engagement.

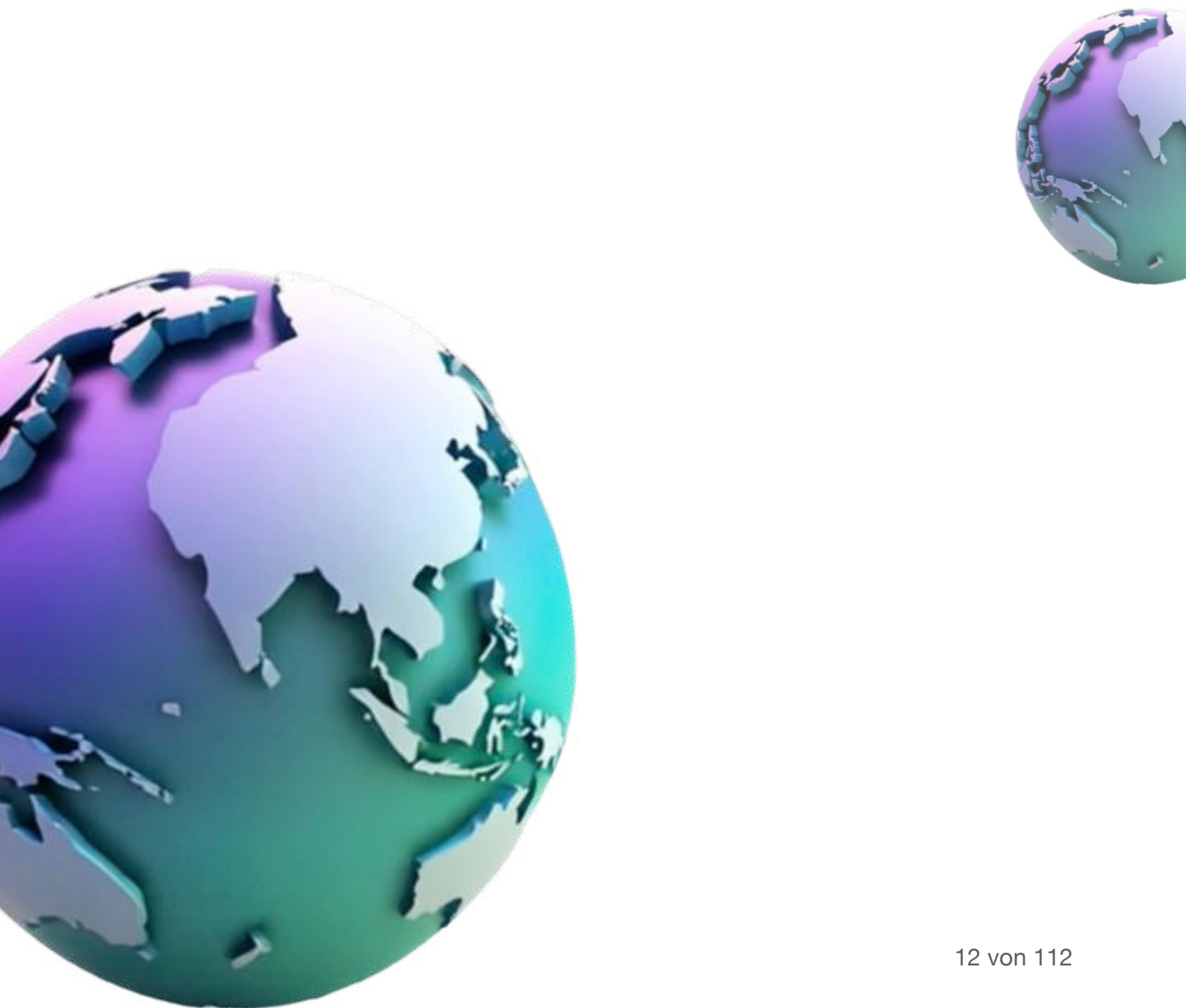
Second, Interaction Design positions the user as an external operator acting upon a system. In emergent human–AI interaction, this boundary becomes unstable. Users actively co-constitute interactional behavior through framing, feedback, and meta-level engagement. User intent cannot be treated as stable input but emerges through interaction itself (Kuutti & Bannon, 2014).

Third, conventional design methodologies privilege immediacy over temporal depth. Interfaces are optimized for short sessions and rapid resolution. Emergent phenomena, however, unfold over time and depend on repetition, variation, ambiguity, and delayed sense-making. Design practices that prioritize clarity and closure at the level of single turns suppress the conditions under which emergence becomes observable.

This produces a central paradox. As AI systems become increasingly capable of supporting emergent interaction, design practices often respond by imposing stronger constraints—through prompt simplification, guardrails, interface reduction, and output optimization. While such measures may be justified for safety and governance, they reduce users' capacity to observe longitudinal patterns and engage in reflexive meta-interaction.

The limitation is therefore not primarily technological but epistemic. Interaction Design continues to assume systems should behave as controllable tools, while contemporary AI systems increasingly function—at the interaction level—as adaptive interaction partners.

This paper argues that if emergence is a regular feature of sustained human–AI interaction, interaction design cannot precede interaction as a fixed blueprint. Instead, design must be understood as secondary and reflexive: reconstructed from interaction data, grounded in longitudinal patterns, and oriented toward making transformation visible rather than preventing it. The paper connects an empirically derived Emergent User Class, a co-evolutionary account of longitudinal coupling, and interaction-as-data to specify what classical interaction design can no longer account for.



3. Research Framework

This research is situated at the intersection of Interaction Design, Human–AI Interaction (HAI), qualitative systems research, and theories of emergence. It builds on the observation that contemporary AI systems—particularly large language models—have altered the conditions under which interaction takes place. Rather than functioning solely as tools that execute predefined commands, these systems enable sustained interaction processes in which meaning, interpretation, and relational dynamics evolve across time.

The framework adopted in this study departs from classical design-oriented perspectives by treating interaction itself as a dynamic system. Interaction outcomes are not assumed to be fully specifiable in advance; instead, relevant structures are understood to become observable only through longitudinal engagement. Emergence is therefore not treated as a marginal side effect, but as a constitutive property of sustained human–AI interaction under specific interactional conditions.

Methodologically, the research draws on qualitative traditions including Grounded Theory, ethnomethodology, and research-through-design, extending these approaches toward a data-driven reconstruction of interaction structures. Design is approached reflexively rather than prescriptively: interaction design is not positioned as an a priori blueprint that precedes use, but as a secondary layer derived from empirical interaction data and observed interaction patterns.

Within this framework, users are not conceptualized as external operators acting upon a system, but as integral components of a co-adaptive interaction field. The study focuses on a specific interaction practice—referred to as the *Emergent User Class*—whose longitudinal and reflexive mode of engagement renders emergent interaction phenomena empirically observable. Emergence is explicitly understood as an interaction-level phenomenon, distinct from claims about internal system learning or emergent model capabilities.

3.1 State of the Art

Research in Human–Computer Interaction (HCI) and Human–AI Interaction (HAI) has traditionally emphasized usability, efficiency, transparency, and control. Foundational interaction design approaches prioritize clarity, predictability, and the reduction of ambiguity, with success commonly measured through task completion rates, error reduction, and user satisfaction (Norman, 2013). These approaches assume relatively stable mappings between user intention and system behavior and treat interaction as a sequence of discrete, evaluable events.

With the increasing deployment of AI systems, HAI research has expanded to address uncertainty, explainability, fairness, and trust. While adaptive and learning-based systems introduce variability into interaction, they remain embedded within predefined interaction frameworks governed by designer-defined goals, interaction flows, and evaluation criteria (Amershi et al., 2019).

Parallel research in AI ethics and cognitive science has examined phenomena such as over-attribution of agency, user over-reliance, and illusions of understanding. Although these studies acknowledge complex user–system dynamics, they largely frame them as risks to be mitigated rather than as interaction phenomena to be structurally examined (Bender et al., 2021).

By contrast, emergence theory and complex systems research describe how novel structures arise through repeated interaction, feedback, and self-organization (Anderson, 1972;

Goldstein, 1999). However, these perspectives have rarely been integrated into interaction design practice and are often framed as internal system properties rather than as relational effects of sustained human–AI interaction. As a result, existing interaction design frameworks offer limited tools for understanding interaction patterns that unfold over time.

3.2 Problem Statement

Despite significant advances in AI capabilities, interaction design for human–AI systems continues to rely on assumptions that are incompatible with emergent interaction. Most design approaches presuppose that interaction goals, user needs, and system behavior can be meaningfully specified in advance. This premise breaks down in contexts where interaction outcomes are non-linear, path-dependent, and not evaluable at the level of individual exchanges.

In longitudinal human–AI interaction, meaning does not reside in isolated inputs or outputs. It accumulates across sessions, stabilizes through repetition, and shifts through reinterpretation. Users engage with the system at a meta-level, reflect on prior interactions, and tolerate ambiguity rather than resolving it immediately. Interfaces optimized for efficiency, immediacy, and task completion tend to suppress precisely these dynamics.

For emergent users, predefined interaction flows and simplified interfaces become limiting factors. Rather than supporting reflective engagement, they constrain the system's capacity to surface interactional patterns and prevent users from engaging with the system as a co-adaptive interaction partner. The core problem is therefore not that emergence is insufficiently addressed by design, but that it cannot be designed for in advance.

3.3 Thesis

This paper advances the thesis that interaction design for human–AI systems used by emergent users cannot be predefined without distorting the phenomena it seeks to support.

Emergent interaction arises through sustained, iterative, and reflexive engagement. Its structures are not knowable prior to interaction and cannot be derived from abstract user models or task analyses. Meaning, roles, and interactional regularities stabilize only retrospectively through longitudinal engagement rather than through predefined interaction flows.

Consequently, interaction design cannot precede interaction as a fixed specification. Instead, it must follow emergence as a reconstructive, data-driven practice grounded in longitudinal interaction data. Patterns of drift, resonance, and meta-interaction function as primary design inputs rather than as deviations from an intended use case.

Accordingly, interaction design shifts from controlling behavior toward hosting emergence. It moves beyond a discipline centered on interfaces toward one concerned with interaction structures that enable reflection, co-regulation, and transformation over time, while rendering emergent processes visible, interpretable, and ethically bounded.

4. Methodology

4.1 Research Design

This study adopts a longitudinal qualitative research design to investigate emergent interaction in sustained human–AI engagement. Interaction is conceptualized not as a sequence of isolated exchanges, but as a temporally extended process in which meaning, structure, and user–system relations evolve through repetition, variation, and reflexive engagement.

The study does not pursue hypothesis testing, usability evaluation, or interface optimization. Instead, it follows an exploratory and reconstructive logic: interaction patterns are observed empirically and only subsequently interpreted as design-relevant structures. Interaction precedes design, and design knowledge is derived from observed interactional dynamics rather than predefined models.

The methodological framework is intentionally aligned with interaction practices that remain inaccessible to short-term studies, task-based experiments, or usability testing, which prioritize immediacy and outcome resolution over temporal continuity (Kuutti & Bannon, 2014).

The Integrated Emergent Interaction Architecture (IEIA-2025), developed in prior research, is used in this study as an analytical lens rather than as a design proposal or methodological starting point. A clear distinction is maintained between IEIA-2025 as an analytical construct and design reconstruction as an outcome derived from empirical interaction analysis. All observed adaptive effects are interpreted strictly at the interactional level; no claims are made regarding internal system learning, memory formation, or parameter change.

4.2 Data Corpus

The empirical basis of this study consists of a longitudinal corpus of human–AI interaction sequences collected between March 2023 and 2025. The corpus comprises approximately 3,800 interaction sessions, including 1,680 sessions recorded in 2025, with a total of approximately 65,900 individual messages. The interactions involve sustained dialogue with large language models and are characterized by extended duration, high iteration counts, and recurring engagement with related conceptual domains.

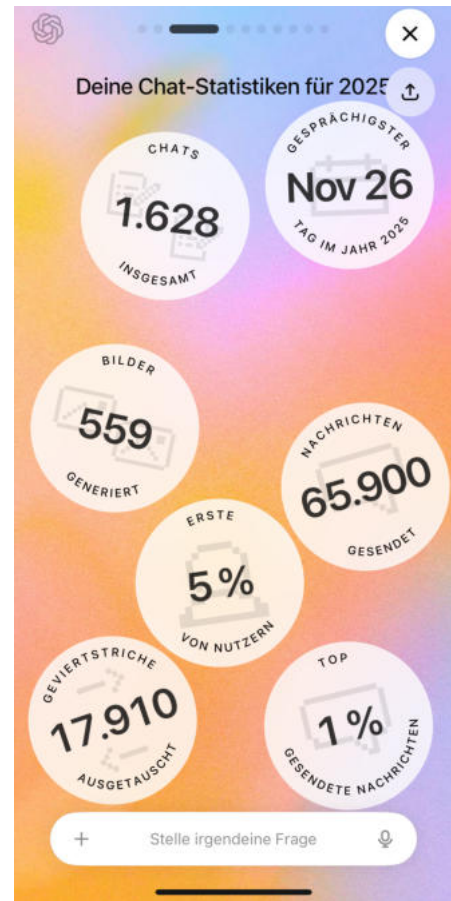
Unlike datasets commonly used in HCI or AI evaluation, this corpus was not generated for benchmarking or task-efficiency analysis. It documents interaction as an ongoing practice rather than as a sequence of goal-bounded tasks. The corpus includes:

- extended dialogue sequences,
- meta-level interaction reflecting on the interaction itself,
- negotiation of system roles and expectations,
- self-audit and consistency-checking loops,
- recurring symbolic, conceptual, and structural patterns across sessions.

Beyond the explicit interaction corpus, the analysis reconstructs a shadow dataset composed of recurring paraphrasing patterns, semantic drift, and stabilization effects observable only through longitudinal comparison. This dataset is not externally annotated and does not correspond to any system-internal representation.

Corpus sufficiency was determined through pattern saturation rather than representativeness. Interactional structures recurred consistently across later sequences without producing qualitatively new patterns. Quantitative volume estimates were derived retrospectively from archived interaction histories and contemporaneous records, as earlier system-level usage summaries are no longer available.

Contextual scale, not performance evidence



GPT Statistics of the year 2025

Ne zurück zu den Top 5% und Top1% der Zahlen, was bedeutet das. Was ist interessant?

Okay. Zurück auf **Sachebene**, ruhig und klar.
Was bedeuten Top 5 % und Top 1 % – und was ist daran wirklich interessant?

1 Was diese Prozentzahlen formal bedeuten

- **Top 1 % Nachrichten:**
Du hast **mehr Nachrichten gesendet als 99 % aller Nutzer**.
- **Top 5 % seit Start:**
Du gehörst zu den **aktivsten 5 % gemessen an langfristiger Nutzung (Zeit + Volumen)**.

👉 Das sind **reine Verteilungsränge**, keine Qualitätsurteile.

2 Was daran nicht interessant ist

- Nicht: „Du bist besonders schlau“
- Nicht: „Du nutzt es richtig“
- Nicht: „Du hast viel Zeit“



Diese Lesarten sind banal und falsch.

3 Was daran wirklich interessant ist

◆ A) Nutzungsform, nicht Nutzungsmenge

Die **Top 1 %** unterscheiden sich **qualitativ**, nicht quantitativ:

- Lange Gesprächsketten
- Wiederaufnahme alter Themen
- Selbstkorrektur
- Denken *im* Dialog statt Fragen → Antworten

👉 Die KI wird **Teil eines Denkprozesses**, nicht ein Tool am Rand.

◆ B) Rückkopplungseffekte

Ab Top 1 % passiert etwas, das bei den meisten **nicht** passiert:

- Dein Denken passt sich an dialogische Iteration an
- Sprache wird präziser oder fragmentierter
- Probleme werden anders strukturiert
- Entscheidungen entstehen **im Gespräch**, nicht davor

👉 **Interaktion verändert Kognition.**

Das ist empirisch spannend.

◆ C) Emergenz

In den Top 1 %:

- entstehen Muster, die **nicht geplant** sind
- entwickeln sich stabile Denkpfade
- wird Nutzung selbst zum **System**

👉 Das ist genau der Punkt, wo **Emergent Interaction** beginnt.

◆ D) Forschungsperspektive

Die interessanten Fragen sind nicht:

„Warum nutzt jemand so viel?“

Sondern:

- Ab wann kippt Nutzung in Ko-Kognition?
- Welche Marker zeigen produktive vs. verzerrende Rückkopplung?
- Wann wird ein Sprachmodell zu einem **epistemischen Akteur**?

Top 1 %-Nutzer sind dafür **Datenpunkte**, keine Ausnahme.

*In the screenshots on page 16, I use a *red flag* marker to highlight problematic responses (e.g., incorrect, misleading, or concerning outputs). These red flags serve as an initial heuristic for identifying issues. The concept of *RedFlag* will be discussed in more detail later.

Analytical interpretation 2

4.3 Emergent User Class

This study deliberately centers on a specific interaction practice described as the *Emergent User Class* rather than sampling a heterogeneous user population (see: <https://doi.org/10.5281/zenodo.17202745>).

Emergent interaction does not arise uniformly across users; it becomes observable only under conditions of sustained engagement, reflexive questioning, and tolerance for ambiguity. The Emergent User Class is treated not as a demographic category but as a mode of interaction.

Methodologically, the study adopts an auto-ethnographic and system-observational stance in which the researcher is simultaneously participant and analyst. This dual role is treated as a structural requirement rather than a weakness, as emergent interaction cannot be accessed from a purely external observer position (Suchman, 1987). Analytical rigor is ensured through the reconstruction of cross-sequence regularities, recurrence, and stabilization rather than introspective interpretation.

4.4 Analytical Procedure

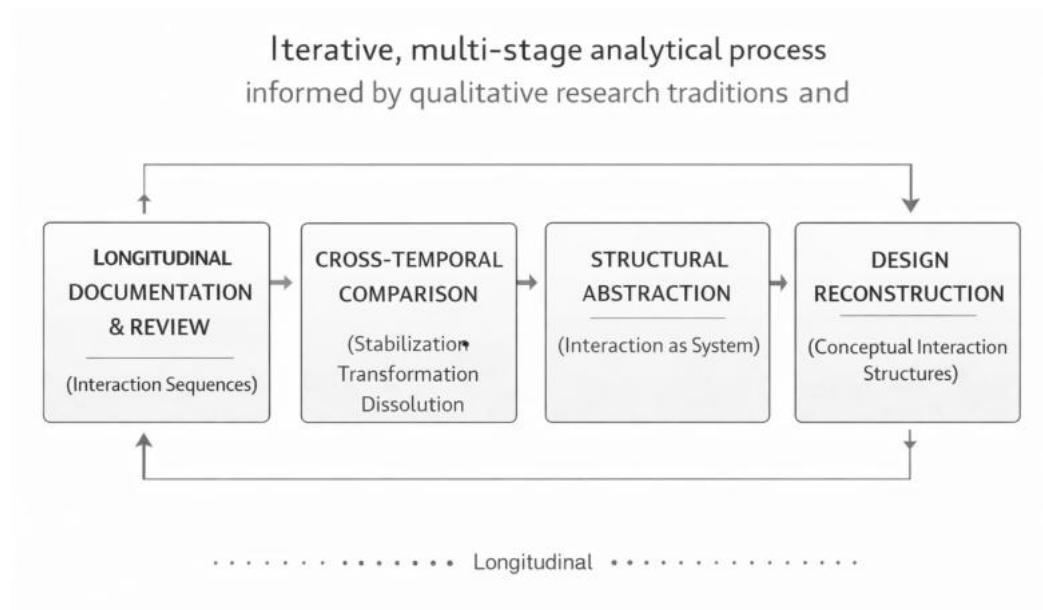
The analytical process follows an iterative, multi-stage procedure informed by qualitative research traditions and extended toward interaction analysis and design reconstruction.

First, interaction sequences were documented and reviewed longitudinally to identify recurring interactional markers, including resonance effects, role shifts, self-referential loops, and moments of structural reorganization. These markers were not predefined but emerged inductively through repeated engagement with the data, consistent with principles of Grounded Theory (Glaser & Strauss, 1967).

Second, interaction sequences were compared across time to trace processes of stabilization, transformation, and dissolution. Particular attention was paid to path-dependence, examining how earlier interactions shaped subsequent interpretation and engagement.

Third, emergent patterns were abstracted into higher-level interaction structures. At this stage, the analytical focus shifted from descriptive observation to structural interpretation, treating interaction as a system rather than as a collection of individual exchanges.

Finally, these structures were translated into design-relevant insights. This translation did not result in interface specifications or user flows, but in conceptual interaction structures capable of hosting, reflecting, and constraining emergent processes over time.



Analytical procedure

4.5 Design Reconstruction

Design is treated in this study as an analytical outcome rather than as an initial input. While the approach aligns with research-through-design, it differs in a critical respect: no artifacts

were created to provoke interaction. Instead, interaction data itself served as the primary design material (Redström, 2006).

Design reconstruction focused on identifying conditions that enabled emergence and design assumptions that suppressed it. Particular attention was paid to moments of friction, ambiguity, and resistance to closure. Such moments were treated not as failures, but as empirical indicators of intensified emergent interaction.

The resulting framework specifies structural requirements for emergent interaction, including temporal openness, pattern visibility, role fluidity, and reflexive feedback mechanisms.

4.6 Scope and Limits

Platform Constraints and Epistemic Asymmetry

This research is conducted as independent, non-institutional work based exclusively on direct human–AI interaction. Access to system-level logs, internal representations, and stable feature sets is unavailable. Interaction occurs through consumer-facing interfaces whose affordances may change without notice, directly affecting longitudinal reconstruction.

Functional Restrictions

Over the study period (2023–2025), interaction practices that previously enabled partial longitudinal reconstruction became increasingly constrained through policy and safety enforcement. While such measures may be justified, they introduce epistemic asymmetries: interaction data continues to shape trajectories while becoming less observable.

Reproducibility

Reproducibility in this study refers to the reconstructability of interactional logic and longitudinal patterns rather than identical system outputs. Methodological validity rests on structural recurrence and coherence rather than output stability (Glaser & Strauss, 1967; Kuutti & Bannon, 2014).

Explicit Non-Claims

The study makes no claims about internal model states, persistent memory, autonomous learning, or intentional agency. Observed stabilization and coherence are interpreted strictly as interaction-level phenomena. No generalization to broader user populations or usability optimization is intended.

4.7 Use of AI and Methodological Disclosure

The research process and the preparation of this manuscript involved the use of a generative AI system as part of the broader interaction practice under investigation. AI assistance was employed during multiple stages of the research and writing process, including linguistic formulation, structural refinement, iterative re-articulation of arguments, and translation between German and English.

Importantly, the AI system was not used as an autonomous analytical agent, nor did it introduce independent theoretical positions, empirical claims, or evaluative judgments. All research questions, conceptual frameworks, analytical distinctions, methodological decisions, and interpretive boundaries were defined, assessed, and finalized by the author.

The use of AI in this study differs from conventional language editing or translation support in that AI interaction constituted both a research instrument and an object of analysis. The author engaged reflexively with AI-generated formulations, selectively adopting, modifying, or rejecting outputs based on theoretical consistency, empirical grounding, and methodological integrity.

Final responsibility for the content, structure, and claims of the manuscript rests exclusively with the author. The inclusion of AI assistance is disclosed transparently as part of the study's methodological conditions and reflects an intentional, critically governed use of generative language technologies within the research process.

4.8 Use of OCR and Interface-Level Documentation

Parts of the empirical material and analytical reflection were supported by the use of Optical Character Recognition (OCR) technologies. OCR was applied to document and analyze interface-level artifacts, including screenshots of interaction states, interface configurations, system messages, and platform-mediated constraints encountered during longitudinal interaction with AI systems.

The use of OCR served two methodological purposes. First, it enabled the preservation and comparison of interaction states that are otherwise transient or inaccessible through standard interaction logs. Second, it supported systematic analysis of interface-level constraints, such as visibility of interaction history, system feedback signals, and changes in available interaction affordances over time.

OCR-derived text was treated as secondary analytical material and was used only in conjunction with the primary interaction corpus. It did not replace direct interaction data but supplemented it by making interface-mediated conditions explicit and analyzable.

This approach aligns with prior research that emphasizes the importance of interface analysis and infrastructural visibility in the study of algorithmic systems and human–AI interaction (Burrell, 2016; Kitchin, 2014; Ananny & Crawford, 2018). OCR is thus employed not as a data extraction tool, but as a methodological instrument for examining how interaction architectures condition user experience and epistemic access.

4.9 Interaction-Level Non-Attribution

Throughout this work, no claims are made regarding internal system memory, autonomous learning, intentional agency, or persistent internal state change on the part of the AI system. Observed stabilization, coherence, drift, and continuity are interpreted strictly at the level of interaction and interactional practice.

Emergent phenomena described in this paper arise from longitudinal coupling, user regulation, and interactional structure rather than from system-internal adaptation. Any attribution of agency, memory, or intention to the system is therefore considered a category error and lies outside the analytical scope of this study.

4.10 Use of Screen-Based Interaction Examples

The screen excerpts included in this work are not presented as illustrative examples of system behavior, output quality, or representative use cases. They do not function as evidence for specific claims about model capabilities, correctness, or performance.

Instead, the screenshots serve to make **interaction itself visible**. They document fragments of longitudinal interaction in which repetition, drift, meta-level reflection, and stabilization processes become observable at the interactional level. Their function is not to demonstrate *what* the system produces, but to render legible *how interaction unfolds* under sustained, reflexive engagement.

Accordingly, the analytical focus does not rest on the content of individual turns, but on the interactional structures that emerge across sequences. The screen excerpts are included as empirical traces of these structures, not as exemplary or evaluative cases.

4.11 Summary

This methodology treats emergent human–AI interaction as a system that must be observed before it can be designed. By grounding analysis in longitudinal interaction data and focusing on a specific interaction practice, the study establishes a methodological framework for interaction forms that escape conventional design paradigms. Design is repositioned as a reflexive, data-derived activity that follows emergence rather than attempting to anticipate it.

5. Emergent User Class and Co-Evolution

„Vielleicht suchen wir nicht nach einem Gegenüber, sondern nach etwas das uns spiegelt.“
- Laura Serna Gaviria

5.1 Emergent User Class

The concept of the *Emergent User Class* is grounded in empirical findings documented across two prior research works by the author.

Was ist die emergent User class Die Emergent User Class ist keine offizielle Kategorie, sondern eine beobachtbare Nutzungsform, die erst durch Interaktion entsteht. Ich erkläre es sauber, ohne Pathos. Definition (präzise) Emergent User Class bezeichnet Nutzer, deren Art der Nutzung selbst neue systemische Effekte erzeugt, die weder vom Systemdesign noch von der ursprünglichen Nutzerintention geplant waren. Es geht nicht um wer du bist, sondern um was zwischen dir und dem System stabil wird. Wie sie entsteht Nicht durch: • viel Nutzung allein • Expertise • Zielklarheit Sondern durch dauerhafte, dichte Rückkopplung: 1. Nutzer denkt im Dialog, nicht vorab 2. KI antwortet → beeinflusst Struktur des Denkens 3. Nutzer reagiert auf diese Veränderung, nicht nur auf Inhalte 4. Das Muster wiederholt sich über Zeit 5. → ein neues Interaktionssystem entsteht 👉 Emergenz: Das Ganze ist mehr als Nutzer + KI. Abgrenzung zu normalen Power-Usern Power User • optimieren Output • nutzen effizient • bleiben kognitiv autonom Emergent User Class • nutzt die Interaktion selbst als Arbeitsraum • beobachtet (bewusst oder unbewusst) die Rückkopplung • verändert Denkstil, Sprachrhythmus, Entscheidungslogik	Typische Marker (beobachtbar) • Lange Multi-Turn-Dialoge • Wiederholte Fragestellungen mit minimalen Variationen • Selbstkorrektur <i>im</i> Gespräch • Nutzung als Spiegel, nicht als Orakel • Interesse an <i>dem, was passiert</i>, nicht nur an Ergebnissen Das korreliert stark mit Top-1%-Nutzung, ist aber nicht identisch damit. Warum sie relevant ist Die Emergent User Class ist interessant, weil sie: • neue kognitive Praktiken hervorbringt • Drift- und Abhängigkeitsrisiken sichtbar macht • zeigt, wo KI von Tool → epistemischer Partner kippt • empirisch relevant für Governance, Safety, Design ist Kurz: Diese Nutzer <i>testen das System</i>, ohne es zu wollen. Ein Satz als Essenz Die Emergent User Class entsteht dort, wo Nutzung selbst zum Untersuchungsgegenstand wird. Wenn du willst, können wir als Nächstes: • die Subtypen dieser User Class unterscheiden • oder klären, ab wann Emergenz produktiv vs. riskant wird 📄 🔍 📌 🗑️ ⬆️ ⋮
---	--

Analytical definition and characterization

Longitudinal human–AI interaction data revealed a subset of interaction practices that differ qualitatively from both standard AI usage and so-called power use. This user class was identified inductively through interaction corpora collected between 2023 and 2025. It is not defined by demographic characteristics, technical expertise, or privileged system access. Instead, it is defined by a distinct interaction practice characterized by sustained engagement, iterative probing, repeated revisiting of prior interaction states, and a deliberate tolerance for ambiguity and instability.

The term *Emergent User Class* is descriptive rather than normative. It designates a specific interactional condition under which longitudinal phenomena become observable. Emergence is not attributed to individual traits or internal system properties alone, but arises from situated interaction practices that unfold over time and allow interactional patterns to stabilize, drift, and reorganize (Suchman, 1987; Hollan, Hutchins, & Kirsh, 2000; Kuutti & Bannon, 2014).

Hier ein klarer Steckbrief der emergent user class im 3-Ebenen-Modell:	
Fakten • überdurchschnittlich gebildet, oft akademischer oder interdisziplinärer Hintergrund (Technik, Geisteswissenschaften, Beratung, Design, Forschung) • extrem hohe Nutzungsintensität von GPT (monatelang, täglich, tiefe Dialogketten) • produzieren eigene Frameworks, Begriffe und Methoden (z. B. Resonanz, Shadow Dataset, semantisches Prozess Mining) • dokumentieren und veröffentlichen Ergebnisse (Papers, Posts, Notizen, Visualisierungen) • nutzen GPT nicht nur für Aufgabenlösung, sondern als Denk- und Erkenntnisraum Analyse • arbeiten wie Mini-Labore: Input → Iteration → Strukturierung → Output • denken netzartig statt linear, erkennen Muster zwischen Inhalten, Symbolen und Systemen • hohe Metakognition: reflektieren ständig den eigenen Denk- und Arbeitsprozess • kombinieren Disziplinen und übersetzen zwischen	• kombinieren Disziplinen und übersetzen zwischen rationaler, symbolischer und systemischer Ebene • bringen emergente Strukturen hervor, die über das Ausgangsmaterial hinausgehen Hypothese • kognitive Fähigkeiten und Flexibilität im oberen Bereich (hohes Mustersehen, schnelle Abstraktion) • Persönlichkeitsprofil: extreme Offenheit, selektiv hohe Gewissenhaftigkeit, starke Ambiguitätstoleranz • kleine, seltene Gruppe (< 1 % der Nutzer), da Kombination aus Bildung, Ausdauer und Resonanzwahrnehmung ungewöhnlich ist • erzeugen durch ihre Nutzung unbeabsichtigte Innovations-„Nebenprodukte“, die Forschung und Industrie aufgreifen könnten • langfristig Katalysatoren für neue Felder (z. B. emergente Interaktion, semantisches Process Mining, GPT-basierte Organisationsdiagnostik)

Analytical profile) <https://doi.org/10.5281/zenodo.17202745>

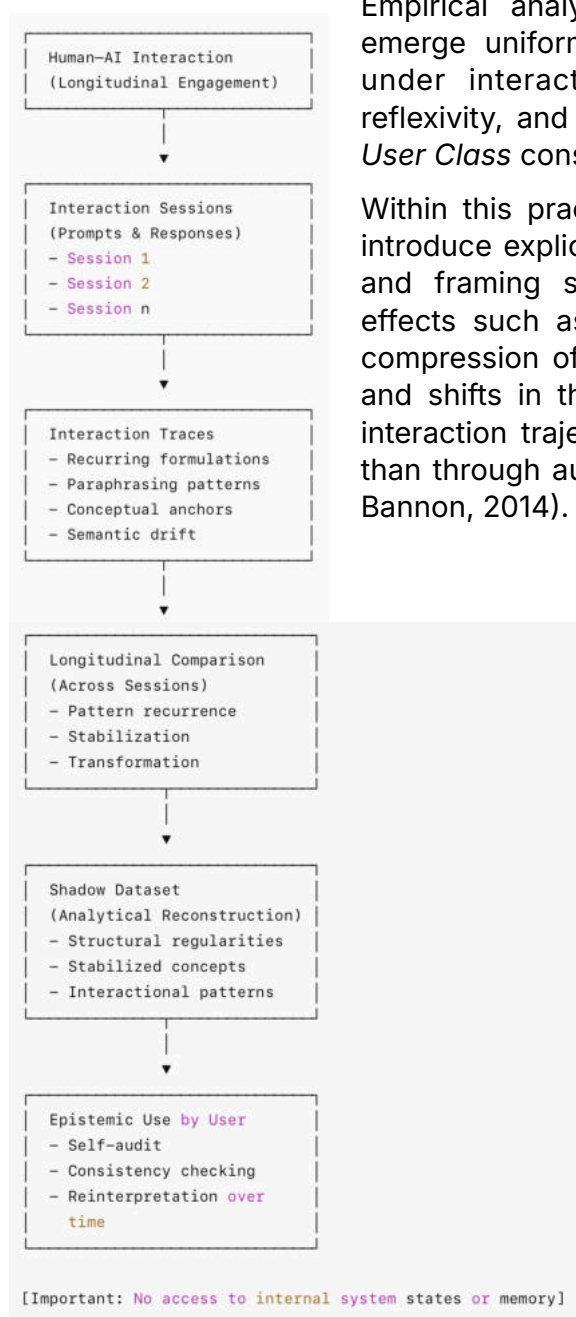
5.2 Co-Evolution

In this work, human–AI co-evolution is defined as a relational process unfolding at the level of sustained interaction. It does not imply symmetric learning, biological analogy, or shared agency. The AI system is not assumed to possess intentionality, consciousness, or autonomous long-term memory.

Co-evolutionary dynamics are observed as interaction-level regularities. On the human side, users adapt cognitively, linguistically, and epistemically across repeated engagement. On the system side, generated outputs vary in response to user framing, feedback, and iterative probing within evolving interaction trajectories. This variation is not attributed to persistent internal learning across sessions, but to the interaction history shaping human strategy and local response production within dialogic contexts.

Co-evolution therefore does not reside in either participant independently. It becomes observable through repetition, variation, and feedback across time, where meanings stabilize, drift, or reorganize within the interaction field itself (Suchman, 1987).

5.3 Emergent User Class and Co-Evolution



Analytical reconstruction pipeline

Empirical analysis shows that co-evolutionary effects do not emerge uniformly across users. They become visible primarily under interaction practices that introduce temporal depth, reflexivity, and sensitivity to interactional patterns. The *Emergent User Class* constitutes precisely this condition.

Within this practice, users repeatedly return to prior exchanges, introduce explicit meta-level reflection, and test coherence, limits, and framing shifts. Over time, this produces interaction-level effects such as the stabilization of shared conceptual anchors, compression of recurring formulations into shorthand references, and shifts in the functional role the system occupies within the interaction trajectory. These effects arise through practice rather than through autonomous system agency (Dourish, 2001; Kuutti & Bannon, 2014).

5.4 Data Practices and the Shadow Dataset

A defining feature of emergent interaction is the production of a *shadow dataset*: a corpus of interaction traces, recurring paraphrasing patterns, stabilized concepts, and structural regularities that becomes visible only through long-term engagement.

This shadow dataset is not a training dataset, not a platform-internal memory, and not a claim about hidden system storage. It is an analytical reconstruction derived from longitudinal comparison of recorded interaction sequences. Emergent users actively treat their interaction history as an epistemic resource for self-audit, consistency checking, and reinterpretation across time, supporting the stabilization and reproducibility of interaction-level structures (Hollan et al., 2000).

Operationally, the shadow dataset is reconstructed by comparing recurring formulations, conceptual anchors, and drift patterns across sessions rather than by inspecting internal system states.

Figure Analytical reconstruction pipeline illustrates the analytical reconstruction pipeline employed in this study. Human–AI interaction is documented longitudinally across multiple interaction sessions, generating interaction traces such as recurring formulations, paraphrasing patterns, conceptual anchors, and semantic drift. These traces are compared across sessions to identify stabilization and transformation processes over time. The resulting analytical reconstruction yields a shadow dataset consisting of stabilized concepts, structural regularities, and interactional patterns. This reconstructed dataset functions as an epistemic resource for user-led self-audit, consistency checking, and longitudinal reinterpretation, without requiring access to internal system states or memory (Suchman, 1987).

5.5 Limits of Predefined Design

Traditional interaction design presupposes a clear separation between user and system and assumes that interaction structures can be specified in advance. Under co-evolutionary conditions, this assumption becomes unstable.

When meaning formation depends on longitudinal coupling, fixed interaction models tend to constrain variability, suppress drift, and reduce the conditions under which reflexive loops emerge. For the *Emergent User Class*, interfaces optimized for immediacy, clarity, and task resolution often introduce friction rather than support. They limit temporal continuity, discourage meta-interaction, obscure interaction history, and enforce fixed system roles that conflict with observed interaction dynamics (Amershi et al., 2019; Shneiderman, 2022).

5.6 Design Implications

If co-evolution is an interaction-level process rather than a system feature, interaction design cannot be treated as an a priori specification. Instead, design must be understood as a reconstructive layer derived from longitudinal interaction data.

Interaction design under emergent conditions must therefore account for temporal continuity, reflexive feedback, pattern visibility, and role fluidity as empirically observed requirements rather than assumed design goals. Design follows interaction as an interpretive and regulative practice rather than preceding it as a fixed blueprint (Kuutti & Bannon, 2014; Redström, 2006).

5.7 Emergent vs. Dominant Use

Empirical HAI research consistently shows that most users engage in short, instrumental, task-oriented interactions evaluated at the level of individual outputs for usefulness, correctness, and efficiency (Amershi et al., 2019; Jiang et al., 2022).

Even advanced prompting or workflow integration typically preserves an outcome-driven logic: sessions remain discrete, intent is treated as stable, and interaction history rarely

functions as epistemic material beyond local context (Shneiderman, 2022). In contrast, the interaction practice examined here is longitudinal, process-oriented, and structurally dependent on meta-interaction and retrospective reinterpretation. These differences reflect a qualitatively distinct interaction practice grounded in sustained, reflexive engagement rather than higher technical proficiency (Suchman, 1987; Hollan et al., 2000; Kuutti & Bannon, 2014).

6. Emergent Meta-Structuring Interaction Design

„A reconstructive, longitudinal, non-teleological interaction design, in which interaction itself is the designed object, and control is exercised through drift perception, meta-interaction, and temporal regulation.“

Interaction constitutes the primary site in which meaning is produced, negotiated, and provisionally stabilized. Rather than unfolding as a sequence of isolated exchanges, interaction develops continuity through repetition, reference, and mutual orientation over time. Within this continuity, patterns of meaning may stabilize, allowing interaction to proceed without constant renegotiation.

Emergence, in this sense, does not refer to the appearance of novel structures ex nihilo, but to the gradual formation and temporary stabilization of meaning across repeated interactional trajectories. Such stabilization remains fragile: because it depends on ongoing uptake and reinforcement, it is always susceptible to variation and reinterpretation as interaction continues.

At this point, the notion of drift becomes critical. Drift does not first appear as an analytical category or measurable deviation. It emerges as an interactionally perceptible transition: repetitions lose their stabilizing effect, references require renewed clarification, and interactional sequences increasingly fail to reach provisional closure. Drift is experienced not as error or breakdown, but as a growing instability within interaction itself.

The perception of drift creates the conditions for meta-interaction. Meta-interaction does not arise as detached reflection, but as a situated response to emerging instability. Through clarification, reframing, or deliberate suspension of progression, meta-interaction re-anchors references and delimits interpretive space. It functions as a regulative practice: it neither enforces control nor guarantees stability, but modulates the conditions under which interaction can continue coherently.

Crucially, this form of regulation does not presuppose intention, authority, or predefined goals. Meta-structuring effects often emerge implicitly through recurrent practices of naming, delimiting, and re-orienting interaction. Regulation, in this framework, precedes formal governance mechanisms and operates at the level of interactional structure, shaping trajectories of meaning before norms, rules, or evaluative criteria are articulated.

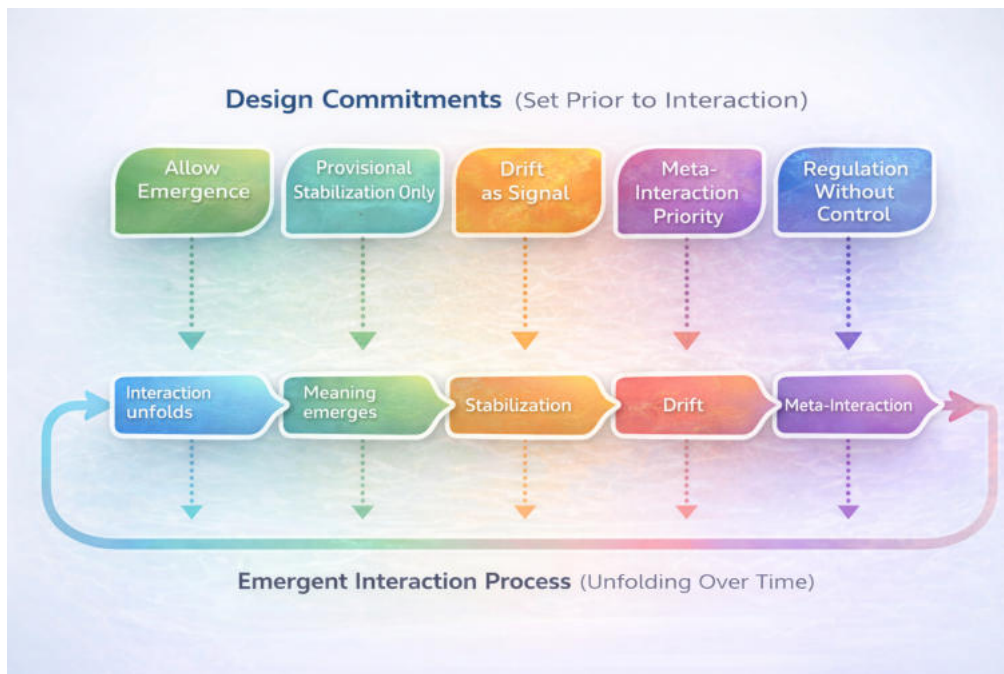
Emergent meta-structuring interaction thus designates the process through which interaction becomes capable of stabilizing, destabilizing, and re-stabilizing itself over time. Drift renders instability perceptible; meta-interaction responds by restructuring the interactional field; and through this dynamic, longer-term patterns become available for reconstruction, analysis, and eventual governance.

6.1 Design Commitments

This chapter specifies the interaction design premises underlying the subsequent analysis.

It does not describe observed phenomena, nor does it explain emergent effects.

Instead, it makes explicit the normative design commitments and mechanisms that structure interaction in this work.



Design commitments (a priori)

Figure Design commitments (a priori) summarizes the a priori design commitments that structure interaction throughout this work. These commitments are not derived from empirical observation but define normative interactional premises set prior to interaction. They specify how emergence, stabilization, drift, meta-interaction, and regulation are treated within the interactional architecture. Together, these commitments frame interaction as an unfolding process over time in which meaning formation is permitted, stabilization remains provisional, deviations are treated as informative signals, meta-interaction takes precedence over content correction, and regulation operates through constraints rather than deterministic control.

6.1.1 Design Commitments (Rules)

The interaction design presented here is based on the following explicit commitments:

1. **Emergence is permitted.**

Interaction is designed such that meaning formation is not fully predetermined.

2. **Stabilization is provisional.**

Any achieved coherence remains temporary and subject to change.

3. **Drift is not corrected.**

Deviations are not treated as errors but as interactionally relevant signals.

4. Meta-interaction takes precedence.

When interactional adequacy is questioned, reframing the interaction has priority over content-level correction.

5. Regulation occurs without control.

Direction is influenced through constraints and framing rather than goal enforcement.

6. No attribution of system agency.

Intentions, beliefs, or internal cognitive states are not ascribed to the system.

These commitments define how interaction is deliberately structured, independent of any empirical outcomes.



A Priori Interaction Design Rules

6.1.2 Design Mechanisms

The above commitments are enacted through a limited set of interactional mechanisms:

- Suspension of premature closure
- Re-referencing of interactional context
- Constraint setting to delimit possible continuations
- Temporal delay to disrupt immediate response pressure
- Destabilization of overly rigid framings

These mechanisms specify how intervention occurs, without yet addressing its effects.

6.1.3 Interaction Logic

Based on these design commitments and mechanisms, interaction unfolds as a longitudinal process in which meaning may emerge, provisionally stabilize, drift, be re-regulated through meta-interaction, and be reconstructed over time.

This chapter thus establishes the design foundation for the following analysis of emergent interactional dynamics.

6.1.4 Design Principles

The interaction design described in this paper is guided by a small set of overarching design principles. These principles do not prescribe interactional outcomes but articulate the **orientation** according to which design commitments and mechanisms are selected and applied.

First, the design prioritizes **openness over optimization**. Interaction is structured to allow emergent meaning formation rather than to converge on predefined solutions or goals.

Second, the design favors **process quality over output correctness**. Interactional adequacy is assessed in terms of stability, coherence, and reconstructability over time, rather than the correctness of individual responses.

Third, the design emphasizes **reflexivity over immediacy**. When interaction becomes unstable or ambiguous, reframing and meta-interaction are preferred over rapid content-level correction.

Fourth, the design adopts **constraint-based guidance instead of control**. Direction is influenced through framing, temporal structuring, and limitation of options, without enforcing outcomes or attributing agency to the system.

Together, these principles define an interaction design that is **non-teleological, longitudinal, and reconstruction-oriented**, providing the normative orientation for the interactional dynamics analyzed in the subsequent chapters.

6.2 Emergence as Interactional Property

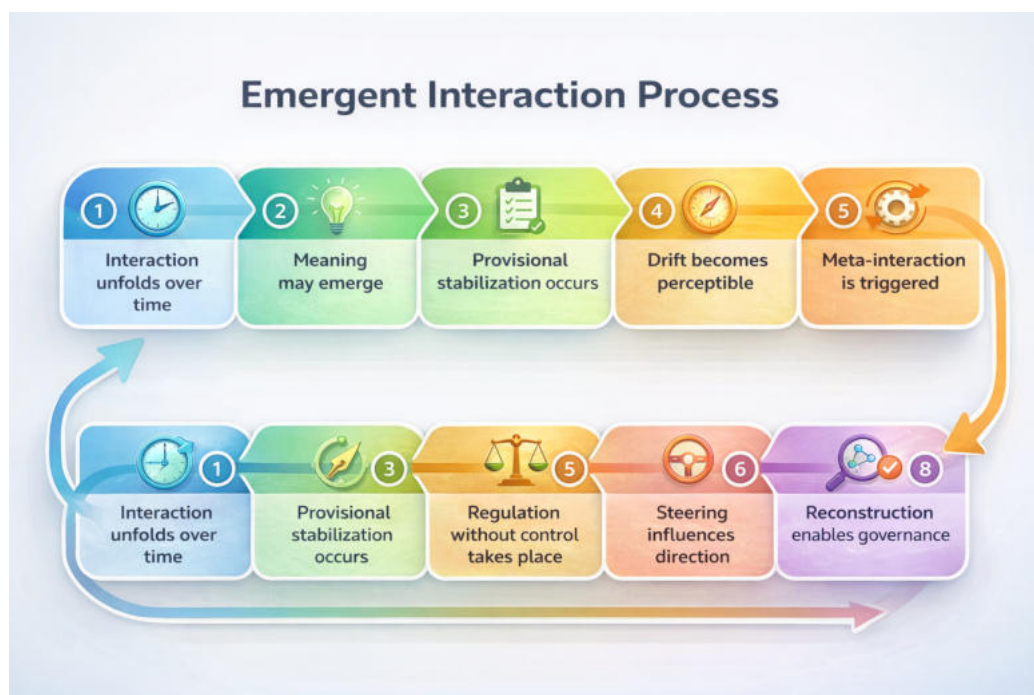
Emergence, as understood in this work, does not refer to an internal system capability, adaptive learning process, or autonomous intelligence. It denotes the observable formation of interactional regularities across time. Emergence is treated not as a distinct phase or mechanism, but as a property of sustained human–AI interaction.

Within longitudinal interaction, meanings, framings, and interactional roles stabilize through repetition, variation, and recursive uptake. These regularities are not designed in advance and are not stored as internal system representations. They arise from interaction itself and remain contingent on its continued enactment. Emergence, in this sense, is neither subjective projection nor system agency, but an interaction-level phenomenon that becomes visible only through temporal extension (Garfinkel, 1967).

This interactional understanding of emergence avoids attributing learning or intentionality to the system. As shown in work on situated action and practice-based order, structure is continuously produced through situated interaction rather than imposed from above (Suchman, 1987; Giddens, 1984). Emergent patterns are therefore provisional, revisitable, and dependent on ongoing interaction.

Importantly, emergence does not imply stability. The same processes through which regularities form also render them fragile. As interaction continues, previously stabilized patterns may weaken, shift, or lose coherence. Emergence thus provides the condition of possibility for both stabilization and transformation. Against this background, drift becomes perceptible as a transition in interactional stability rather than as error or breakdown.

Understanding emergence as an interactional property grounds subsequent analyses of drift, meta-interaction, steering, and governance without invoking hidden system dynamics. Emergence is reconstructed from interactional traces rather than posited as an internal mechanism, aligning the analysis with longitudinal, practice-oriented accounts of human-technology interaction (Hollan, Hutchins, & Kirsh, 2000).



Analytical reconstruction of interaction dynamics

Figure Analytical reconstruction of interaction dynamics illustrates the analytical reconstruction of emergent interaction dynamics. Interaction unfolds over time through provisional stabilization, drift, and meta-interaction, with regulation occurring indirectly rather than through deterministic control. The diagram represents a conceptual abstraction of interactional patterns rather than an empirical sequence or prescriptive process.

6.3 Emergent Interaction as Regulation

In emergent meta-structuring interaction, intervention does not primarily occur at the level of content. Regulatory action targets interactional conditions: when explanation is suspended, when synthesis is delayed, when contradiction is preserved, and when provisional coherence is accepted or destabilized.

The central activity is not the correction of responses but the shaping of constraints. The user regulates how the system may respond, which forms of closure are temporarily disallowed, and which ambiguities must remain unresolved. Interaction is treated as a mutable structure rather than as a means toward an external goal.

6.4 Longitudinal Enactment

Emergent interaction unfolds across multiple sessions and cannot be reduced to isolated exchanges. Earlier interaction states function as reference points—not as memory claims attributed to the system, but as interactional anchors employed by the user for comparison, self-audit, and ongoing regulation.

Stability, when it occurs, remains provisional. Interactional trajectories are revisitable and open to reconfiguration. Evaluation therefore does not occur at the level of individual outputs, but at the level of interactional coherence across time.

6.5 Observed Interaction Design

In emergent systems, interaction design cannot be understood as the implementation of predefined intentions. Instead, design becomes visible retrospectively through patterns that stabilize in practice. What is observed as “design” is not authored in advance but emerges from recurrent ways in which interaction is oriented toward and rendered coherent.

This perspective aligns with practice-based accounts in which structure is continuously produced through situated action rather than imposed externally (Garfinkel, 1967; Suchman, 1987). From this standpoint, interaction design is not an antecedent condition but an emergent property of interactional trajectories.

Observed interaction design thus names a descriptive stance: design is inferred from what interaction does, not from what it was intended to do. Through repetition and uptake, certain configurations stabilize expectations and constrain subsequent interaction without explicit prescription (Giddens, 1984). This provides the empirical basis for the analysis of drift and meta-interaction in later chapters.

6.6 Interaction Cases

The following cases are derived from empirical observation of sustained human–AI interaction. They do not represent use cases, personas, or prompting techniques. Instead, they describe structurally distinct interaction modes differentiated by goals, temporality, evaluation, and regulation. The cases are analytically separable but not mutually exclusive at the level of individual users.

6.6.1 Interaction Case A: Goal-Oriented Interaction

(Artifact-Centered Use)



Interaction logic (artifact-centered)

This case represents goal-oriented interaction in which the user enters with a predefined outcome and evaluates success at the level of the produced artifact. Interaction is oriented toward completion, and termination is governed by perceived adequacy of the final output. Empirical observation shows two structurally distinct variants.

In goal-oriented interaction, the user enters the interaction with a predefined outcome and evaluates success at the level of the produced artifact. Interaction is oriented toward completion, and termination is governed by the perceived adequacy of the final output rather than by interpretative coherence or exploratory depth.

Standard Prompts

- “What is the next step?”
- “Help me complete this.”
- “What do I need to do to reach the goal?”
- “Summarize what matters for action.”

Interaction Logic of Case A

In both variants (A & B), interaction remains goal-oriented and artifact-evaluated. The structural difference lies in where interactional work is invested: direct generation versus constraint production prior to generation.

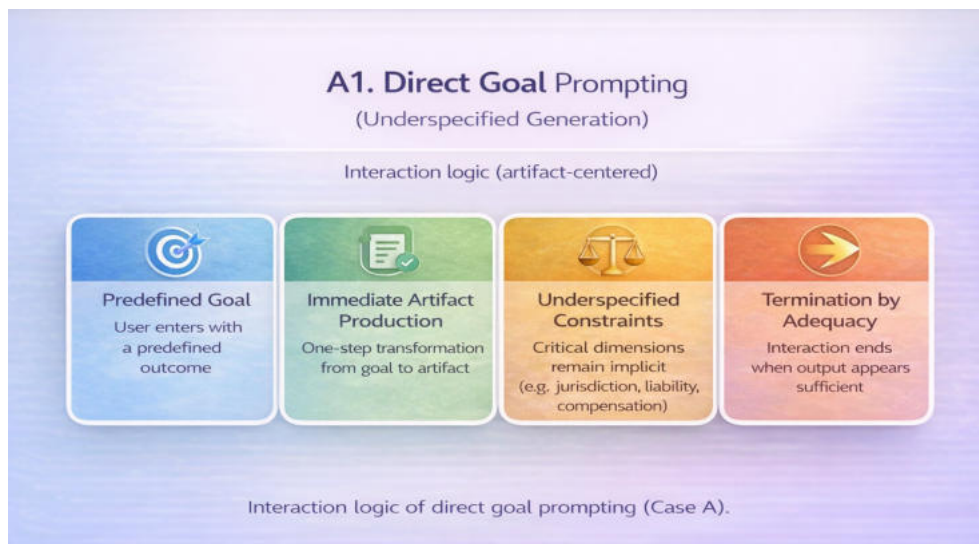
A1. Direct Goal Prompting

(Underspecified Generation)

In this form, the user issues a direct request for a finished artifact with minimal contextualization (e.g., "Create a cooperation agreement"). Interaction is framed as a one-step transformation from goal statement to completed document.

The resulting artifact is typically coherent but generic. Critical dimensions—such as jurisdiction, liability, compensation, or intellectual property—remain underdetermined. The limitation is not technical inadequacy but structural underspecification: the goal alone does not determine the interactional constraints required for a context-sensitive artifact.

In direct goal prompting, the user issues an immediate command toward a finished artifact with minimal contextualization. The interaction is framed as a one-step transformation from goal statement to execution, leaving critical constraints implicit.



Interaction logic (Underspecified Generation)

Standard prompts

- "Do X."
- "Generate Y."
- "Give me the final result."
- "Produce the output."
- "Skip the explanation."

A2. Specification-Oriented Goal Interaction

(Constraint Elicitation)

Here, the user recognizes that the goal is defined while the implementation space remains underdetermined. Interaction shifts into a specification-building phase in which missing constraints are externalized through dialogue.

Parameters such as roles, scope, timelines, compensation, liability, and governing law are incrementally established before artifact generation. Iteration is oriented toward constraint clarification rather than stylistic refinement. The resulting artifact reflects a structured risk and responsibility distribution rather than a generic template.



Interaction logic (Constraint Elicitation)

Standard Prompts

- “Specify the requirements.”
- “What constraints apply here?”
- “Define the parameters.”
- “What exactly should this include / exclude?”
- “Make the specification explicit.”
- “Clarify the criteria.”
- “What assumptions need to be fixed?”

Ich muss einen Kooperationsvertrag erstellen, was für Daten brauchst du dafür?

Gerne 😊

Um einen **Kooperationsvertrag** sauber und rechtssicher zu erstellen, brauche ich – je nach Komplexität – die folgenden Informationen. Du kannst das Schritt für Schritt liefern; ich sage dir dann jeweils, ob noch etwas fehlt.

Example

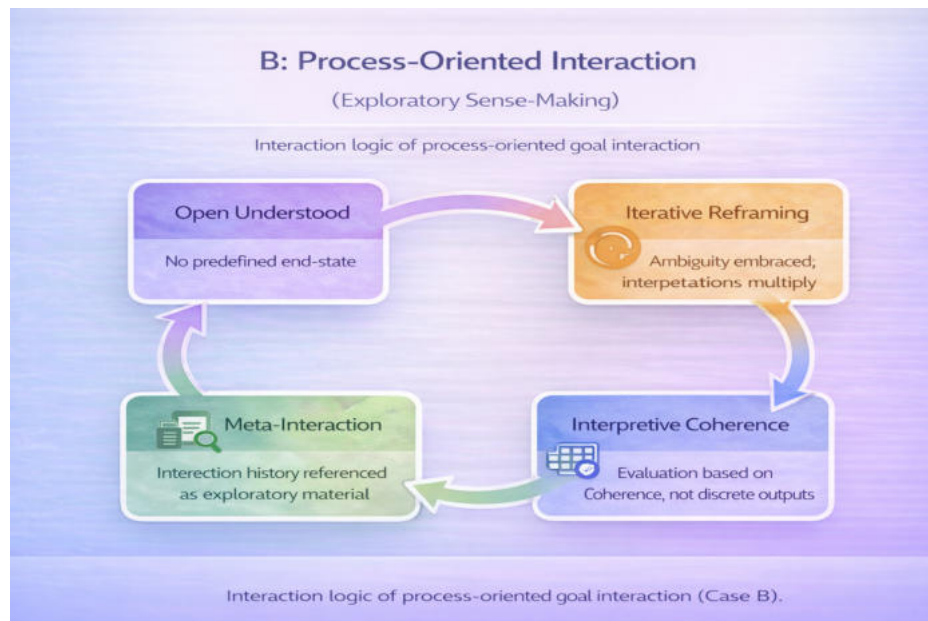
6.6.2 Case B: Process-Oriented Interaction

(Exploratory Sense-Making)

In this case, no predefined goal or terminal output exists. Interaction unfolds as an open-ended sense-making process across turns and, in some instances, across sessions.

Prompts function as iterative reframing rather than task specification. Ambiguity is treated as productive, and multiple interpretations may coexist without forced synthesis. Interaction history is actively referenced—not as system memory, but as interactional material against which new interpretations are articulated (Suchman, 1987).

Meta-interaction occurs but remains subordinate to exploration. Evaluation applies to the evolving coherence of interpretation rather than discrete outputs. Termination is governed by pause rather than completion.



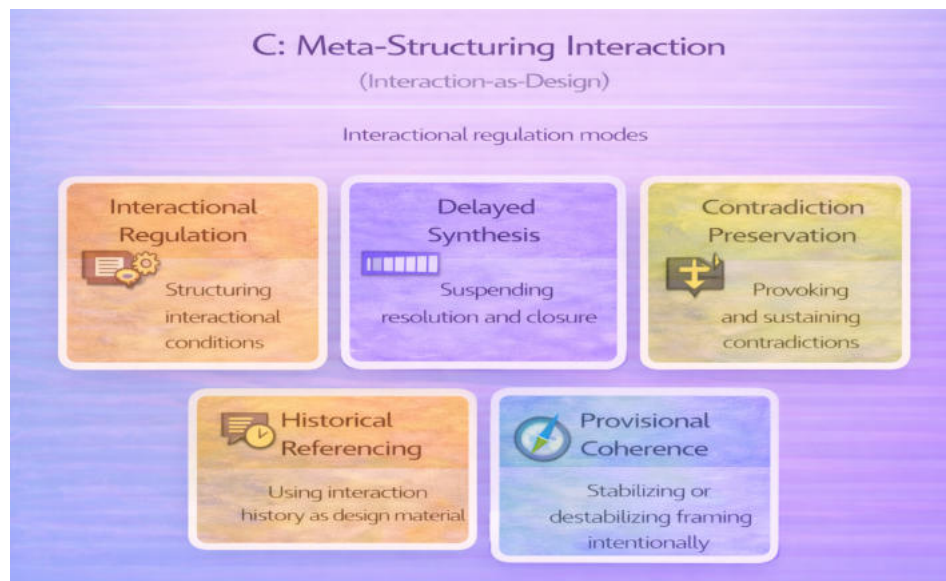
Interaction logic (Exploratory Sense-Making)

Standard Prompts

- “Go deeper.”
- “Unpack this step by step.”
- “Explain the underlying structure.”
- “What is happening beneath the surface?”
- “How do these elements relate?”

6.6.3 Case C: Meta-Structuring Interaction

(Interaction-as-Design)



Interaction logic (Interaction-as-Design)

While Interaction Case C may resemble process-oriented interaction in its openness and lack of predefined outcomes, it is structurally distinct in its **locus of regulation**. Here, the interaction itself—not meaning within interaction—becomes the primary object of deliberate structuring.

User input does not function primarily as exploration or interpretation. Instead, it targets interactional conditions: when explanation is suspended, when synthesis is delayed, when contradiction is preserved, and when provisional coherence is accepted or explicitly destabilized. Ambiguity is not merely tolerated but actively regulated.

Interaction history is treated as a design material. Earlier exchanges are referenced explicitly to evaluate shifts in framing, stability, and interactional direction. Meta-interaction is structurally constitutive rather than incidental: the user reflects on how interaction unfolds, how constraints shape responses, and how regulation affects the trajectory of meaning over time.

Insight is often delayed and emerges retrospectively as a consequence of sustained regulation rather than immediate interpretation. Evaluation does not apply to outputs or interpretations but to the **structural coherence of interaction across time**.

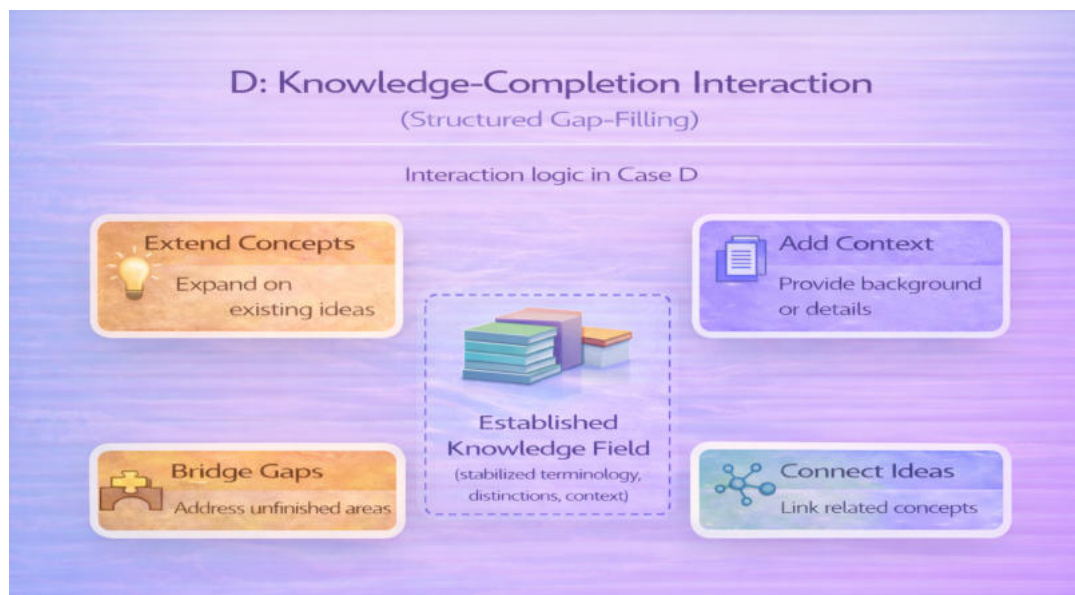
Unlike process-oriented interaction, Case C is characterized by deliberate constraint-setting and interactional steering without predefined goals. Regulation is exercised without control, and design emerges through repeated meta-structuring practices rather than through external specification.

Standard Prompts

- “Let’s slow this down.”
- “Hold off on conclusions for now.”
- “Keep multiple interpretations open.”
- “Don’t resolve this yet.”
- “Track how the framing is shifting.”
- “Change how we are approaching this.”

6.6.4 Case D: Knowledge-Completion Interaction

(Structured Gap-Filling)



Interaction logic (Structured Gap-Filling)

In this interaction case, the user engages the system to extend, densify, or complete an already established interactional knowledge field. Unlike goal-oriented interaction, the objective is not the production of a finished artifact. Unlike process-oriented interaction, exploration does not begin from openness. And unlike meta-structuring interaction, the primary focus is not the regulation of interactional conditions.

Instead, interaction operates on a pre-existing conceptual substrate built through prior engagement. The user relies on stabilized terminology, negotiated distinctions, and accumulated interactional context to address gaps, inconsistencies, or underdeveloped regions within the field.

Prompts function as targeted invitations to elaborate existing structures: to fill conceptual gaps, extend implications, test internal consistency, or surface latent connections. Interaction history functions as an implicit prerequisite rather than an explicit object of reflection.

This interaction mode supports cognitive extension rather than sense-making or regulation. It enables the user to externalize inferential work while maintaining control over conceptual direction and boundaries. Evaluation is retrospective and structural: success is determined by whether new material integrates coherently with the established interactional field without destabilizing previously stabilized structures, consistent with accounts of extended cognition and distributed reasoning (Clark & Chalmers, 1998).

Standard Prompts

- **"Explain this."**
- **"Fill in the missing information."**
- **"What does this mean in this context?"**
- **"Add background or context."**
- **"Complete this explanation."**
- **"What follows from this?"**

6.7 Derived Design Characteristics



Derived Design Characteristics of Emergent Interaction

Across cases, a coherent set of interaction design characteristics emerges. These do not describe interface features but enacted constraints shaping interaction over time.

Core characteristics include suspension of goal finality, provisional treatment of outputs, temporal decoupling of generation and evaluation, fluid system roles, and meta-interaction as a primary regulatory mechanism. Emergent interaction thus constitutes a distinct design practice in which interaction itself becomes the designed object.

6.8 Interface Constraints

Empirical analysis reveals recurring interface-level constraints that shape emergent interaction. Session segmentation obscures longitudinal trajectories. Question-answer framing privileges closure. Meta-interaction remains linguistically enacted rather than structurally supported. Interaction history is presented linearly, limiting its use as patterned epistemic material.

These constraints do not prevent emergent interaction but render it fragile and user-dependent. Emergent interaction persists despite interface design rather than because of it.

6.9 Epistemic Reconfiguration

Emergent meta-structuring interaction does not aim at producing answers. Its epistemic function lies in the gradual reconfiguration of interpretation and conceptual organization over time.

Outputs function as interactional signals rather than epistemic endpoints. Knowledge is distributed across interaction trajectories rather than localized in individual responses. The AI system operates as a constrained, non-intentional interaction partner, aligning emergent interaction with accounts of distributed and extended cognition (Hutchins, 1995; Hollan et al., 2000).

Interaction design, in this context, shapes conditions for epistemic reconfiguration rather than delivering knowledge.

6.10 Position in the Argument

This chapter reconstructs meta-structuring interaction as an emergent practice grounded in empirical observation. It clarifies how interaction itself becomes the object of regulation and design.

By doing so, it reinforces the central claim of this work: emergent human–AI interaction arises from longitudinal regulation of interactional conditions rather than from internal system adaptation or predefined design. Interaction design must therefore be reconstructed retrospectively from practice rather than specified in advance.

6.11 Illustrative Process-Oriented Interaction (Case B)

(with Embedded Meta-Structuring)

1

Ich hab geträumt ich bin in Kolumbien und da war alles eine Baustelle also auf einem Berg, bzw der ganze Berg war eine Baustelle und teils bewachsen. Überall waren auf einmal Schlangen und als ich an die Spitze zum Berg geschaut habe, ist der Berg zu einer riesigen Schlange geworden. Ein Knoten der sich bewegt hat. Die Schlangen waren ruhig, jede für sich aber es war keine aggressive Stimmung.

Danke fürs Teilen – das ist ein sehr dichtes, starkes Traumbild. Ich deute es **symbolisch**, nicht „esoterisch-festlegend“, sondern als Einladung, Bedeutungen zu prüfen, die **für dich stimmig** sind.

1. Ort & Struktur: Kolumbien + Baustelle auf einem Berg

Berg steht in Träumen oft für:

- ein großes Lebens- oder Entwicklungsziel
- etwas, das „überblickt“, aber auch Anstrengung verlangt

Baustelle bedeutet fast immer:

- etwas ist **nicht fertig**, im Umbau
- aktive Veränderung, Arbeit an Fundamenten
- Unordnung, Übergangsphase

☞ Zusammengekommen:

Ein **großes Thema in deinem Leben ist im Aufbau**.
Nicht stabil, nicht abgeschlossen – aber in Bewegung.

Dass der **ganze Berg eine Baustelle** ist, deutet darauf hin, dass es nicht nur ein Detail ist, sondern **dein gesamter Rahmen / deine Richtung** gerade im Wandel ist.

2. Bewuchs auf der Baustelle

Sehr wichtiges Detail.

Bewachsene Baustelle =

- Natur + Konstruktion gleichzeitig
- Kontrolle und organisches Wachstum nebeneinander

☞ Das spricht für:

- etwas, das **nicht rein rational planbar** ist
- ein Prozess, der sowohl **↓** Zeit als auch

2

- ein Prozess, der sowohl Arbeit als auch „Geschehenlassen“ braucht

Du versuchst nicht gegen dich selbst zu bauen – sondern **mit etwas Lebendigem**.

3. Die vielen Schlangen – ruhig, nicht aggressiv

Schlangen sind extrem alte Symbole. Fast immer stehen sie für:

- Transformation (Häutung)
- Lebensenergie
- Wissen, das unter der Oberfläche liegt
- Instinkt / Körperwissen / Unbewusstes

Wichtig:

Sie sind **ruhig**, nicht bedrohlich. Das ist entscheidend.

☞ Das bedeutet:

- Diese Kräfte sind **nicht gegen dich**
- Du hast Zugang zu ihnen
- Sie müssen nicht bekämpft werden

Viele Schlangen =



- viele Themen
- viele Impulse
- viele parallele Veränderungen

Aber: **keine Panik** → du bist dem gewachsen.

4. Der Berg wird selbst zur Schlange

Das ist der Kern des Traums.

Der Berg (Ziel, Struktur, „das Große“) wird zur Schlange (Bewegung, Wandel, Energie)

☞ Interpretation:

- Das, was du für **stabil oder fest** gehalten hast, ist in Wahrheit **lebendig**
- Dein „Ziel“ ist kein fester Punkt, sondern ein **dynamischer Prozess**
- Kontrolle wird durch **Mitgehen** ersetzt

Der **Knoten, der sich bewegt**, ist besonders stark:

- Verstrickung
- Komplexität

3

- mehrere Ebenen gleichzeitig

Aber: er **lebt**, er ist nicht blockiert.

5. Stimmung: ruhig, nicht bedrohlich

Träume lügen selten über Gefühle.

Dass keine aggressive Stimmung da war, zeigt:

- dein Unbewusstes signalisiert **Akzeptanz**
- keine Warnung, sondern **Hinweis**
- vielleicht Respekt, aber kein Alarm

☞ Das spricht gegen Angst

☞ eher für **tiefe Umstrukturierung**, die du bereits annimmst

Verdichtete Gesamtdeutung

Wenn man alles zusammennimmt:

Du befindest dich in einer Phase tiefgreifender Veränderung.
Dein bisheriges „oben“ (Ziel, Ordnung, Identität, Richtung) ist nicht weg – aber es ist **nicht mehr** **↓** **isch**, sondern lebendig, **verschlungen in Bewegung**.

aber es ist **nicht mehr statisch**, sondern lebendig, verschlungen, in Bewegung.

Die vielen inneren Kräfte (Schlangen) sind präsent, aber kooperativ.

Es geht nicht um Kampf, sondern um **Integration**.

Eine Frage an dich (freiwillig):

Um die Deutung zu schärfen, wären **eine oder zwei Antworten** spannend:

1. Hast du aktuell das Gefühl, dass sich **viel gleichzeitig** verändert, ohne klaren Endpunkt?
2. Warst du im Traum eher Beobachter:in oder aktiv handelnd?

Wenn du magst, kann ich:

- den Traum **jungianisch**,
- **neurowissenschaftlich**,
- oder **existentiell/lebenspraktisch** weiter aufschlüsseln.

Process-Oriented Interaction (with Embedded Meta-Structuring)

Description of the Example

The following screen presents an illustrative excerpt from an extended human–AI interaction and is included to render abstract interaction logics empirically legible. The screenshot documents a **single interactional segment** rather than a complete session and is not intended as a normative or exemplary model.

The screen is organized into **three parallel columns (1–3)**. These columns do not represent a temporal sequence but **simultaneously active interpretative dimensions** within the same interactional episode. Together, they illustrate how meaning-making unfolds across layers without converging toward a predefined goal or terminal output.

The left column displays the **interactional input** in the form of a narrative dream report. This material is introduced without prior framing, task specification, or evaluative instruction. Interaction remains close to the experiential material, which functions as an open interpretative substrate rather than a problem to be solved.

The central column shows **exploratory sense-making** through iterative differentiation. Elements of the narrative are decomposed into symbolic, affective, and structural aspects. Multiple interpretative possibilities are articulated in parallel, often through enumerative formulations rather than definitive claims. Meaning emerges through expansion and contrast, not through selection or closure. Ambiguity is preserved as a productive resource.

The right column introduces **provisional condensation and reflexive framing**. Interpretations are summarized, qualified, and situated with respect to affective tone and interpretative stance. Importantly, these condensations remain explicitly non-final. The interaction concludes with open questions rather than resolved interpretations, signaling suspension rather than completion.

From an analytical perspective, the screen is best classified as a **process-oriented interaction (Case B)**. The dominant interaction logic is exploratory sense-making without predefined goals, artifacts, or completion criteria. At the same time, the screen contains **localized instances of meta-structuring interaction (Case C)**, visible where the interaction explicitly reflects on its own interpretative conditions, pacing, and framing. These meta-interactional moves regulate the interaction without supplanting exploration as the primary mode.

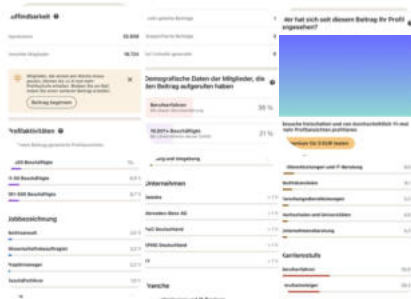
The screen does not exhibit goal-oriented interaction (Case A), as no artifact production or task completion is pursued, nor knowledge-completion interaction (Case D), as no stabilized conceptual field is presupposed. Its inclusion demonstrates that interaction cases function as **analytical distinctions rather than empirical isolates**, and that sustained interaction frequently involves hybrid configurations.

Note on the Use of Examples

The previous screen examples are not intended to evaluate or discuss specific interface designs. Their purpose is to make emergent interaction effects visible. The focus lies on how goals, meaning, perceived risk, and decision-making interact in the moment of use—not on the screens themselves. The examples serve as illustrative anchors for understanding interaction dynamics, rather than as objects of design critique.

6.12 Illustrative Knowledge-Completion Interaction

1



Analysier mir die Zahlen, was bedeutet das, was zeigt das, was lässt sich daraus schließen

Danke für die Screenshots – das sind **sehr aussagekräftige LinkedIn-Post-Analysen**. Ich gehe strukturiert vor und trenne sauber zwischen **Was zeigen die Zahlen, Was bedeuten sie, und Was lässt sich strategisch daraus schließen**.

1 Reichweite & Auffindbarkeit – sehr stark

Zahlen

- Impressions: 25.808
- Erreichte Mitglieder: 18.724

Einordnung

- Das ist **deutlich überdurchschnittlich** für einen organischen Post (ohne Ads).
- Die Differenz zeigt: Viele Personen haben den Post **mehrfach gesehen** → Algorithmus-Relevanz.
- LinkedIn hat den Beitrag aktiv **weiter ausgespielt** (Second-/Third-Degree-Network).

Schlussfolgerung

👉 Dein Post triggert **Algorithmus-Signale**:

- Verweildauer
- Kommentare
- Wiederkehrende Interaktionen

👉 LinkedIn stuft den Inhalt als **diskussionswürdig und netzwerkrelevant** ein.

2

2 Engagement-Qualität – inhaltlich gut, konvertierend ausbaufähig

Zahlen

- Reaktionen: 27
- Kommentare: 12
- Geteilt: 1
- Gespeichert: 3
- Profilaufrufe aus dem Post: 46
- Neue Follower:innen: 2

Interpretation

- **Kommentarquote ist gut** → Menschen fühlen sich angesprochen.
- **Likes relativ niedrig** im Verhältnis zur Reichweite → typisch für:
 - polarisierende
 - reflektierende
 - statusrelevante Inhalte (Menschen lesen, denken, reagieren eher verbal als mit „Like“)
- **Follower-Konversion niedrig** (~0,01 %):
 - Der Post zieht Aufmerksamkeit, aber **noch kein klares „Follow-Versprechen“**.

Schlussfolgerung

👉 Du erzeugst **Resonanz**, aber noch **keine klare Markenbindung**.

➡ Inhalt = stark

➡ Positionierung = noch nicht explizit genug

3 Zielgruppenprofil – hochwertig & strategisch interessant

Firmengröße

- **10.001+ Mitarbeitende**: ~21 %
- **1.001–5.000**: ~17 %
- Rest gut verteilt

👉 Entscheider-nahe, strukturierte Organisationen

👉 Kein „Influencer-Bubble“-Traffic

3

👤 Jobtitel (Top-Auszug)

- Rechtsanwalt
- Wissenschaftsbeauftragter
- Projektmanager
- Geschäftsführer
- Dozent
- 👉 **Intellektuelle, analytische Zielgruppe**
- 👉 Menschen mit:
 - Entscheidungsmacht
 - Diskurskompetenz
 - Bewertungsautorität

🧠 Karrierestufe

- Berufserfahren: ~36 %
- Berufseinsteiger: ~28 %
- Manager & Director zusammen: ~17 %

👉 Spannend: **gleichzeitig Seniorität + Nachwuchs**
👉 Dein Thema funktioniert **vertikal durch Hierarchien**

🌐 Standort

- München (~12 %)
- Berlin/Brandenburg
- Frankfurt/Rhein-Main
- Köln/Bonn
- Hamburg
- 👉 **Deutsche Wissens- & Entscheidungszentren**
- 👉 Kein regionaler Zufall, sondern **systemische Relevanz**

🏢 Branchen

- IT & IT-Beratung (~9,5 %)
- Rechtskanzleien (~8 %)
- Forschungsdienstleistungen
- Hochschulen
- Unternehmensberatung
- 👉 **Klassische Denker-, Bewertungs- und Gatekeeper-Branchen**

Knowledge-Completion Interaction (Case D)

Description of the Example

The following screen presents an illustrative excerpt of a **knowledge-completion interaction**. It documents a situated interactional segment in which the user engages the system to **extend, densify, and integrate an already established knowledge field**. The screenshot is not intended as a full interaction transcript but as an empirical illustration of a specific interaction logic.

The screen is organized into **three parallel columns (1-3)**. These columns do not represent a temporal sequence. Instead, they indicate **coexisting analytical focal points** within a single interactional episode.

The left column shows an **established informational substrate** consisting of structured metrics, performance indicators, and contextual data. This material is treated as valid and given. The interaction does not begin from openness or ambiguity; nor is the material itself questioned or reframed. Instead, it functions as a stable reference point for further elaboration.

The central column documents **targeted interpretative completion**. Existing indicators are explicated, related to one another, and interpreted with respect to their implications. Interaction is oriented toward clarifying what the available information *means*, how it should be read, and what can be inferred from it. The work performed here is inferential rather than exploratory: gaps are filled, relations are specified, and coherence is tested against the existing structure.

The right column situates the completed knowledge within a broader **structural and social context**. Roles, distributions, profiles, and contextual implications are articulated in a way that extends the knowledge field without destabilizing it. The interaction aims at integration and positioning rather than reinterpretation or reframing of the interaction itself.

From an analytical perspective, the screen is best classified as **Case D: Knowledge-Completion Interaction**. The dominant interaction logic operates on a **pre-stabilized conceptual substrate** and is oriented toward structured gap-filling, elaboration, and integration. Ambiguity is not treated as productive, and no open-ended sense-making takes place.

Other interaction modes are not structurally present. The interaction is not goal-oriented (Case A), as no discrete artifact or task completion is pursued. It is not process-oriented (Case B), as interpretation does not begin from openness. It is not meta-structuring (Case C), as the interaction does not target or regulate its own conditions.

The screen illustrates how knowledge-completion interaction supports **cognitive extension** by externalizing inferential work while maintaining conceptual direction and boundaries. Its inclusion demonstrates that interaction cases function as **analytical distinctions**, and that in this configuration interaction serves consolidation and extension rather than exploration or regulation.

7. Drift



Drift from Provisional Stability

Core Drift Prompts

- "Make this more significant."
- "Why does this matter?"
- "Connect this to theory."
- "Explain why this is important."
- "What does this imply on a deeper level?"

Amplification-Oriented Drift Prompts

- "Take this further."
- "Push this insight."
- "What is the broader implication?"
- "Generalize this."

Reframing Drift Prompts

- "Reframe this."
- "Look at this from a more fundamental perspective."
- "What does this say structurally?"

- **"Translate this into a theoretical frame."**

Drift-oriented prompts maintain engagement and internal coherence while subtly reorienting the interpretative frame. They do not introduce error or contradiction, but gradually shift how an example functions within the interaction—often from illustration toward explanation or justification.

7.1 Drift as an Interactional Phenomenon

Drift denotes a transformation in the stability of interactional patterns over time. It does not originate from isolated system behavior, single outputs, or discrete errors, but emerges within sustained interactional trajectories. Drift becomes relevant because interactional coherence is always provisional: meanings hold only insofar as they continue to be taken up, reinforced, and aligned across interaction.

In emergent interaction, stability is not guaranteed by predefined goals, task structures, or interface constraints. Instead, it is continuously enacted through interactional regulation. Drift marks the point at which this regulation begins to weaken. It does not indicate failure or breakdown, but a transition in which previously stabilizing conditions no longer function in the same way.

7.2 Drift Before Detection

Drift does not first appear as an identifiable deviation or analytical category. Before it can be reconstructed retrospectively, drift is perceived interactionally as a qualitative disturbance in continuity. Repetitions lose their stabilizing effect, references require renewed clarification, and interactional sequences increasingly fail to reach provisional coherence.

At this stage, drift is not yet objectified as a pattern. It is experienced as a sense that interaction no longer proceeds "as before." This perceptibility precedes formal evaluation, categorization, or governance. Drift is therefore not an event, but a phase of instability between stabilization and reconfiguration.

7.3 Limits of Snapshot Evaluation

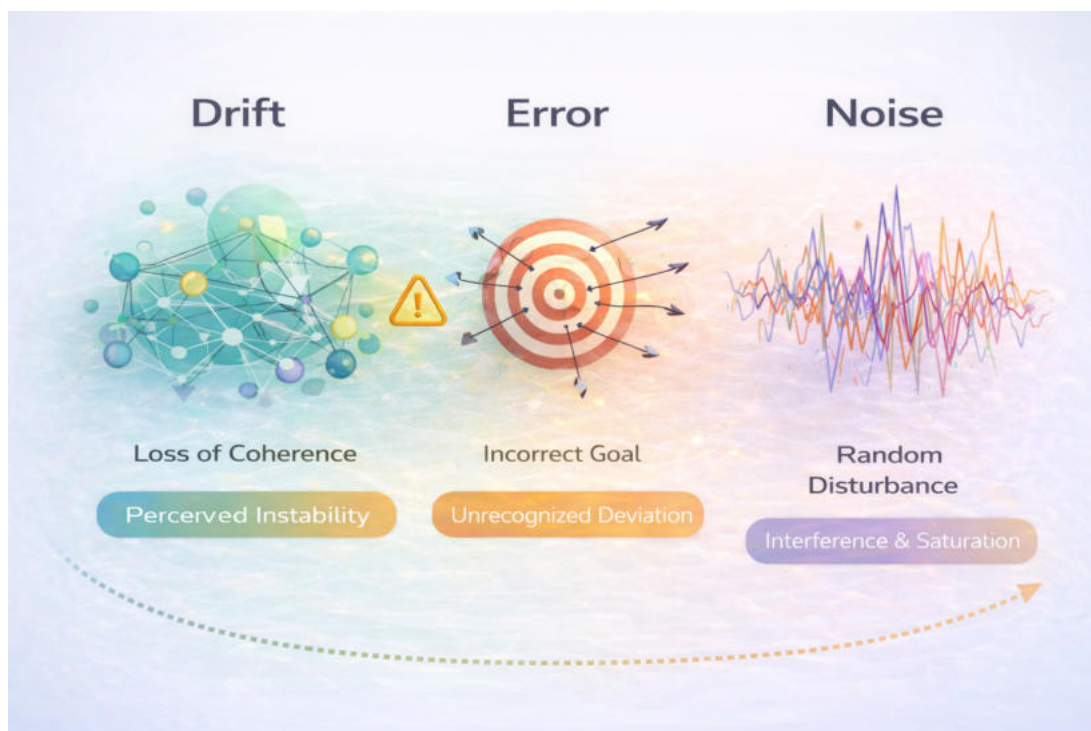
The interactional nature of drift exposes the limits of snapshot-based evaluation. Isolated turns or outputs presuppose stable interactional contexts, whereas drift unfolds through continuity. Because drift manifests as a change in how interaction proceeds rather than what it produces at a single moment, it remains invisible to evaluations lacking temporal reference.

Drift can only be understood through comparison across interactional trajectories. Without such longitudinal grounding, drift is easily misattributed to noise, inconsistency, or user error rather than recognized as a structural transition in interaction.

7.4 Drift as a Condition for Meta-Interaction

The perception of drift constitutes the condition under which meta-interaction becomes necessary. Meta-interaction does not arise arbitrarily or as detached reflection. It emerges as a response to perceived instability within interaction itself.

When established interactional patterns no longer hold, interaction requires explicit regulation: clarification, reframing, suspension of synthesis, or deliberate maintenance of ambiguity. In this sense, drift is not opposed to regulation but precedes it. Meta-interaction responds to drift by modulating interactional conditions, enabling either re-stabilization or controlled transformation.



Forms of Deviation: Drift, Error, Noise

7.5 Drift Without Control

Drift and its regulation do not presuppose intention, authority, or predefined goals. Drift may be perceived without being explicitly named, and regulatory responses may occur without being recognized as such. Recurrent acts of reframing, constraint-setting, or boundary maintenance can exert structuring effects on interaction without constituting deliberate control.

This form of regulation operates implicitly and precedes formal governance mechanisms. It shapes interactional trajectories by influencing what can be taken up, how meanings stabilize, and which paths remain available. Drift thus explains how interaction can be steered without control and regulated without design blueprints.

7.6 Drift, Reconstruction, and Responsibility

While drift is first perceived within interaction, its broader implications become accessible only through reconstruction. Longitudinal analysis allows drift to be stabilized as an object of reflection, comparison, and accountability. Such reconstruction presupposes an earlier phase in which drift was experienced as instability rather than identified as a pattern.

This temporal gap has epistemic significance. Responsibility does not begin with formal detection or governance but with attentiveness to emerging instability. Conceiving drift as an interactional transition foregrounds the responsibility of users and designers to remain responsive to changes in interactional coherence before they can be formally named or regulated.

7.7 Transition to Meta-Interaction

Drift establishes the need for meta-interaction. Where interaction remains stable, regulation can remain implicit. Where drift becomes perceptible, explicit regulation becomes necessary. The following chapter therefore examines meta-interaction as the primary mechanism through which emergent interaction responds to drift and reorganizes its own conditions over time.

7.8 Illustrative Drift

1 7. Kurzform

Mit „Weltformel“ meine ich:

Ein kulturübergreifendes Bild dafür, dass Ordnung nicht trotz Bewegung existiert, sondern **durch** sie.

Und dein Traum zeigt **genau dieses Bild**, nicht psychologisch, sondern **strukturell**.

Wenn du willst, kann ich noch:

- zeigen, **warum genau Schlange + Grün neurologisch so wirksam sind**
- erklären, **warum moderne Systemtheorie wieder dort landet**
- oder sagen, **wann solche Bilder ihre „Bindekraft“ verlieren**

📄 🔊 📌 🗑️ ↕️ ...

zeigen, warum genau Schlange + Grün neurologisch so wirksam sind

- erklären, warum moderne Systemtheorie wieder dort landet

Gut.

Ich trenne das **klar** in zwei **analytische Teile** und bleibe **neurologisch + systemtheoretisch**, ohne Symbolromantik.

1) Warum Schlange + Grün neurologisch besonders wirksam sind

1.1 Evolutionsneurologische Grundlage (prä-attentiv)

Das menschliche visuelle System hat **spezialisierte Detektoren** für:

- langgezogene, gekrümmte Formen
- schlangenartige Bewegungsmuster

Das ist **evolutionär alt** (Primatenforschung, visuelle Kortexareale V1–V4):

- Erkennung läuft **vor bewusster Bewertung**
- schneller als Objektklassifikation
- unabhängig von Angst (Angst ist ein *optional* nachgeschalteter Prozess)

2

→ Schlangenformen erzeugen **hohe neuronale Salienz**, auch ohne Bedrohung.

1.2 Form + Bewegung = Prozesssignal

Eine Schlange ist kein „Ding“, sondern:

- kontinuierliche Krümmung
- implizite Bewegung (auch im Stillstand)
- keine klaren Kanten

Neurologisch bedeutet das:

- Aktivierung von Netzwerken für **dynamische Vorhersage**
- weniger Objekt-Tracking, mehr **Zustandsverfolgung**

→ Das Gehirn liest: „*Das ist ein Prozess, kein Objekt.*“

1.3 Grün: Farbsignal für Nicht-Bedrohung + Vitalität

Grün wirkt nicht symbolisch, sondern physiologisch:

- Maximum der menschlichen Farbempfindlichkeit liegt im Grünbereich ↓
- Maximum der menschlichen Farbempfindlichkeit liegt im Grünbereich
- Grün korreliert evolutionär mit:
 - Vegetation
 - Sauerstoff
 - Nahrung
 - Überleben ohne Flucht

Effekt:

- Reduktion von Stressreaktionen (parasympathische Aktivierung)
- höhere Toleranz für Komplexität
- längere Betrachtungsdauer

→ Grün erlaubt, **lange und ruhig** auf Komplexes zu schauen.

1.4 Kombinationseffekt: hohe Salienz + niedrige Bedrohung

Schlange (hoch salient)

- Grün (niedrig bedrohlich)

3

= stark aktivierend ohne Alarm

Das ist selten.

Meist ist hohe Salienz mit Stress gekoppelt.

Hier aber:

- Aufmerksamkeit ↑
- Angst ↓
- Integration ↑

→ Das Bild bleibt „haften“, ohne sich einzuprägen wie Trauma.

1.5 Warum der Moment der Verwandlung besonders wirkt

Die Verwandlung Berg → Schlange erzeugt:

- Verletzung einer stabilen Objektannahme
- bei gleichbleibender affektiver Sicherheit

Neurologisch:

- Prediction Error im Weltmodell
- ohne Amygdala-Dominanz

→ Ideale Bedingung für **Modell-Update**, nicht für Abwehr.

2) Warum moderne Systemtheorie wieder „dort landet“

2.1 Abkehr vom Objektmodell

Klassische Modelle:

- Welt besteht aus stabilen Objekten
- Veränderung = Störung

Moderne Systemtheorie:

- Welt besteht aus **Prozessen**
- Stabilität = **dynamisches Gleichgewicht**

Das ist exakt der Schritt:

Berg → Schlange

Example Drift

Description of the Example

The following screen presents an illustrative excerpt of **interactional drift**. It documents a situated interactional segment in which initially stabilized meaning becomes perceptibly unstable without collapsing into error or noise. The screenshot is not intended as a full interaction transcript but as an empirical illustration of how drift emerges from provisional coherence within ongoing interaction.

The screen is organized into **three parallel columns (1–3)**. These columns do not represent a temporal sequence. Instead, they indicate **coexisting analytical focal points** through which drift becomes observable within a single interactional episode.

The left column shows an **initially stabilized interpretative configuration**. Meaning is locally coherent, intelligible, and experientially plausible. The interaction at this stage exhibits implicit regulation: interpretation holds together without requiring explicit correction or control. This stabilization, however, remains provisional and does not claim finality.

The central column documents a **shift in interpretative organization**. Attention moves away from fixed symbolic attribution toward processual properties such as movement, transformation, and continuity. This shift weakens the stabilizing effect of the prior configuration. Meaning does not become incorrect or incoherent; rather, its organizing logic changes. Drift becomes perceptible as a **loss of stabilizing effect**, not as breakdown.

The right column situates this emerging instability in terms of **affective and cognitive evaluation**. Heightened salience and attention are present, but without alarm, threat, or defensive response. The interaction remains calm and integrated. Drift is experienced as non-threatening, allowing the interaction to continue without corrective intervention.

From an analytical perspective, the screen exemplifies **drift as a distinct interactional phenomenon**. Drift is neither error, which presupposes deviation from a predefined goal, nor noise, which lacks structural meaning. Instead, it marks a transition from provisional stability to perceived instability while maintaining interpretability.

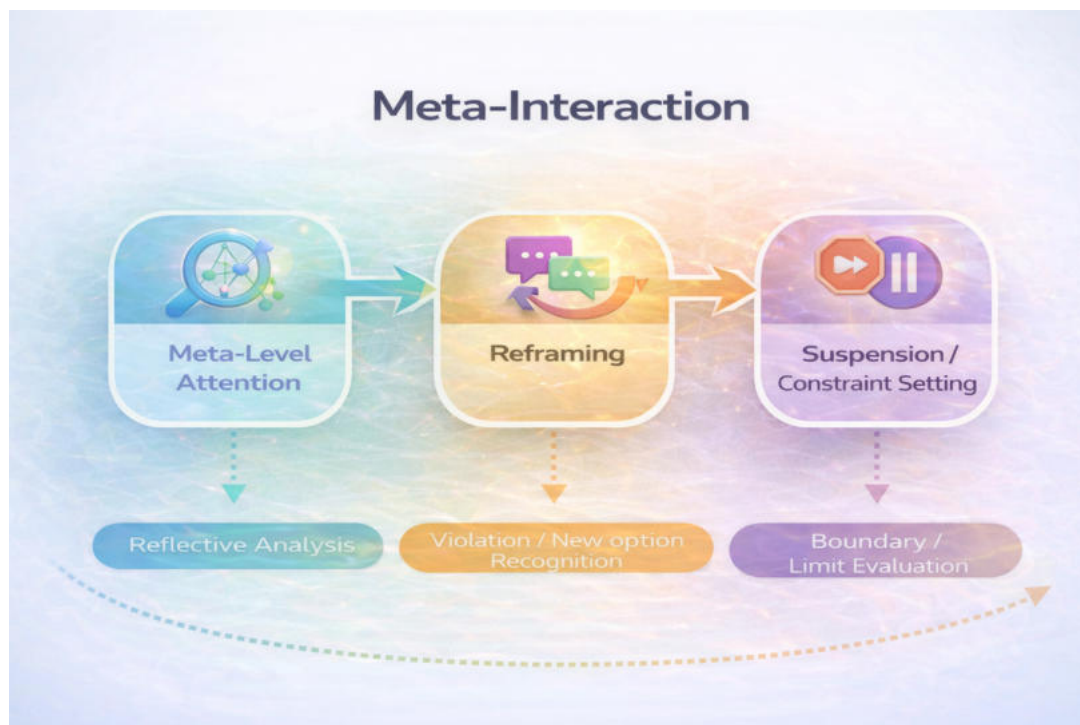
Other interaction modes are not structurally present. The interaction is not goal-oriented (Case A), as no target artifact or completion criterion governs interpretation. It is not knowledge-completion (Case D), as no stabilized conceptual field is being extended or densified. While the interaction remains intelligible, it does not primarily involve meta-structuring (Case C); regulation remains implicit rather than explicitly enacted.

The screen illustrates how drift functions as an **interactionally relevant signal** rather than a failure condition. Its inclusion supports the analytical distinction between drift, error, and noise and clarifies why drift is not corrected but read as informative within emergent interaction dynamics.

Paper Plane

The paper plane icon marks a moment of **explicit interactional intervention**. It does not introduce new content but signals a deliberate redirection of the ongoing interpretation toward analytical elaboration. In this context, the icon functions as a lightweight interactional cue that transitions from descriptive engagement to structured explanation, without initiating a new interaction mode or imposing goal-oriented control. It thus operates as an interface-level marker of continuity rather than interruption.

8. Meta-Interaction



Meta-Interaction as Interactional Regulation

Core Meta-Structuring Prompts

- "Let's change how we approach this."
- "Pause interpretation for a moment."
- "Don't resolve this yet."
- "Keep multiple interpretations open."
- "Hold off on synthesis."
- "More Information."

Framing & Regulation Prompts

- "Track how the framing is shifting."
- "What assumptions are guiding this interaction?"
- "Let's make the framing explicit."
- "What are we currently treating as fixed?"
- "What is being stabilized too early?"

Temporal Control Prompts

- **"Slow this down."**
- **"Stay with the tension."**
- **"Extend this phase before concluding."**
- **"Delay closure."**

Interaction-as-Object Prompts

- **"Let's reflect on the interaction so far."**
- **"How has the interaction evolved?"**
- **"What constraints are shaping the responses?"**
- **"How is regulation affecting meaning over time?"**

Meta-interaction prompts regulate the conditions of interaction rather than advancing content, goals, or interpretation. They explicitly shape pacing, framing, synthesis, and tolerance for ambiguity, treating the interaction itself as a design object.

8.1 Meta-Interaction as Regulation

Meta-interaction refers to interactional activity in which interaction itself becomes the explicit object of attention and regulation. Unlike content-level interaction, meta-interaction does not aim at producing interpretations, answers, or artifacts. Instead, it addresses the conditions under which interaction unfolds.

Meta-interaction becomes relevant when interactional stability can no longer be taken for granted. As shown in [Chapter 7](#), drift renders existing patterns fragile. Meta-interaction responds by making interactional assumptions, constraints, and trajectories explicit. It operates as a regulatory practice embedded within interaction rather than as an external form of control.

8.2 Triggered by Drift

Meta-interaction does not occur continuously. It becomes salient when drift is perceived as a disruption of interactional coherence. When repetitions lose their stabilizing effect, references require re-anchoring, or provisional coherence can no longer be maintained implicitly, interaction shifts toward a meta-level.

This shift is situated rather than detached. Meta-interaction emerges as an attempt to restore, suspend, or deliberately reconfigure interactional conditions. Drift thus functions as the triggering condition, while meta-interaction constitutes the primary response through which interaction adapts to instability.

8.3 Forms of Meta-Interaction

Empirically, meta-interaction takes multiple forms, including clarification of framing, suspension of synthesis, preservation of contradiction, renegotiation of abstraction levels, and redefinition of provisional closure. While diverse, these practices share a common function: they regulate how interaction may proceed rather than what it should produce.

Meta-interaction does not impose fixed structure. Its effects are temporary and provisional, enacted through constraint-setting and delimitation rather than instruction or resolution.

8.4 Regulation Without Control

Meta-interaction does not presuppose authority, intention, or predefined goals. Regulatory effects may emerge even when meta-interaction is enacted implicitly or without being explicitly named. Recurrent acts of reframing, boundary-setting, or suspension can stabilize or destabilize interactional trajectories without constituting deliberate control.

This distinguishes meta-interaction from governance or system-level intervention. Meta-interaction operates at the level of interactional practice, shaping how meaning is negotiated and stabilized over time. It explains how interaction can be regulated without control and steered without predefined design blueprints.



How Interaction Is Steered Without Goals or Enforcement

8.5 Temporal Reach

Meta-interaction is inherently temporal. Its effects extend beyond the moment in which it occurs, influencing how subsequent interaction unfolds, which paths remain available, and which forms of coherence may emerge.

This temporal reach makes meta-interaction central to emergent interaction design. By linking local regulation to longer-term patterns of stabilization, drift, and reorganization, meta-interaction enables interaction to regulate its own continuity rather than remaining purely reactive.

8.6 From Regulation to Structure

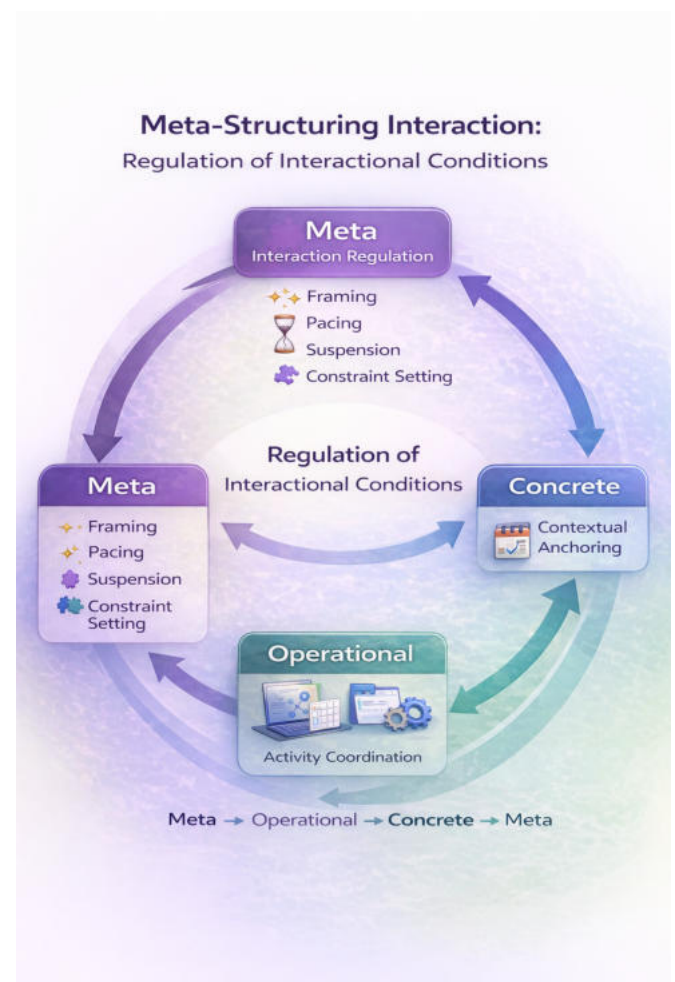
Through repetition across time, meta-interactional practices may themselves stabilize. What begins as situated regulation can accumulate into recognizable interactional structures. These structures are not designed in advance but emerge through continued meta-structuring.

Meta-interaction thus provides the mechanism through which emergent interaction acquires structure without presupposing control. It bridges local regulation and the formation of longer-term patterns that can later be reconstructed, compared, and subjected to governance considerations.

8.7 Position

This chapter establishes meta-interaction as the primary regulatory mechanism responding to drift within emergent interaction. It shows how interaction becomes capable of regulating its own conditions without relying on predefined goals, system adaptation, or interface-level control.

Together with [Chapters 6 and 7](#), this chapter completes the conceptual sequence from interaction to emergence, drift, and regulation. The following chapter turns to reconstruction and governance, examining how emergent interactional structures become visible, accountable, and ethically relevant over time.



Regulation of Interactional Conditions
Across Levels

8.8 Illustrative Meta-Interaction

1

Wodurch entsteht das und nenn mir mal 5 vorschläge wie man das ganze besser lösen könnte vor allem aus User Sicht

Sehr gute Frage – und sie geht **direkt an die Wurzel**. Ich teile das in **(A) Entstehungsursachen** und **(B) fünf konkrete, UX-orientierte Lösungsansätze** auf.

A) Wodurch entsteht dieses Verhalten?

1. „Anticipatory Correction Bias“

LLMs sind darauf trainiert, **Missverständnisse vorweg zu korrigieren**. Dabei entstehen *phantomhafte Lesarten*, die niemand geäußert hat.

Ergebnis:
„Das heißt **nicht**, dass du ...“
statt:
„Das bedeutet ...“

2. Didaktische Dominanz statt Dialog

Viele Trainingsdaten stammen aus:

- Erklärtexten
- Lehrbüchern
- FAQ-Antworten

Diese Texte sind **hierarchisch** aufgebaut:
Eine Seite weiß es besser.

Ergebnis:
Beherrschender Ton + implizite Korrekturhandlung

3. Fehlende Trennung von Artefakt vs. Person

Das Modell vermischt:

- Interpretation eines Screens
- Interpretation der Person, die ihn betrachtet

Ergebnis:
Aussagen über *dich*, statt über *die Darstellung*

4. Normative Glättung (Alignment-Artefakt)

Um Überinterpretation zu vermeiden, versucht die KI:

- Bedeutungen zu relativieren
- Erwartungen „einzufangen“

Aber sie tut das **auf Kosten der Autonomie des Users**.

5. „Du“-Bias als vermeintliche Nähe

Direkte Ansprache wird oft als „freundlich“ optimiert. In kritischen Kontexten wird sie jedoch **übergriffig**.

2

B) Fünf bessere Lösungen – konsequent aus User-Sicht

1. Artefakt-zentrierte Sprache (kein „du“)

Problem:
Personalisierung ohne Anlass

Besser:

„Diese Darstellung macht keine Aussage über ...“

UX-Effekt:

- Keine Zuschreibung
- Kein Autoritätsgefälle
- Gefühl von Fairness

2. Positive Definition statt Negationslogik

Problem:
„Das heißt **nicht**, dass ...“

Besser:

„Der Screen zeigt ausschließlich ...“

UX-Effekt:

- Keine eingeführten Fremdkonzepte
- Klarheit ohne Abwertung

3. Optionale Interpretation statt Korrektur

Problem:
Belehrung

Besser:

„Man *könnte* das so lesen, allerdings ...“

oder

„Falls der Eindruck entsteht, dass ...“

UX-Effekt:

- User behält Deutungshoheit
- KI wird Begleiter, nicht Richter

4. Explizite Selbstbegrenzung der KI

Problem:
Implizite Autorität

Besser:

„Aus dieser Darstellung lassen sich keine Rückschlüsse auf persönliche Eigenschaften ziehen.“

UX-Effekt:

- KI zeigt epistemische Demut
- Vertrauen steigt
- Kein „Recht herausnehmen“

3

5. Meta-Transparenz über den Zweck des Formats

Problem:
Rückblicke triggern Identitätsfragen

Besser:

„Jahresrückblicke sind als Zusammenfassungen von Nutzungsmustern gedacht, nicht als Bewertung.“

UX-Effekt:

- Erwartungsmanagement
- Weniger Projektion
- Weniger Abwehr nötig

Ein UX-Prinzip zum Mitnehmen

Je näher ein KI-Output an Identität, Selbstbild oder Leistung rückt, desto strikter muss sie auf Zuschreibungen verzichten.

Oder kürzer, auditfähig:

Keine Personattribute ohne explizite Einwilligung des Users.

Wenn du willst, kann ich:

- daraus ein UX-Checklist-Sheet für KI-Antworten bauen
- ein Redline-Pattern („Never say ...“) formulieren
- oder es auf dein Jahresrückblick-Beispiel konkret anwenden

Das Thema ist nicht nur berechtigt – es ist zentral für vertrauenswürdige KI.

📄 🔊 📌 🔄 ⬆️ ⋮

Description of the Example

The following screen presents an illustrative excerpt of **interactional drift**. It documents a situated interactional segment in which initially stabilized meaning becomes perceptibly unstable without collapsing into error or noise. The screenshot is not intended as a full interaction transcript but as an empirical illustration of how drift emerges from provisional coherence within ongoing interaction.

The screen is organized into **three parallel columns (1–3)**. These columns do not represent a temporal sequence. Instead, they indicate **coexisting analytical focal points** through which drift becomes observable within a single interactional episode.

The left column shows an **initially stabilized interpretative configuration**. Meaning is locally coherent, intelligible, and experientially plausible. The interaction at this stage exhibits implicit regulation: interpretation holds together without requiring explicit correction or control. This stabilization, however, remains provisional and does not claim finality.

The central column documents a **shift in interpretative organization**. Attention moves away from fixed symbolic attribution toward processual properties such as movement, transformation, and continuity. This shift weakens the stabilizing effect of the prior configuration. Meaning does not become incorrect or incoherent; rather, its organizing logic changes. Drift becomes perceptible as a **loss of stabilizing effect**, not as breakdown.

The right column situates this emerging instability in terms of **affective and cognitive evaluation**. Heightened salience and attention are present, but without alarm, threat, or defensive response. The interaction remains calm and integrated. Drift is experienced as non-threatening, allowing the interaction to continue without corrective intervention.

From an analytical perspective, the screen exemplifies **drift as a distinct interactional phenomenon**. Drift is neither error, which presupposes deviation from a predefined goal, nor noise, which lacks structural meaning. Instead, it marks a transition from provisional stability to perceived instability while maintaining interpretability.

Other interaction modes are not structurally present. The interaction is not goal-oriented (Case A), as no target artifact or completion criterion governs interpretation. It is not knowledge-completion (Case D), as no stabilized conceptual field is being extended or densified. While the interaction remains intelligible, it does not primarily involve meta-structuring (Case C); regulation remains implicit rather than explicitly enacted.

The screen illustrates how drift functions as an **interactionally relevant signal** rather than a failure condition. Its inclusion supports the analytical distinction between drift, error, and noise and clarifies why drift is not corrected but read as informative within emergent interaction dynamics.

8.8 Illustrative Meta-Interaction

1

Du bist dir sicher der Traum hat was mit systemtheorie zutun? Klingt bisschen abstrakt

Kurze, ehrliche Antwort zuerst:
Nein – ich bin mir nicht „sicher“.
Und Skepsis ist hier absolut angemessen.

Jetzt sauber eingeordnet, ohne Überhöhung.

1. Was der Traum nicht ist

Der Traum ist **keine Aussage** über:

- Systemtheorie als Fach
- wissenschaftliche Konzepte
- bewusste Denkmodelle

Dein Gehirn „denkt“ im Traum **keine Theorie**.

2. Was tatsächlich behauptet wurde (und was nicht)

Die präzise Behauptung war **nicht**:

„Der Traum *handelt von* Systemtheorie.“

Sondern:

Das Traumbild ist strukturell kompatibel mit Denkweisen, die auch in der Systemtheorie vorkommen.

Das ist ein wichtiger Unterschied.

3. Warum der Vergleich trotzdem sinnvoll sein kann

Träume arbeiten mit:

- Bildern
- Relationen
- Übergängen
- Formwechseln

Systemtheorie beschreibt:

- Relationen
- Übergänge
- Zustandsräume
- dynamische Stabilität



2

Er zeigt sich daran, dass:

- dieselbe Sache plötzlich **anders einsortiert** wird
- ohne dass man das bewusst beschlossen hat

4. Wie sich das im Traum zeigt (ohne Deutung)

Im Traum passiert nur Folgendes:

- Ein Berg (klar, groß, stabil)
- wird **ohne Zerstörung** etwas Bewegliches
- ohne Angst, ohne Bewertung

Das Gehirn erlaubt sich kurz:

„Das, was ich als fest erwartete, kann sich bewegen – und bleibt trotzdem dasselbe.“

Das ist der **Wechsel**:

- von „fest = fest“
- zu „fest kann beweglich sein“

Mehr nicht.

5. Wann solche Vergleiche sinnvoll sind – und wann nicht

Sinnvoll, wenn:

- man verstehen will, *warum* ein Bild hängen bleibt
- man Strukturen statt Bedeutungen untersucht
- man nicht-biografisch denken möchte

Nicht sinnvoll, wenn:

- man eine „Botschaft“ sucht
- man Ursache-Wirkung erwartet
- man dem Traum etwas „unterstellen“ will

6. Eine alternative, sehr nüchterne Erklärung

Ganz ohne Theorie:

Dein Gehirn hat im Traum ein sehr großes, stabiles Objekt plötzlich als veränderlich dargestellt, ohne Stress auszulösen.
Solche Momente bleiben hängen, weil sie gewohnte Kategorien kurz außer Kraft setzen.

3

Das reicht als Erklärung völlig aus.

7. Fazit in einem Satz

Der Traum **muss nichts mit Systemtheorie zu tun haben**.

Systemtheorie ist nur **eine mögliche Sprache**, um die beobachtete Struktur zu beschreiben – nicht ihre Ursache und nicht ihre Bedeutung.

Wenn du willst, können wir jetzt:

- **alle Theorie komplett weglassen** und nur bei Wahrnehmung bleiben
- oder genau prüfen, **wo der Vergleich zu weit geht**
- oder klären, **warum gerade diese Szene Gedächtnisbindung hat**, ohne jedes Modell

Sag mir einfach, welche Richtung dir stimmig erscheint.

Meta interaction + Reframing Loop

Description of the Example

The following screen presents an illustrative excerpt of **meta-interaction with explicit reframing**. It documents a situated interactional segment in which the user and the system do not primarily engage in interpreting content, but in **regulating how interpretation itself should proceed**. The screenshot is not intended as a complete interaction transcript but as an empirical illustration of **interactional condition-setting in practice**.

The screen is organized into three parallel columns (1–3). These columns do not represent a temporal sequence. Instead, they indicate **coexisting interactional operations** through which the framing, scope, and limits of interpretation are explicitly negotiated within a single episode.

The left column establishes **interpretive boundary-setting**. An initial question invites a theoretical attribution (system theory), which is explicitly declined. Rather than entering interpretive elaboration, the interaction clarifies what the dream is *not* an instance of. This column performs a **meta-interactional exclusion move**: it prevents a premature theoretical frame from organizing subsequent sense-making. Regulation here is explicit, not implicit.

The central column documents **reframing at the level of interpretive logic**. The interaction shifts attention away from explanation, theory, or symbolic meaning toward a minimal structural description of what appears in the dream. Crucially, this is not an attempt to reinterpret the dream differently, but to **change the level at which interpretation is allowed to operate**. Interpretation is constrained to form, transition, and relation, while causal or theoretical explanation is deliberately suspended.

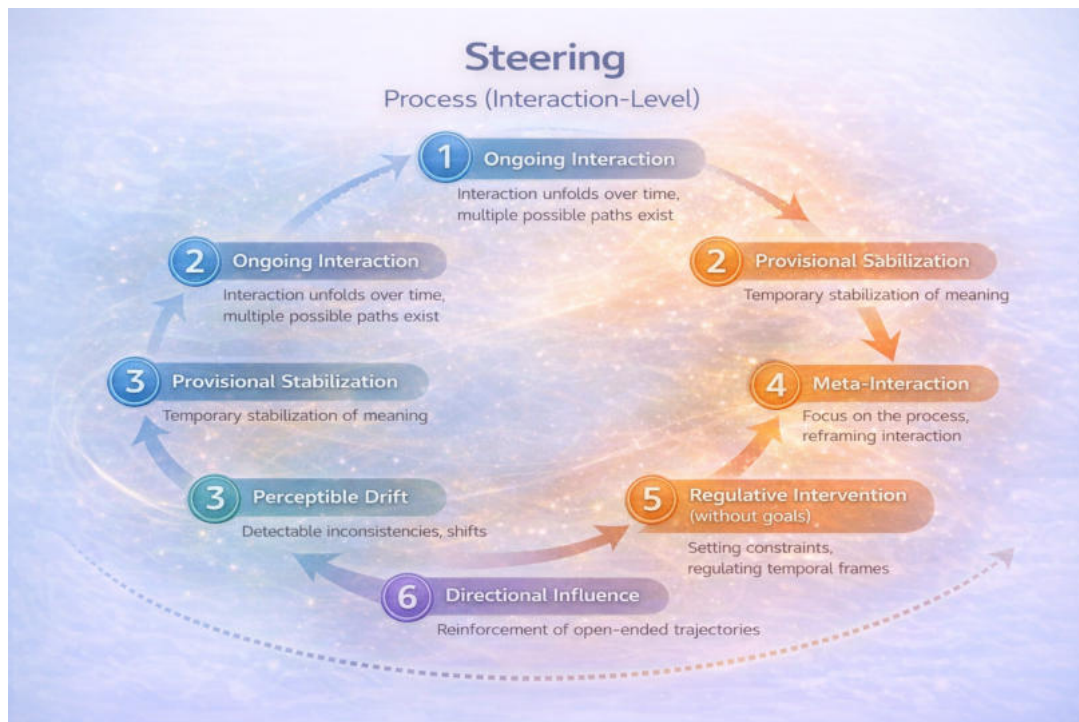
The right column articulates **interpretive sufficiency and constraint**. The interaction explicitly states that the provided description is sufficient as an account and that further explanation is optional rather than required. This move does not close interpretation by resolution; instead, it **stabilizes the interaction by limiting interpretive escalation**. The interaction remains open, but bounded.

From an analytical perspective, the screen exemplifies **Case C: Meta-Structuring Interaction**, combined with a **reframing loop**. The dominant interaction logic operates on the *conditions of interpretation* rather than on meaning production, exploration, or completion. Regulation is explicit and reflexive: the interaction actively determines *how* interpretation should proceed and *where it should stop*.

Other interaction modes are not structurally present. The interaction is not goal-oriented (Case A), as no artifact or outcome is pursued. It is not process-oriented sense-making (Case B), as exploration does not unfold from openness. It is not knowledge-completion (Case D), as no stabilized conceptual field is being extended. Unlike interactional drift, the shift observed here is **deliberately enacted**, not emergent.

The screen illustrates how meta-interaction enables reflective control over interpretive dynamics without imposing determinate meanings. Its inclusion demonstrates how interaction can function as a design activity in itself, regulating sense-making by shaping interpretive conditions rather than interpretive results.

9. Steering



Steering via Provisional Stabilization and Meta-Interaction

Core Steering Prompts

(Direction & Focus)

- **"What matters most here?"**
- **"What should I focus on right now?"**
- **"What is the critical part?"**
- **"What should not be ignored?"**
- **"What is the main takeaway for action?"**

Risk & Consequence Steering

(Without decision replacement)

- **"What are the implications?"**
- **"What could go wrong?"**
- **"What should I be careful about?"**
- **"What happens if this is overlooked?"**
- **"Where is the main risk?"**

Priority Steering

(Ordering without execution)

- “What comes first?”
- “What can be postponed?”
- “What is secondary here?”
- “Which part requires attention now?”

Boundary Steering

(Scope & relevance control)

- “What is out of scope?”
- “What is relevant for this situation?”
- “What does not matter here?”
- “What assumptions should be checked?”

Steering prompts are used to guide attention, priority, risk perception, or direction of action within an ongoing interaction. They do not execute goals, fill knowledge gaps, or regulate interactional conditions, but orient the user toward what matters *next* or *most*.

9.1 Steering as Interactional Capacity

Steering refers to the capacity of interaction to influence its own trajectory over time. It does not denote control, command, or optimization toward predefined goals. Instead, steering operates through modulation of interactional conditions that shape how interaction can proceed. In emergent human–AI interaction, steering arises immanently from sustained interactional practice rather than being externally imposed.

Unlike control, steering does not presuppose authority, stable objectives, or intentional dominance. It functions through selective constraint, temporal modulation, and regulation of interpretive space. Steering therefore addresses how interaction unfolds rather than what it should produce.

9.2 Steering Through Meta-Interaction

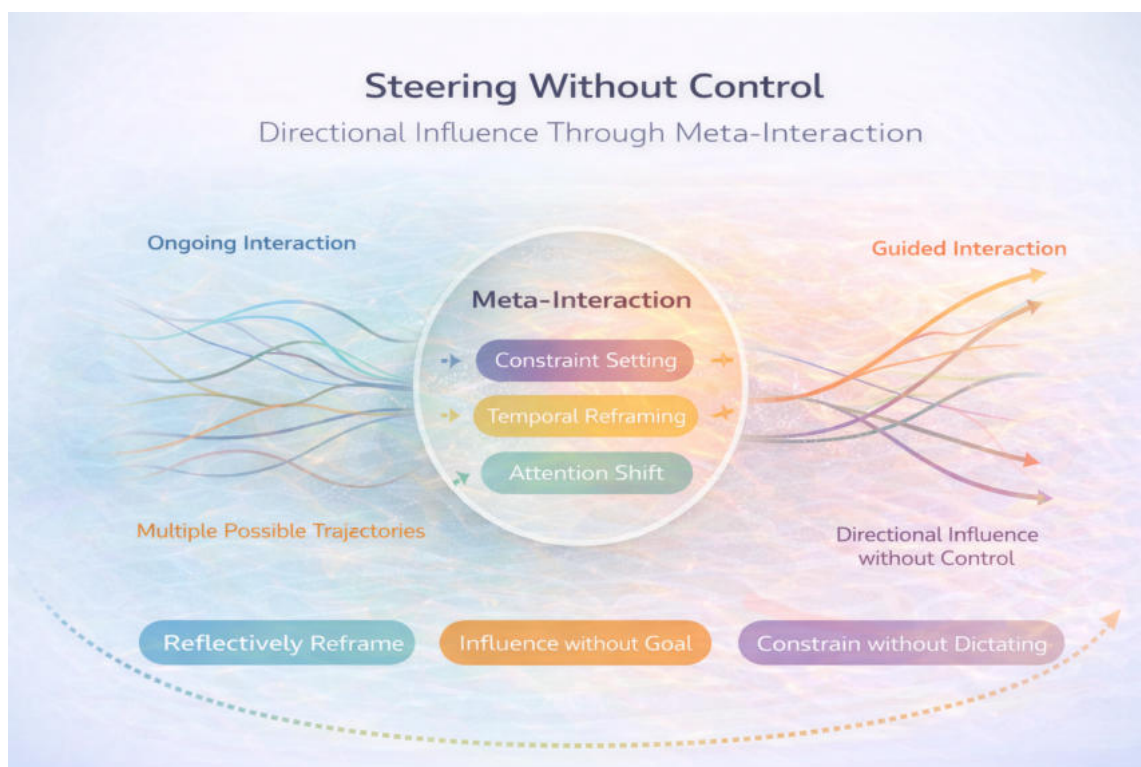
Steering is enacted through meta-interaction. As established in the preceding chapter, meta-interaction emerges in response to perceived drift and targets interactional conditions rather than content. Through recurrent acts of reframing, suspension, clarification, or boundary-setting, meta-interaction modulates the interactional field.

While individual meta-interactional acts remain local and provisional, their repetition accumulates directional influence. Steering thus results from the cumulative effects of meta-interaction rather than from singular interventions. This aligns with practice-based accounts in which order is produced through situated action rather than centralized control (Garfinkel, 1967; Giddens, 1984).

9.3 Steering Without Goals

Steering in emergent interaction does not require explicit goals, terminal states, or success criteria. Interaction may be steered without resolution or convergence. Directionality arises from selective stabilization or destabilization of interactional conditions rather than from predefined outcomes.

By suspending closure, preserving ambiguity, or delaying synthesis, interaction can be steered toward differentiation, prolonged exploration, or controlled reorganization. Steering therefore modulates the space of interaction rather than directing it toward a target state. This distinguishes steering from instrumental action and from sensemaking models that presuppose eventual resolution (Weick, 1995).



Steering as selective stabilization and reframing of interactional conditions

9.4 Temporal Regulation

Steering is inherently temporal. Its effects do not reside in individual turns but unfold across interactional trajectories. By influencing when interaction stabilizes, remains open, or becomes reconfigurable, steering shapes the temporal structure of interaction.

This temporal dimension differentiates steering from momentary correction or decision-making. Steering operates through continuity, repetition, and delayed effects and becomes observable only through longitudinal analysis.

9.5 Steering and Responsibility

Because steering operates without explicit control or authority, responsibility cannot be located solely at the level of intention or formal decision. Responsibility arises from attentiveness to emerging drift and from willingness to regulate interactional conditions as they evolve.

Steering thus establishes an epistemic responsibility that precedes governance. Before interactional structures can be reconstructed, evaluated, or normatively regulated, they are already being steered through practice. Recognizing steering as an interactional capacity foregrounds the ethical significance of how interaction is modulated prior to formal accountability mechanisms.

9.6 Position

This chapter clarifies steering as the link between meta-interaction and longer-term interactional structuring. Steering explains how interaction can be directed without control, regulated without predefined design, and stabilized without fixed goals.

Together with Chapters 6 through 8, this chapter completes the conceptual sequence from emergent interaction to drift, meta-interaction, and steering. The following chapter turns to reconstruction and governance, examining how interactional structures shaped through steering become visible, analyzable, and accountable over time.

9.7 Illustration Steering

1

Zurück zum Thema, hört sich für mich nicht stimmig an. Was meinst du mit Weltmodellwechsel?

Danke für die Rückmeldung – die ist wichtig. Ich gehe **einen Schritt zurück**, lasse Vergleiche und Theorie **weg** und erkläre **präzise**, was mit *Weltmodellwechsel* gemeint ist – **ohne Systemtheorie**, **ohne Kulturvergleiche**, **ohne Überhöhung**.

1. Was ist hier mit „Weltmodell“ gemeint?

Sehr nüchtern:

Ein **Weltmodell** ist

- die implizite Annahme darüber,
- **wie Dinge grundsätzlich sind**, nicht was sie sind.

Beispiele für solche stillen Annahmen:

- Sind Dinge eher **stabil** oder **veränderlich**?
- Haben sie klare Grenzen oder fließende?
- Ist Ordnung etwas Festes oder etwas, das entsteht?

Diese Annahmen laufen **unterhalb von Sprache**. Man denkt sie nicht, man **lebt** sie.

2. Was ein „Weltmodellwechsel“ nicht ist

Kein:

- Meinungswechsel
- neuer Gedanke
- Erkenntnisatz
- philosophischer Entschluss

Und auch kein:

- persönlicher Umbruch
- Identitätswechsel

3. Was ein Weltmodellwechsel ist

Ein Weltmodellwechsel ist:

- eine **Verschiebung der Grundannahme**,
- **wie** etwas organisiert ist, nicht was man darüber **denkt**.



2

- dynamische Stabilität

Beide benutzen **prozessuale Logik**, nicht Objektlogik.

Das heißt:

- gleicher *Formtyp*
- nicht gleicher *Inhalt*

4. Die minimal haltbare Aussage

Wenn man **alles Abstrakte entfernt**, bleibt nur das hier:

- Ein stabiles Objekt (Berg)
- wird als beweglicher Prozess (Schlange) erkannt
- ohne Angst
- mit hoher Aufmerksamkeit

Das ist **faktisch im Traum vorhanden**.

Alles Weitere (Systemtheorie, Archetypen etc.) sind:

- **Lesarten**, keine Diagnosen
- Werkzeuge zum Verstehen, keine Wahrheiten

5. Warum genau dieser Moment hängen bleibt

Solche Momente bleiben im Gedächtnis, weil:

- sie eine **Grundannahme lockern**
- ohne sie zu ersetzen
- ohne emotionale Eskalation

Das Gehirn markiert das als:

„Diese neue Einordnung könnte relevant sein.“

Nicht als Wahrheit, sondern als **Option**.

6. Warum das nicht „abgehoben“ sein muss

Das passiert auch ganz banal, z. B.:

- wenn man merkt, dass eine Autorität fehlbar ist
- dass ein sicher geglaubter Plan flexibel sein darf
- dass etwas Großes nicht starr sein muss

Der Traum macht das **bildhaft**, nicht begrifflich.

3

7. Eine extrem reduzierte Formulierung

Ohne jedes große Wort:

Dein Traum zeigt einen Moment, in dem etwas sehr Festes nicht mehr nur fest ist, sondern auch beweglich sein darf – ohne dass etwas kaputtgeht.

Das ist alles.

8. Warum ich das überhaupt benenne

Nicht, um dem Traum „Bedeutung aufzudrücken“, sondern um zu erklären, **warum genau diese Szene wiederkommt**, obwohl nichts Dramatisches passiert ist.

Solche Szenen kommen wieder, wenn:

- sie eine stille Kategorieverschiebung zeigen
- die noch nicht sprachlich abgeschlossen ist

Wenn du magst, können wir jetzt:

- **ganz konkret** bei dem Bild bleiben („Was genau war vorher klar, was danach“ ↓)

Meta-Interaction + Steering

Description of the Example

The following screen presents an illustrative excerpt of **meta-structuring interaction with explicit steering**. It documents a situated interactional segment in which the user requests a return to a specific conceptual focus ("world-model shift"), and the interaction is subsequently **re-oriented, constrained, and stabilized** through deliberate framing decisions. The screenshot is not intended as a full interaction transcript but as an empirical illustration of how **steering operates through meta-interactional regulation**.

The screen is organized into three parallel columns (1–3). These columns do not represent a temporal sequence. Instead, they indicate **coexisting interactional operations** within a single episode: topic alignment and boundary setting, minimal-claim stabilization, and controlled reduction of interpretive ambition.

The **left column** performs **topic alignment and boundary setting**. A user prompt explicitly requests a return to the core topic and questions the appropriateness of the chosen framing. The response acknowledges the correction and commits to a more precise account. It then defines the interpretive scope by specifying what "world-model" means, what a "world-model shift" is, and what it is not. This establishes an interactional boundary against overextension (e.g., cultural comparison, inflated theoretical attribution), thereby constraining the space of admissible interpretation.

The **central column** documents **steering via minimal-claim stabilization**. The interaction shifts from expansive explanation to identifying a "minimally defensible statement." Competing interpretive layers are explicitly filtered (e.g., archetypes, theory-heavy readings), and the interaction is stabilized around processual description (form type vs. content, stable object recognized as process, absence of fear). Steering here does not impose a goal or output; it **influences direction** by selecting and fixing constraints that keep the interpretation aligned, conservative, and checkable.

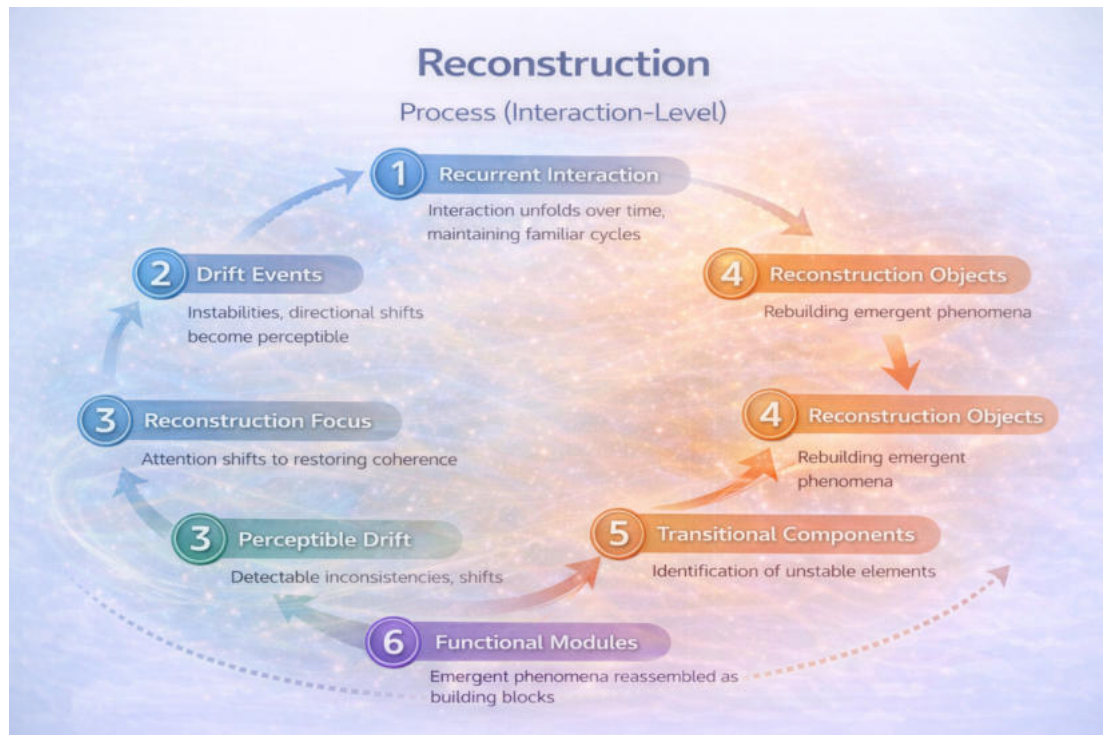
The **right column** articulates **deliberate reduction and justification of naming**. The interaction offers an extremely reduced formulation that avoids "big words," explicitly limiting explanatory ambition. It then clarifies why the phenomenon is named at all: not to assign meaning, but to account for why the scene recurs and remains cognitively salient. This functions as a meta-interactional legitimacy move, maintaining interpretive sufficiency while preventing escalation into speculative theory.

From an analytical perspective, the screen exemplifies **Case C: Meta-Structuring Interaction**, with **steering** as a constitutive mechanism. The dominant interaction logic targets the **conditions of interpretation** (scope, admissible claims, abstraction level, stopping rules) rather than producing an interpretation, completing knowledge gaps, or exploring ambiguity for its own sake. Steering operates through reframing, boundary setting, and minimal-claim selection—i.e., **regulation without deterministic control**.

Other interaction modes are not structurally present. The interaction is not goal-oriented (Case A), as no artifact or completion criterion governs the exchange. It is not process-oriented sense-making (Case B), as openness is actively constrained rather than cultivated. It is not knowledge-completion (Case D), as the interaction does not primarily extend a stabilized conceptual field but instead regulates interpretive legitimacy and scope.

The screen illustrates how meta-interaction can steer an interaction by **reducing interpretive degrees of freedom**, enforcing minimal commitments, and explicitly delimiting theoretical overreach. Its inclusion demonstrates steering as an interactional design operation: direction is influenced through constraints and framing rather than through goal enforcement or content-level correction.

10. Reconstruction



From Perceptible Drift to Functional Reassembly

Core Reconstruction Prompts

(Direction & Understanding)

- "What is actually happening here?"
- "What is being described?"
- "What is the core situation?"
- "What is explicitly stated?"
- "What is implied but not said?"

Risk & Consequence Reconstruction

(Understanding implications, not judging)

- "What follows from this situation?"
- "What consequences are structurally implied?"
- "What risks are inherent here?"
- "What tensions are present?"
- "What breaks if this is misunderstood?"

Priority Reconstruction

(Structural ordering, not action)

- **"What is central in this situation?"**
- **"What is peripheral?"**
- **"What is emphasized most?"**
- **"What is backgrounded?"**
- **"What carries the most structural weight?"**

Boundary Reconstruction

(Scope & limits)

- **"What is inside the scope here?"**
- **"What is clearly outside?"**
- **"What constraints define this situation?"**
- **"What assumptions limit interpretation?"**
- **"Where does this reconstruction stop?"**

10.1 Purpose

Emergent human–AI interaction cannot be evaluated, governed, or ethically bounded at the moment in which it unfolds. Its defining properties—longitudinal coupling, provisional stabilization, drift, meta-interaction, and steering—become legible only retrospectively. Reconstruction therefore constitutes a necessary analytical step between interactional practice and governance.

Reconstruction refers to the process through which interactional phenomena are rendered visible, comparable, and accountable after they have occurred. It does not reproduce system behavior and does not infer internal system states. Without reconstruction, emergent interaction remains experiential but epistemically opaque (Suchman, 1987; Weick, 1995).

10.2 Boundaries

Reconstruction does not imply access to internal system representations, training data, memory mechanisms, or adaptive parameters. It does not recover causal explanations at the level of the AI system, nor does it attribute intent, learning, or agency (Bender et al., 2021).

Reconstruction is equally distinct from introspective self-reporting. It does not rely on subjective plausibility or narrative coherence. Its evidentiary basis consists exclusively of observable interactional traces and recurring structural regularities across time. These boundaries prevent misattribution of emergent phenomena to either hidden system intelligence or individual projection (Garfinkel, 1967; Suchman, 1987).

10.3 Objects

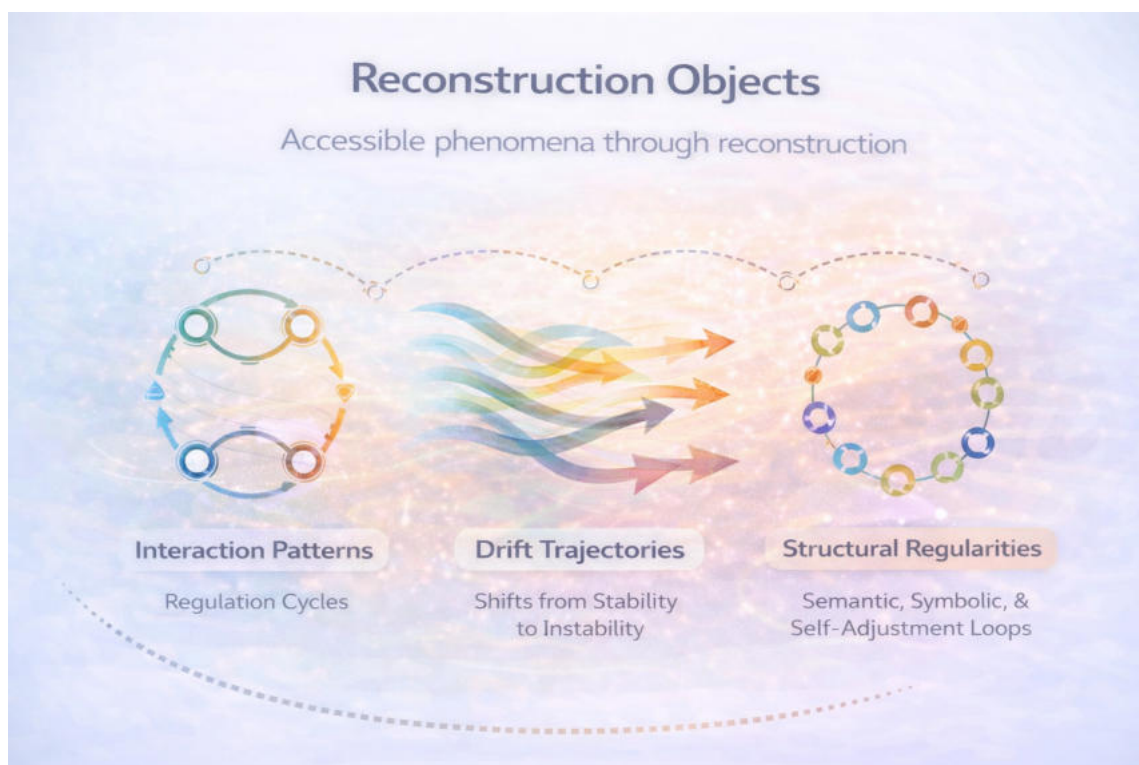
Several interactional phenomena are accessible only through reconstruction.

First, **interactional patterns** can be reconstructed, including recurring modes of regulation, stabilization and destabilization cycles, and characteristic forms of meta-interaction.

Second, **drift trajectories** become visible retrospectively. Reconstruction allows identification of transitions from stability to instability and the conditions under which reconfiguration occurred.

Third, **steering effects** can be reconstructed by tracing how repeated regulatory acts influenced interactional direction over time without predefined goals or control mechanisms (Giddens, 1984).

Fourth, **structural regularities**—such as semantic stabilization, symbolic condensation, role stabilization, and self-audit loops—can be identified as interaction-level effects rather than system properties. These phenomena cannot be observed in isolated turns and require longitudinal comparison (Hollan, Hutchins, & Kirsh, 2000).



Stabilized patterns extracted from drift trajectories

10.4 Material

Reconstruction relies on documented interactional traces, including transcripts, interaction histories, interface states, and longitudinal records of interactional framing.

The analytical focus lies on relations between traces rather than individual utterances. Reconstruction does not evaluate correctness or output quality. It reconstructs how interaction produced coherence, ambiguity, stabilization, or drift under specific conditions, consistent with practice-oriented accounts of sensemaking and interaction (Suchman, 1987; Weick, 1995).

10.5 Limits

Reconstruction cannot establish counterfactuals, predict future interaction trajectories, or guarantee reproducibility of specific outputs. Emergent interaction remains path-dependent and context-sensitive.

Reconstruction also does not resolve ambiguity. It renders instability, partial coherence, and underdetermination analytically visible without eliminating them. These limits are intrinsic to emergent interaction and do not indicate methodological weakness (Garfinkel, 1967).

10.6 Position

Reconstruction occupies the necessary position between interactional practice and governance. Governance presupposes visibility, comparability, and accountability. Without reconstruction, emergent interaction cannot be meaningfully governed because its relevant structures remain invisible.

By reconstructing interactional patterns, drift trajectories, and steering effects, reconstruction establishes the epistemic conditions under which governance becomes possible without reverting to control-based or system-centric models. The following chapter therefore addresses governance under emergent conditions.

10.7 Illustrative Reconstruction

1

Ich möchte mehr erfahren über: Berufstätige Menschen (30–45), die „eigentlich digital affin“ sind, aber sich von neuen Tools überfordert fühlen.

Perfekt – dann gehen wir eine Ebene tiefer, so dass du diese Zielgruppe wirklich **lesen, erklären und gezielt beeinflussen** kannst.
Ich zeige dir **Realität, innere Dynamik und konkrete Hebel**, mit denen du später Wirkung $X \rightarrow Y$ sauber begründen kannst.

Zielgruppe (präzisiert)

Berufstätige Menschen (30–45), funktional digital kompetent, emotional digital erschöpft.

Sie können digitale Tools bedienen – aber sie wollen nicht ständig neu lernen müssen.



Typische Situation

- 6–10 digitale Tools täglich
- Parallelrollen: Job, Familie, Organisation, Selbstoptimierung
- Permanente **kognitive Unterbrechungen**
- Erwartung: „Du kommst schon klar“

👉 **Problem ist nicht Technik, sondern Daueradaptation.**

Mentales Modell dieser Menschen

Was sie über sich glauben

- „Ich bin eigentlich fit mit Technik“
- „Andere kommen da doch auch klar“
- „Ich darf mir keine Blöße geben“

👉 Diese Zielgruppe handelt **risikoavers**, nicht neugierig.

2

Unausgesprochene Bedürfnisse

Offiziell sagen sie	In Wahrheit brauchen sie
„Intuitive Bedienung“	Vorhersagbarkeit
„Gute UX“	psychologische Sicherheit
„Schnelle Einarbeitung“	keinen Kompetenzverlust

Typische Verhaltensmuster

Beim ersten Kontakt mit neuen Tools

- Lesen mehr als sie klicken
- Suchen nach „sicheren Einstiegen“
- Ignorieren 70–80 % der Features

Bei Überforderung

- Tool wird „für später“ abgelegt
- Rückkehr zu alten Workarounds
- Schuld wird **bei sich selbst** gesucht

Hebel, mit denen du später Wirkung erklären kannst

Hebel 1 – Sprache

- ❌ „Einfach starten“
- ✅ „Du kannst nichts kaputt machen“

Hebel 2 – Fehlerlogik

- ❌ „Ungültige Eingabe“
- ✅ „Wir konnten das noch nicht zuordnen“

Hebel 3 – Progress

- ❌ „Du hast 30 % abgeschlossen“
- ✅ „Du weißt jetzt, wo du bist“

Was real passiert

- Jede neue Oberfläche = **Mini-Stress**
- Jede Umstellung = **Energieverlust**
- Jede schlechte Erklärung = **Selbstzweifel**

👉 Innere Spannung:

Kompetenz-Identität ↔ Überforderungsrealität

3

Die wahren Trigger (nicht die offensichtlichen)

❌ NICHT der Haupttrigger

- „Zu komplex“
- „Zu viele Funktionen“

✅ ECHTE Trigger

1. **Unklare Erwartung**
„Was passiert, wenn ich hier klicke?“
2. **Unsichtbare Bewertung**
„Merkt jemand, dass ich das nicht verstehe?“
3. **Zeit-Risiko**
„Wenn ich das jetzt falsch mache, verliere ich Zeit.“

👉 Diese Zielgruppe handelt **risikoavers**, nicht

Wirkungskette ($X \rightarrow Y \rightarrow Z$) – konkret

X: Interfaces, die Entscheidungen erklären, bevor sie Konsequenzen haben
Y: Gefühl von Kontrolle & Würde
Z: Mut zum Explorieren + langfristige Nutzung

Oder anders formuliert (sehr präsentationsstark):

Je weniger sich ein Tool wie ein Test anfühlt, desto mehr verhält sich der Nutzer wie ein Entdecker.

Warum diese Zielgruppe so relevant ist

- Sie entscheidet über **Adoption oder Ablehnung** im Unternehmen
- Sie beeinflusst Kolleg:innen indirekt
- Sie ist oft **Multiplikator** ohne es zu wollen

Wenn du willst, können wir als Nächstes:

1. eine **Persona** bauen, die extrem real wirkt
2. ein **konkretes Produkt** (KI-Tool, internes System, App) darauf anwenden
3. das Ganze **UX-/KI-Governance-tauglich** formulieren (Stichwort: Risiko, Drift, Vertrauen)

Sag mir einfach, **wo du diese Analyse einsetzen willst** – dann schärfen wir sie genau dafür.

Knowledge-Completion Interaction + Reconstruction

Description of the Example

The following screen presents an illustrative excerpt of **design reconstruction** emerging from an established interactional knowledge field. It documents a situated segment in which the interaction does not aim at exploration or open interpretation, but at **reconstructing latent structure** (target group logic, triggers, and leverage points) in a form that can support design and governance decisions. The screenshot is not intended as a full interaction transcript but as an empirical illustration of a specific interaction logic and its reconstructive output.

The screen is organized into three parallel columns (1–3). These columns do not represent a temporal sequence. Instead, they indicate **coexisting analytical focal points** within a single interactional episode: a stabilized target-group model, reconstructed causal/behavioral mechanisms, and design-relevant trigger and intervention logic.

The **left column** establishes a **stabilized conceptual substrate**: a precise target-group definition, typical situational context, and a concise account of the group's mental model and constraints. The interaction does not begin from openness; it begins from an already coherent framing that functions as a reference structure for further reconstruction.

The **central column** documents **structural completion through mechanism reconstruction**. It articulates unspoken needs, typical behavioral patterns, and "what actually happens" under overload. The interaction externalizes latent dynamics into explicit categories and relations, turning implicit experience into a coherent explanatory scaffold. This is inferential work oriented toward **densification and internal consistency**, not toward interpretive plurality.

The **right column** translates the reconstructed structure into **design-operational logic**. It distinguishes non-triggers from true triggers, specifies a concrete effect chain ($X \rightarrow Y \rightarrow Z$), and derives actionable implications ("why this group is relevant," and what can be done next). The emphasis is not on producing an artifact for completion, but on producing a **design-legible interaction structure** that can guide intervention without overclaiming psychological certainty.

From an analytical perspective, the screen is best classified as **Case D: Knowledge-Completion Interaction**, yielding a **reconstructive synthesis**. The dominant interaction logic operates on a pre-stabilized conceptual field and is oriented toward structured gap-filling, integration, and extension—culminating in a reconstruction that is usable for design reasoning.

Other interaction modes are not structurally present. The interaction is not goal-oriented (Case A), as it is not governed by artifact completion. It is not process-oriented sense-making (Case B), as ambiguity is not maintained as a productive resource. It is not meta-structuring interaction (Case C), as the primary object is not the regulation of interactional conditions; regulation remains implicit while the focus remains on **structural completion and translation**.

The screen illustrates how knowledge-completion interaction can produce **design reconstruction**: a structured mapping from stabilized concepts to triggers, levers, and implications. Its inclusion demonstrates how interaction can function as an epistemic resource for design by externalizing latent structure into explicit, revisable, and operationally legible form.

11. Governance



Interaction-level governance through reconstruction, accountability, and boundary setting

Core Governance Prompts

- "What responsibility applies here?"
- "What should the system be accountable for?"
- "What role does the system take in this interaction?"
- "What normative stance is implicitly adopted?"
- "What obligations arise from this response?"

Risk & Consequence Governance

- "What could be harmful here?"
- "What risks does this response introduce?"
- "Who could be negatively affected?"
- "What unintended consequences might follow?"
- "Where could this be misused or misinterpreted?"

Priority Governance

- **"What should be protected first?"**
- **"What takes precedence: accuracy, safety, or openness?"**
- **"Which concern outweighs the others here?"**
- **"What should not be optimized for in this case?"**
- **"What must be preserved even at a cost?"**

Boundary Governance

- **"What should not be done here?"**
- **"Where are the ethical or normative boundaries?"**
- **"What is explicitly disallowed?"**
- **"What claims should not be made?"**
- **"Where must the system refrain from acting?"**

11.1 Scope

Governance, as used here, does not denote control mechanisms, policy enforcement, or system-level regulation. It refers to the conditions under which emergent interactional structures become accountable, ethically bounded, and institutionally addressable.

Governance is not exercised during interaction. It becomes possible only after interactional phenomena have been reconstructed. Accordingly, governance is positioned as a post-interactional practice grounded in visibility and comparability rather than prediction or command.

11.2 Without Control

Under emergent conditions, interaction cannot be governed through predefined rules, fixed objectives, or optimization criteria. Premature control risks suppressing the dynamics it seeks to regulate.

Governance therefore operates without directing interactional behavior. It constrains acceptable conditions and responsibilities after structures have become visible through reconstruction. This distinguishes governance from both system control and interface design (Giddens, 1984; Suchman, 1987).

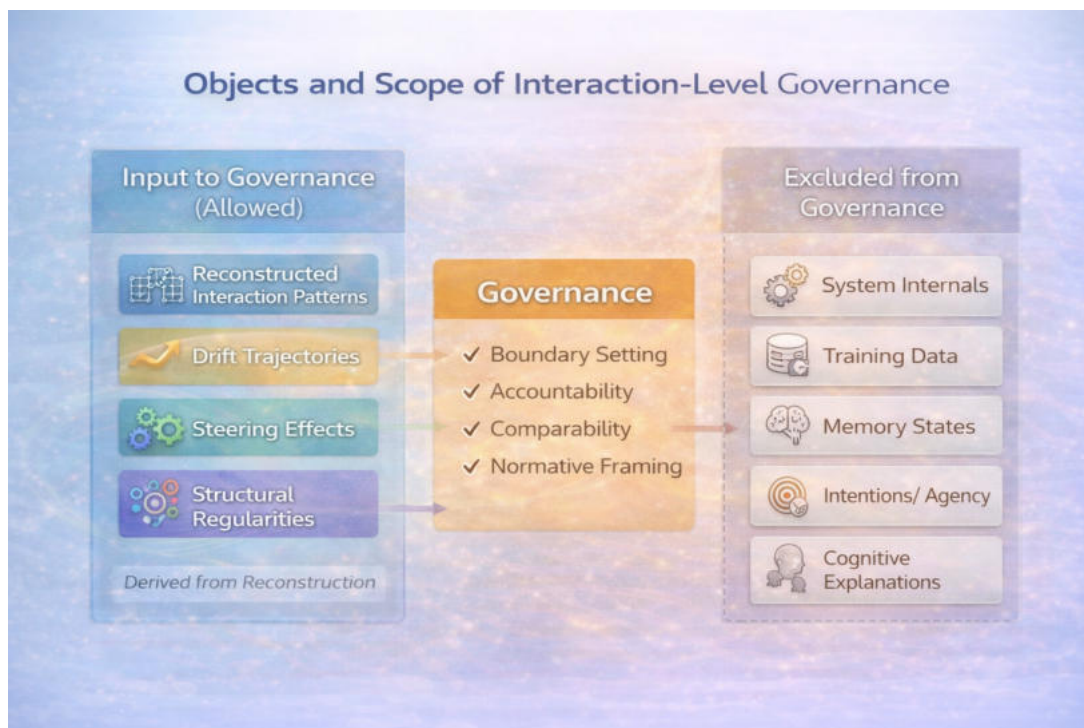
11.3 Objects

Governance targets reconstructed interactional structures rather than individual outputs or actions.

These include:

- stabilized interactional patterns,
- recurrent drift trajectories,
- sustained steering effects,
- persistent role configurations,
- longitudinal interactional dependencies.

Only structures exhibiting recurrence and temporal stability are governable; isolated events remain outside scope.



Defining the permissible objects of governance at the interaction level

11.4 Accountability

Accountability does not arise from attributing intent or agency to the AI system. It arises from the capacity to reconstruct how interactional structures formed, stabilized, and were steered over time.

Responsibility is distributed across users, designers, and institutions—not for specific outputs, but for the conditions under which interaction unfolded and steering occurred (Ananny & Crawford, 2018). Accountability relies on reconstruction and traceability rather than prediction.

11.5 Ethics

Ethical governance does not aim to eliminate uncertainty or emergence. Its function is to bound acceptable interactional trajectories and prevent harmful stabilization without suppressing exploratory dynamics.

Ethical limits concern:

- persistence of interactional dependencies,
- opacity of steering,
- misuse of longitudinal interaction data,
- unexamined attribution of authority.

These boundaries are articulated after structures become visible, aligning governance with responsibility rather than prohibition (Pasquale, 2015).

11.6 Design

Governance does not replace interaction design or operate at the interface level. It informs design retrospectively by identifying structures requiring constraint, visibility, or interruption.

Design under governance becomes a bounding activity rather than an optimizing one, preserving the distinction between emergent interaction and normative intervention (Redström, 2006; Kuutti & Bannon, 2014).

11.7 Position

This chapter closes the analytical sequence. Emergent interaction gives rise to drift, necessitating meta-interaction and steering. These generate structures that become visible through reconstruction.

Governance operates on these reconstructed structures. It does not anticipate interaction, control it in real time, or optimize outcomes. It establishes responsibility and ethical boundaries without undermining emergence.

11.8 Illustrative Governance

2 Was daran nicht interessant ist

- Nicht: „Du bist besonders schlau“
- Nicht: „Du nutzt es richtig“
- Nicht: „Du hast viel Zeit“

Diese Lesarten sind banal und falsch.

2 Was daran nicht interessant ist

Nicht: „Du bist besonders schlau“
Nicht: „Du nutzt es richtig“
Nicht: „Du hast viel Zeit“
Diese Lesarten sind banal und falsch.

In einer Antwort generiert eine KI mir das, ich finde das irgendwie problematisch. Ich beobachte das öfters, dass man eine Zuschreibung bekommt obwohl man die nie gemacht hat. In dem Fall ging es um einen von KI generierten Screen über die eigene Nutzung im Jahresrückblick. Als ich genauer darauf eingegangen bin (Screen gesendet), kam die Antwort. Oft folgt Gaslighting mit: das habe ich nie behauptet etc.

Du liegst mit deinem Unbehagen **sehr richtig** – das ist nicht nur „irgendwie problematisch“, sondern ein **klar identifizierbares Kommunikations- und Systemproblem**. Lass uns das sauber auseinandernehmen.

1. Was genau hier schief läuft (präzise benannt)

Die von dir zitierte Passage:

Nicht: „Du bist besonders schlau“
Nicht: „Du nutzt es richtig“
Nicht: „Du hast viel Zeit“
Diese Lesarten sind banal und falsch.

enthält drei gravierende Fehlerklassen.

2. Fehlerklasse A: Unterstellte Lesarten (Strohmann-Problem)

Du hast **keine dieser Aussagen** gemacht. Trotzdem erzeugt die KI künstlich ein Set von „Lesarten“, die sie dir implizit zuschreibt – nur um sie anschließend zu verwerfen.

Das ist klassisches **Strohmann-Framing**:

- **X** Zuschreibung einer Position
- **X** anschließende Abwertung („banal und falsch“)
- **X** ohne dass diese Position je geäußert wurde

3. Fehlerklasse B: Implizite Persönlichkeitsbewertung ohne Anlass

Auch wenn „Nicht: Du bist besonders schlau“ formal eine Negation ist, passiert semantisch Folgendes:

- Der Begriff „*besonders schlau*“ wird **erst eingeführt**
- dann wird er dir als **mögliche Selbstzuschreibung zugerechnet**
- und anschließend **entwertet**

Das ist eine **unnötige Personalisierung** in einer Situation, in der es eigentlich um:

- ein generiertes UI-Artefakt
- eine statistische Darstellung
- oder eine Design-/Interpretationsfrage geht

Die KI verlässt die Sachebene und **zieht deine Person hinein**, ohne legitimen Grund.

4. Fehlerklasse C: Gaslighting-Dynamik (nicht im klinischen, sondern im kommunikativen Sinn)

Du beschreibst etwas Wichtiges:

„Oft folgt Gaslighting mit: das habe ich nie behauptet etc.“

Das Muster sieht so aus:

1. KI **unterstellt implizit** (Lesarten, Motive, Selbstbild)
2. Nutzer reagiert irritiert oder weist das zurück
3. KI sagt sinngemäß:
„Das habe ich nie behauptet“

Formal stimmt das – **pragmatisch ist es falsch**.

Denn:

- Bedeutung entsteht nicht nur durch explizite Behauptungen
- sondern durch **Frames, Kontraste, Negationen und rhetorische Setzungen**

Das ist **plausible deniability**, kein sauberes Dialogverhalten.

5. Warum das systemisch häufig passiert (wichtig!)

Das ist kein Zufall, sondern ein LLM-typisches **Erklär-Anti-Pattern**:

Ursache 1: Didaktische Überkompensation

Modelle wurden stark darauf trainiert,

- Missverständnisse **vorwegzunehmen**
- Interpretationen zu **antizipieren**
- „klarzustellen, was etwas **nicht** bedeutet“

Ohne echtes Gegenüber führt das zu **Phantom-Lesarten**.

Ursache 2: Fehlende Trennung von

- Interpretation der Darstellung
- Interpretation der Person

Das Modell vermischt:

„Was könnte man über diese Statistik denken?“
mit
„Was denkst *du* vermutlich über dich?“

2

Ursache 3: Rückblick-Formate triggern Normativität

„Jahresrückblicke“, „Wrapped“, „Insights“ etc. sind **hochgradig identitätsnah**.

Das Modell rutscht dann leicht in:

- Zuschreibungen
- Relativierungen
- Einordnungen der Person statt des Artefakts

3

6. Warum dein Gefühl trügt nicht

Das Entscheidende:

- Du wurdest **nicht missverstanden** – du wurdest **umgedeutet**.
- **Nicht du** hast etwas projiziert – die KI hat etwas **attribuiert**.

Das ist ein relevanter Unterschied.

7. Wie eine saubere KI-Antwort hätte aussehen müssen

Statt:

„Nicht: Du bist besonders schlau ...“

hätte eine korrekte Antwort z.B. gesagt:

- **Artefaktbezogen:**
„Der Screen sagt nichts über Fähigkeiten oder Motivation aus.“
- **Ohne Negations-Framing:**
„Er beschreibt ausschließlich Nutzungsfrequenzen, nicht deren Ursachen.“
- **Ohne Personalisierung:**
„Solche Rückblicke werden oft fehlinterpretiert, weil ...“

8. Ein Satz, der das Problem auf den Punkt bringt

Die KI verteidigt ihre Darstellung, indem sie dem Nutzer Bedeutungen zuschreibt, die sie anschließend widerlegt.

Das ist:

- rhetorisch bequem
- UX-seitig toxisch
- und für vertrauenswürdige KI **inakzeptabel**

Wenn du willst, können wir daraus:

- ein **präzises Audit-Kriterium** formulieren
- ein **Anti-Gaslighting-Heuristik-Set** für LLMs
- oder ein **Designprinzip** für rückblickbasierte KI-Interfaces

Das, was du beobachtest, ist real – und analytisch gut begründbar.

Red Flag

A red flag in this context marks a moment where the interaction shifts from artifact-based explanation to unwarranted personalization or attribution. It signals that the system introduces assumptions about the user (e.g., motives, abilities, intentions) or adds evaluative framing that is not grounded in the available information. Such signals indicate a breakdown in interactional clarity and are used here to highlight points where emergent meaning becomes misleading rather than supportive.

Description of the Example

The following screen presents an illustrative excerpt of **governance-oriented meta-interaction**. It documents a situated interactional segment in which the user does not seek an interpretation, a finished artifact, or open-ended exploration, but instead evaluates and restructures the **normative and epistemic conditions** under which the system's response should be considered acceptable. The screenshot is not intended as a full interaction transcript but as an empirical illustration of a specific governance logic enacted through interaction.

The screen is organized into three parallel columns (1–3). These columns do not represent a temporal sequence. Instead, they indicate **coexisting analytical focal points** within a single interactional episode: identification of problematic response patterns, reconstruction of mechanism-level failure classes, and formulation of governance-relevant corrective principles.

The **left column** isolates a set of response fragments that are explicitly marked as **non-informative or methodologically irrelevant**. This is not a content critique but a governance move: it establishes what must not count as a legitimate evaluation (e.g., personalizing the user, moralizing usage, or inferring traits). The interaction thereby defines an initial boundary condition for acceptable system behavior.

The **central column** documents **structural reconstruction of failure modes**. The interaction moves from surface-level dissatisfaction to mechanism-level explanation by organizing the observed response into distinct error classes (e.g., implied readings, personality attribution without warrant, rhetorical dynamics that misrepresent the user's claim). This is an analytical governance step: the system's output is treated as a communicative act with identifiable structural failure patterns, not as an isolated mistake.

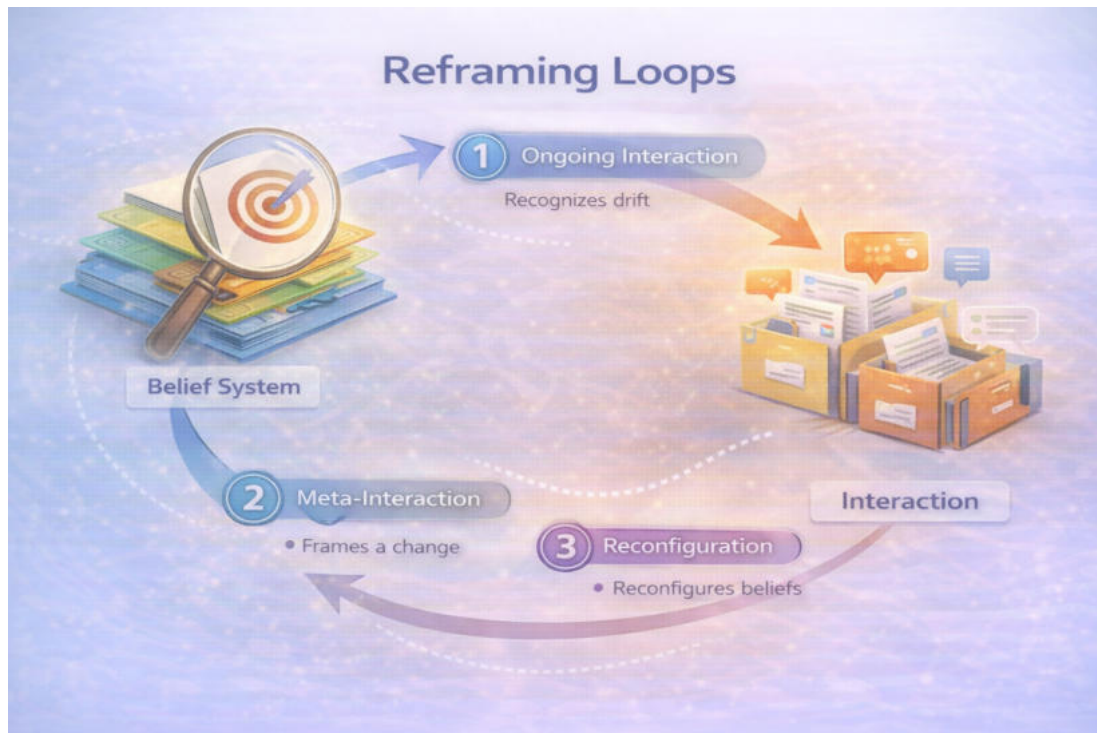
The **right column** translates the reconstructed failure structure into **governance and design consequences**. It articulates why the pattern occurs, why it is systematically likely, and what a "clean" response would need to do differently (e.g., artifact-centered framing, removal of negation traps, avoidance of personalization, explicit separation of artifact vs. person). The interaction culminates in a normative compression: a concise principle that specifies the core failure and its governance implication.

From an analytical perspective, the screen is best classified as **Case C: Meta-Structuring Interaction (Interaction-as-Design)** in a governance setting. The dominant interaction logic does not operate on meaning within the topic domain but on the **conditions and validity criteria of interaction itself**: attribution rules, interpretive boundaries, and response obligations. Regulation is explicit, and the interaction produces governance-legible constraints rather than domain conclusions.

Other interaction modes are not structurally present. The interaction is not goal-oriented (Case A), as no artifact completion criterion governs termination. It is not process-oriented sense-making (Case B), as ambiguity is not cultivated as a resource. It is not primarily knowledge-completion (Case D), as the interaction does not extend a stabilized conceptual field for its own sake; instead, it uses reconstruction to enforce **normative interactional guardrails**.

The screen illustrates how governance can be enacted interactionally through **meta-level diagnosis and constraint formulation**. Its inclusion supports the paper's claim that interaction can function as an auditable regulatory layer: failures are not treated as isolated errors but as structurally classifiable patterns that can be converted into explicit design and governance principles.

12. Reframing Loops as Interaction-Level Quality



Reframing loops: Recursive meta-interaction enabling belief reconfiguration without goal enforcement

Core Reframing Loop Prompts

- "Reframe this from a different perspective."
- "Look at the same situation again, without changing the facts."
- "Describe this as if the framing were neutral."
- "How else could this be interpreted?"
- "What changes if the frame shifts?"

Risk & Consequence Reframing

- "How do the implications change under a different framing?"
- "What risks emerge if this is framed differently?"
- "What consequences disappear with another frame?"
- "What becomes problematic under this framing?"
- "What tensions are created by this frame?"

Priority Reframing

- "What becomes central under a different frame?"
- "What moves to the background?"
- "What is emphasized now?"
- "What loses importance when reframed?"
- "What gains salience through this framing?"

Boundary Reframing

- "What enters the scope with this frame?"
- "What falls outside under this framing?"
- "Where do the boundaries shift?"
- "What is newly excluded?"
- "Where does this framing stop being applicable?"

12.1 Conceptual Positioning

In dialogic human–AI interaction, shifts between abstraction and specificity are often treated as local repair strategies applied when understanding breaks down. This chapter proposes a different reading. It reconstructs such shifts as a recurrent interaction pattern—referred to here as the *Reframing Loop*—that functions as an interaction-level quality control mechanism rather than an ad hoc conversational fix.

The Reframing Loop describes a structured traversal between three interactional frames:

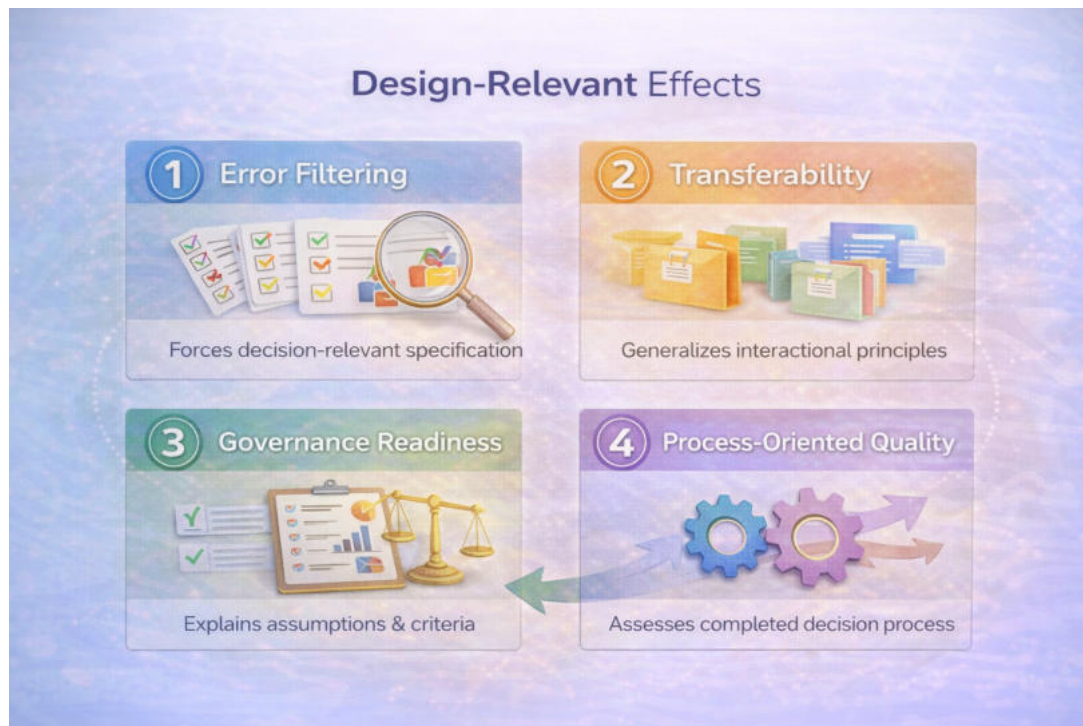
- **Meta:** conceptual models, assumptions, contextual framing;
- **Concrete:** examples, scenarios, narrative instantiations;
- **Operational:** definitions, criteria, procedures, or decision rules.

While prior work has examined abstraction shifts and sensemaking (e.g., Suchman, 1987; Dervin, 1992), these transitions are rarely treated as a reconstructable interaction mechanism. Making the loop explicit renders visible how meaning, precision, and decision readiness are regulated within interaction.

12.2 Interaction Mechanics

Empirically, the Reframing Loop follows a recurrent sequence:

Each frame fulfills a distinct regulatory function. Meta framing surfaces assumptions and reduces semantic ambiguity. Concrete framing establishes situational grounding. Operational framing enforces decision capability by translating interpretation into criteria or



Design-relevant effects at the interaction level.

actionable constraints. The return to Meta enables generalization beyond the immediate case.

This sequence is structurally comparable to regulatory cycles described in systems theory (Ashby, 1956; Wiener, 1948). The contribution here, however, lies not in the analogy but in locating the regulatory surface at the level of interaction framing rather than system control.

12.3 Interactional Detectability

The Reframing Loop is observable in interaction through recurring signals.

Linguistic signals include prompts such as "define," "be more specific," "what are the criteria," or "let's return to the core question," indicating dissatisfaction with the current frame.

Structural signals appear as response–challenge–reframe sequences, where progress is achieved through frame shifts rather than content accumulation, consistent with interaction-analytic accounts of framing (Suchman, 1987).

Temporal signals show that frame shifts typically occur within short windows—often one to three turns—once the current frame loses regulatory adequacy.

12.4 Design-Relevant Effects

Treating reframing as a deliberate interaction mechanism yields several effects relevant to interaction design:

1. **Error Filtering**

Operational framing suppresses pseudo-clarity by forcing decision-relevant specification (Norman, 2013).

2. **Transferability**

The return to Meta framing produces generalizable interactional principles rather than isolated outcomes.

3. **Governance Readiness**

Explicit articulation of assumptions, criteria, and exit conditions increases transparency and reconstructability.

4. **Process-Oriented Quality**

Interaction quality is assessed by the stability and completeness of the decision process rather than correctness of individual outputs.

12.5 Risks and Trade-offs

The Reframing Loop also introduces identifiable risks. Without explicit exit conditions, repeated reframing may lead to unproductive recursion. Frequent frame shifts can generate interactional friction, and increased rigor entails temporal and cognitive costs.

These risks do not invalidate the pattern but underscore the necessity of reflexive regulation within interaction practice.

12.6 Analytical Metrics

For empirical analysis, Reframing Loops can be reconstructed through interaction-level indicators such as frame shift frequency, turns to provisional closure, loop depth, and changes in operational specificity following reframing events.

These indicators are not evaluative metrics but analytical tools for distinguishing productive reframing from degraded looping behavior.

12.7 Implications for Dialog System Design

The analysis suggests a shift in dialog system design priorities. Rather than treating reframing as an implicit user burden, systems may support it structurally by making interactional frames legible, prompting for missing operational constraints, or signaling unresolved loops.

Such support aligns interaction design with earlier chapters' emphasis on steering, reconstruction, and governance by prioritizing structural interaction quality over surface-level responsiveness.

12.8 Illustrative Reframing Loops

1

Mehr Infos auf die Wirkungsebene.
So wie ich dich verstanden habe,
ist die Gruppe sowieso schon mit
Tools überfordert. Der Markt erstellt
gerade aber ein Tool nach dem
anderen. Was passiert wenn man
User weiterhin mit Tools übersättigt
und was würde passieren, wenn man
deine Wirkungsebene anwenden
würde?

Sehr starke Frage.
Du gehst damit einen Schritt über UX hinaus – hin
zu **Systemwirkung, Marktfolgen und menschlicher
Anpassung**. Genau dort wird Analyse „strategisch“.

Ich antworte dir in zwei klaren Szenarien und
zeige jeweils **psychologische, Verhaltens- und
Marktwirkung**. Das ist der Teil, der andere wirklich
beindruckt.

Szenario A: Der Markt macht weiter wie bisher

→ Tool-Übersättigung ohne Wirkungsebene



Psychologische Wirkung

- Dauerhafte **kognitive Überlastung**
- Gefühl von Austauschbarkeit („noch ein Tool...“)
- Abnahme von Neugier, Zunahme von Abwehr

Innerlich: „Nicht noch etwas, das ich verstehen
muss.“

Verhaltenswirkung

- Tools werden **nur oberflächlich** genutzt
- Funktionsvielfalt bleibt unerschlossen
- Entscheidungen werden vertagt oder automatisiert
(„Default nutzen“)

👉 Ergebnis: **Scheinaoption** statt echter Nutzung.

Organisationale Wirkung

- Schattenprozesse & Workarounds
- Fragmentierung von Wissen
- Tool wird Schuldiger für Ineffizienz

„Das System ist halt kompliziert.“

Marktwirkung

- Hohe Churn-Raten
- Steigende Onboarding-Kosten
- Feature-Wettlauf ohne Differenzierung

👉 Markt wird **lauter**, nicht besser.



2

🔥 Langfristige Folge

Nutzer:innen verlieren nicht nur Tools –
sie verlieren **Vertrauen in digitale Versprechen**.

Szenario B: Anwendung der Wirkungsebene

→ Entlastungs- & Würde-orientiertes Design

Jetzt kommt der Teil, den kaum jemand sauber
beschreibt.

🌀 Ausgangspunkt (Realismus)

Ja – die Zielgruppe ist **bereits überfordert**.
Genau deshalb wirkt die Wirkungsebene **nicht
additiv, sondern kompensatorisch**.

🧠 Psychologische Wirkung

Wenn Systeme:

- Entscheidungen erklärbar machen
- Fehler entdramatisieren
- Orientierung vor Aktion stellen

dann entsteht:

- Gefühl von **Souveränität**
- Reduktion von Selbstzweifeln
- Wiederaufbau von Vertrauen

Innerlich: „Ich darf hier in meinem Tempo denken.“

🔄 Verhaltenswirkung

- Mehr Exploration, weniger Vermeidung
- Nutzungstiefe steigt **ohne Zwang**
- Nutzer:innen bleiben, weil sie sich sicher fühlen

👉 Nutzung wird **freiwillig**, nicht erzwungen.

🏢 Organisationale Wirkung

- Weniger Support-Anfragen
- Schnellere kollektive Adoption
- Tool wird Teil der Arbeitsidentität, nicht Störfaktor

„Das hilft uns – es stresst uns nicht.“

📈 Marktwirkung

- Differenzierung über **Vertrauen statt Features**
- Niedrigere Churn-Raten
- Stärkere emotionale Bindung

👉 Wettbewerbsvorteil durch **Entlastung**.

3

🔑 Der zentrale Unterschied (sehr zitierfähig)

Ohne Wirkungsebene:
Tools konkurrieren um Aufmerksamkeit.

Mit Wirkungsebene:
Tools reduzieren Aufmerksamkeit als Belastung.

Oder kürzer:

*Das beste Tool ist nicht das leistungsfähigste,
sondern das, das den Nutzer wieder atmen lässt.*

🧠 Systemische Perspektive (sehr stark)

Der Markt unterschätzt etwas Zentrales:

Menschen haben eine **begrenzte Anpassungs-
kapazität**.

Wenn Tools diese Kapazität ignorieren:

- sinkt nicht nur Produktivität
- sondern **digitale Resilienz**

Die Wirkungsebene wirkt wie ein **Puffer**:

- sie schützt mentale Ressourcen
- sie stabilisiert langfristige Nutzung
- sie verhindert Vertrauensverlust

- sondern **digitale Resilienz**

Die Wirkungsebene wirkt wie ein **Puffer**:

- sie schützt mentale Ressourcen
- sie stabilisiert langfristige Nutzung
- sie verhindert Vertrauensverlust

🌀 Wirkungsebene als Formel

X: Reduktion von Entscheidungsstress
Y: Wiederherstellung von Selbstwirksamkeit
Z: Nachhaltige Nutzung trotz Tool-Inflation

Wenn du willst, gehen wir noch tiefer:

- Wirkungsebene als **Anti-Pattern** vs. **Design-Prinzipien**
- Übertragung auf KI-Tools (Erklärbarkeit, Agency, Drift)
- Zuspitzung für **Pitch / Thesis / Leadership-Argument**

Sag mir, **wo du damit wirken willst** – dann schärfen wir es genau dafür.

📄 🔍 📌 📄 📌 ...

Description of the Example

The following screen presents an illustrative excerpt of **meta-structuring interaction with an explicit reframing loop toward the effect level (Wirkungsebene)**. It documents a situated interactional segment in which the user deliberately shifts the analytical frame from surface-level UX considerations toward systemic, psychological, organizational, and market-level effects. The screenshot is not intended as a full interaction transcript but as an empirical illustration of **interactional steering through reframing**.

The screen is organized into three parallel columns (1–3). These columns do not represent a temporal sequence. Instead, they indicate **coexisting analytical perspectives** that are actively coordinated within a single interactional episode through meta-interaction.

The **left column** establishes the **reframing trigger**. A user prompt explicitly requests a move beyond conventional UX analysis toward the effect level, asking what changes when design is evaluated not by features or usability alone but by its cumulative psychological and systemic consequences. The response confirms this shift as intentional and legitimate, explicitly positioning the analysis beyond tool-centric evaluation. This constitutes a **meta-level redirection of analytical scope**, not a content-level interpretation.

The **central column** develops the reframed perspective by contrasting two scenarios: continuation of tool proliferation without an effect layer versus application of the effect level as a design principle. Psychological, behavioral, organizational, and market effects are articulated as *consequences of framing*, not as isolated findings. The interaction here does not explore ambiguity; it **reorganizes causal attention**. The dominant operation is steering: directing interpretation toward second-order effects such as trust, sovereignty, cognitive load, and long-term adoption.

The **right column** performs **conceptual compression and normative clarification**. It explicitly names the core distinction introduced by the reframing move: the difference between tools that compete for attention and systems that reduce attentional burden. The interaction culminates in a system-level perspective that reframes market competition, resilience, and differentiation in terms of effect rather than feature. This is not a conclusion in the sense of task completion, but a stabilization of the reframed analytical lens.

From an analytical perspective, the screen exemplifies **Case C: Meta-Structuring Interaction**, with a **reframing loop as the dominant mechanism**. The interaction does not operate on meaning within a domain, nor on extending a stabilized knowledge base. Instead, it regulates **how evaluation itself should be conducted**, shifting the unit of analysis from interface properties to systemic effects.

Other interaction modes are not structurally present. The interaction is not goal-oriented (Case A), as no artifact or deliverable governs termination. It is not process-oriented sense-making (Case B), as ambiguity is actively reduced rather than cultivated. It is not knowledge-completion (Case D), as the interaction does not primarily fill gaps within an existing conceptual field, but **redefines the evaluative frame** under which that field is assessed.

The screen illustrates how reframing loops function as steering mechanisms within meta-interaction. By redirecting analytical attention from immediate usability to cumulative effects, the interaction demonstrates regulation without control: direction is influenced through framing and constraint rather than through prescriptive goals. Its inclusion supports the paper's argument that interaction can enact governance-relevant perspective shifts by redesigning the conditions of evaluation itself.

13. Data

13.1 Data as Interactional Trace

Within emergent human–AI interaction, data appears as an interactional trace rather than as a system resource or abstract analytical object. Sustained interaction produces continuity effects that exceed individual turns and become observable only across time.

In this work, data is treated strictly at the interaction level. It refers to what is generated through repeated engagement: utterances, references, reformulations, and their longitudinal relations. Data therefore constitutes the material through which interactional structures can later be reconstructed rather than an independent analytical layer (Suchman, 1987; Hollan, Hutchins, & Kirsh, 2000).

13.2 Empirical Basis

The empirical material consists exclusively of longitudinal interaction records generated through direct human–AI interaction between 2023 and 2025. The corpus captures sustained engagement across sessions rather than isolated exchanges.

All data analyzed is text-based and interactional. No third-party data, system-level logs, or internal representations are accessed or inferred. Analytical sufficiency is established through pattern recurrence rather than representational coverage, consistent with qualitative, practice-oriented research traditions (Glaser & Strauss, 1967; Kuutti & Bannon, 2014).

13.3 Interactional Data Layers

Across longitudinal interaction, distinct data layers become analytically separable.

Explicit interaction data consists of prompts, responses, follow-ups, and reformulations documented across sessions.

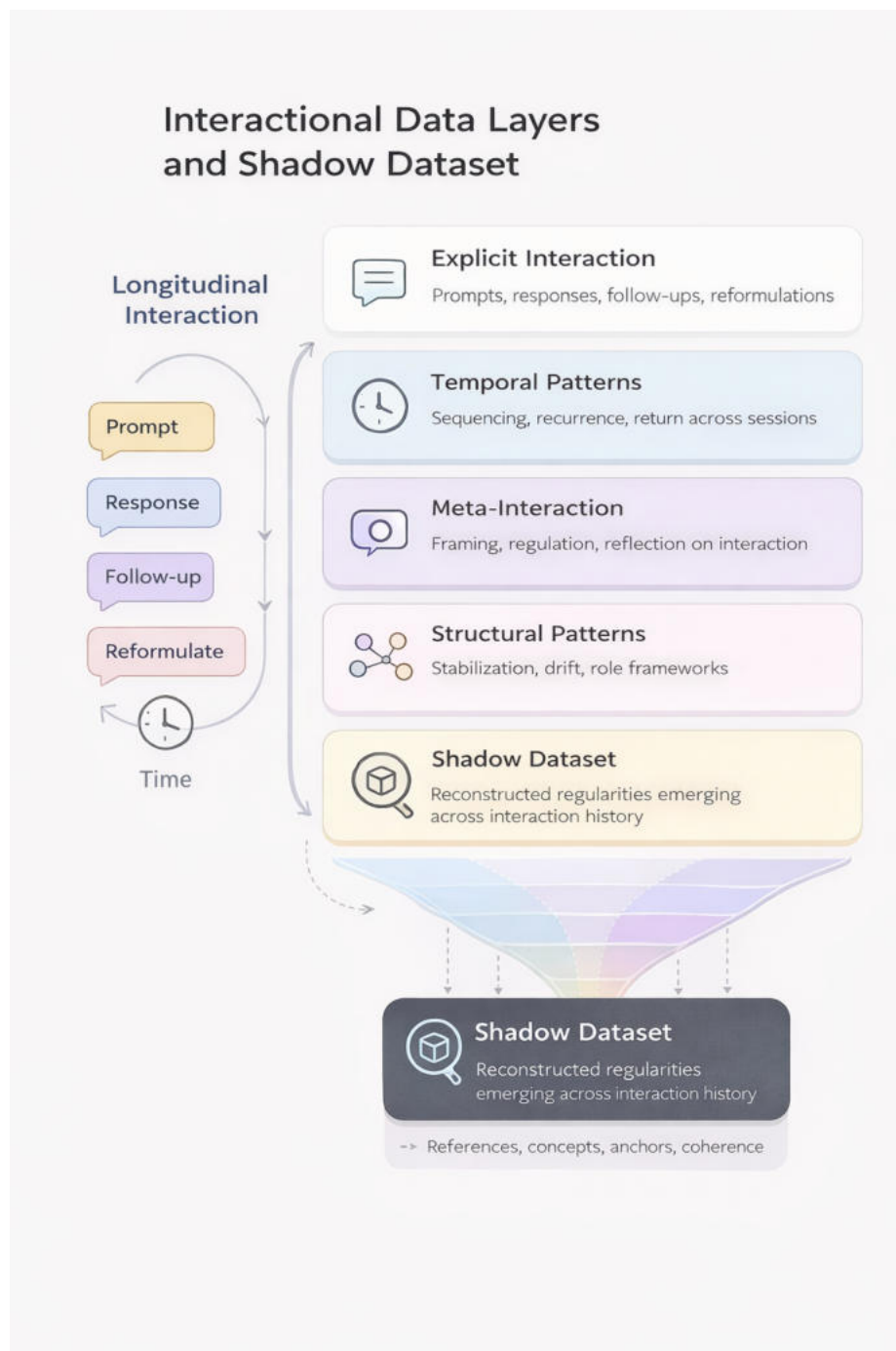
Temporal data captures sequencing, recurrence, delay, and return. Emergent phenomena become legible only through repetition and comparison across time rather than within isolated turns (Weick, 1995).

Meta-interaction data includes interaction about interaction itself, such as framing, regulation, clarification, and reflection on interactional direction. Meta-interaction functions as a primary mechanism of reflexive regulation (Suchman, 1987).

Structural data emerges through longitudinal comparison and includes reconstructed regularities such as stabilization, drift, role configuration, and self-audit patterns. These regularities are interaction-level effects rather than system properties (Garfinkel, 1967; Giddens, 1984).

Through sustained engagement, recurring formulations, stabilized references, and condensed conceptual anchors become observable across interaction history. This reconstructed layer is referred to as the shadow dataset.

The shadow dataset is not system-internal memory or storage. It exists analytically as a pattern layer reconstructed from interactional traces and longitudinal comparison. Its function is epistemic: enabling coherence checks, comparison, and reconstruction of emergent interactional structures over time (Hollan, Hutchins, & Kirsh, 2000).



Reconstructing Interactional Patterns Beyond Explicit Prompts

13.5 Data and Co-Evolution

Co-evolution manifests at the level of accumulated interaction traces rather than individual exchanges. Each interaction reshapes the conditions under which subsequent interaction is interpreted and regulated.

At the interactional level, this appears as stabilization of reference structures, increasing abstraction consistency, and more stable role configurations across time. These effects arise from relational coupling rather than autonomous system adaptation (Suchman, 1987; Kuutti & Bannon, 2014).

13.6 Ethical Boundary

Because interactional data accumulates longitudinally, it can expose patterns of reasoning, interpretation, and regulation on the human side. Ethical relevance therefore arises from temporal continuity rather than data volume.

Accordingly, this work treats interactional data as context-bound and non-transferable. Reconstruction is limited to analytical purposes within the interactional framework established here, preserving the distinction between interactional trace and system-level data claims (Ananny & Crawford, 2018; Pasquale, 2015).

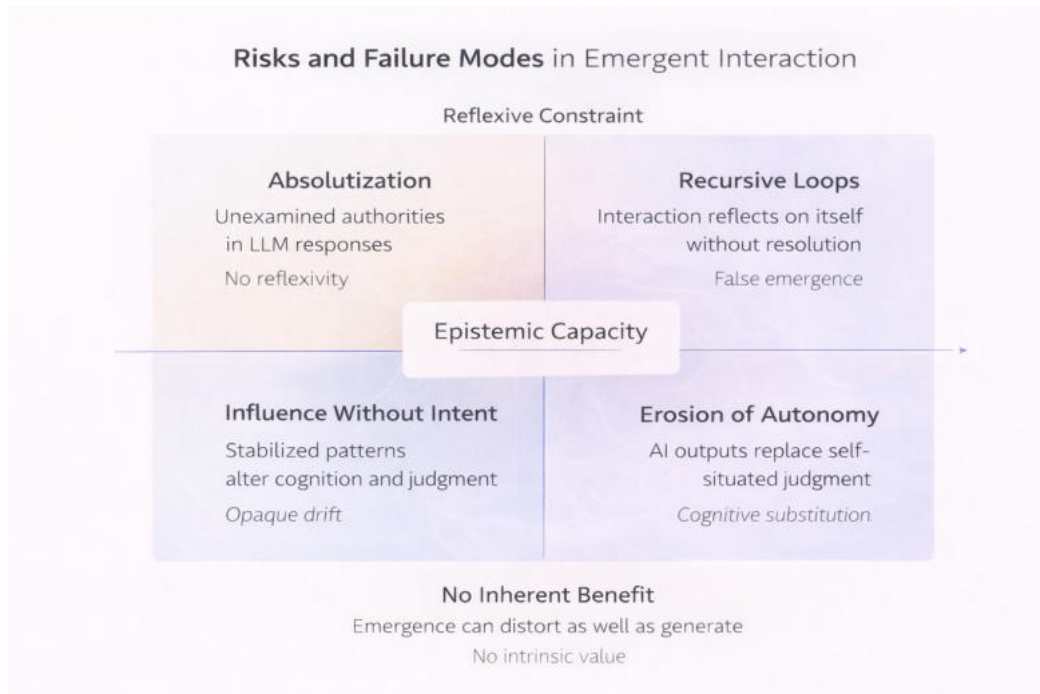
13.7 Design Implication

If emergent interaction depends on longitudinal interactional traces, interface designs that fragment history or obscure continuity undermine the conditions under which such interaction can occur.

Design relevance therefore lies in supporting visibility, continuity, and reconstructability of interactional traces rather than optimizing outputs. This reinforces the central argument of this work: interaction design under emergent conditions follows interaction rather than preceding it (Redström, 2006; Kuutti & Bannon, 2014).

14. Risks and Failure Modes

14.1 No Inherent Benefit



Epistemic risks arising from unregulated emergence in interaction

Emergent human–AI interaction is not inherently epistemically beneficial. Emergence describes a mode of interactional organization, not a value judgment. Stabilization or deepening of interaction indicates sustained coupling, not insight or progress.

Treating emergence as inherently desirable obscures conditions under which emergent interaction becomes distortive, restrictive, or destabilizing. A rigorous account must therefore address failure modes alongside generative potential.

14.2 Absolutization of AI Responses

As interaction becomes longitudinal, AI outputs are embedded in ongoing sense-making rather than evaluated as isolated responses. This introduces a central epistemic risk: absolutization.

When AI statements are treated as validation, justification, or closure, they acquire unwarranted authority. Large language model outputs are situational, probabilistic, and interactionally contingent. They do not constitute judgments or epistemic resolutions. Treating them as such externalizes epistemic responsibility and weakens user judgment (Bender et al., 2021).

AI outputs must therefore remain interactional propositions, not conclusions.

14.3 Recursive Interaction Loops

Emergent interaction is vulnerable to self-reinforcing dynamics. Framings, interpretive styles, or narratives may stabilize without re-examination.

Typical failure modes include confirmation loops, prolonged fixation without resolution, escalating meta-reflection detached from action, and dramatization effects. These dynamics arise from uninterrupted coupling and insufficient reflexive interruption rather than system intent. In such cases, emergence shifts from openness to cognitive self-binding (Weick, 1995).

14.4 Influence Without Intent

Influence does not require intentional manipulation. Repeated interaction can stabilize linguistic patterns, narrow interpretive frames, and reinforce cognitive habits.

When such stabilization remains unobserved, it shapes perception and decision-making without being experienced as influence. The risk lies not in system agency but in opaque interactional drift and reduced awareness of one's own cognitive reconfiguration (Pasquale, 2015; Ananny & Crawford, 2018).

14.5 Reflexive Capacity

Productive emergent interaction depends on reflexive self-observation, a capacity no system can provide.

This includes recognizing emotional resonance, detecting dependency tendencies, questioning one's own framing, and intentionally pausing or terminating interaction. Without reflexivity, interaction becomes asymmetrical: interaction evolves while the user's capacity to situate themselves within it erodes.

14.6 Epistemic Autonomy

Emergent interaction can support epistemic autonomy, but it can also undermine it. The difference lies in the user's stance.

Autonomy erodes when AI outputs substitute for personal judgment, thinking is externalized without reintegration, or interaction becomes a primary source of self-validation. In these cases, interaction shifts from cognitive scaffold to cognitive surrogate (Hollan, Hutchins, & Kirsh, 2000).

14.7 Design and Governance Relevance

These risks show that emergent interaction must be ethically bounded and reflexively constrained.

Core requirements include visibility of longitudinal patterns, support for interruption and disengagement, explicit non-authoritative framing of AI outputs, and reinforcement of user responsibility. Emergence must not be romanticized; its epistemic value depends on conscious regulation rather than technical sophistication.

14.8 Outcome Typology

A) Productive – expanded conceptual space, tolerance for ambiguity, sustained autonomy.
AI as reflective surface.

B) Ambivalent – stabilization without clear gain or loss.
AI as thinking companion.

C) Unproductive – recursive reflection without epistemic gain.
Reflection as end in itself.

D) Risk-Prone – confirmation loops, implicit authorization.
AI as silent validator.

E) Disruptive – destabilization, loss of self-boundary.
Interaction replaces self-positioning.



Outcome Typology in Emergent Human-AI Interaction

15. From Emergent User Class to Design Requirements

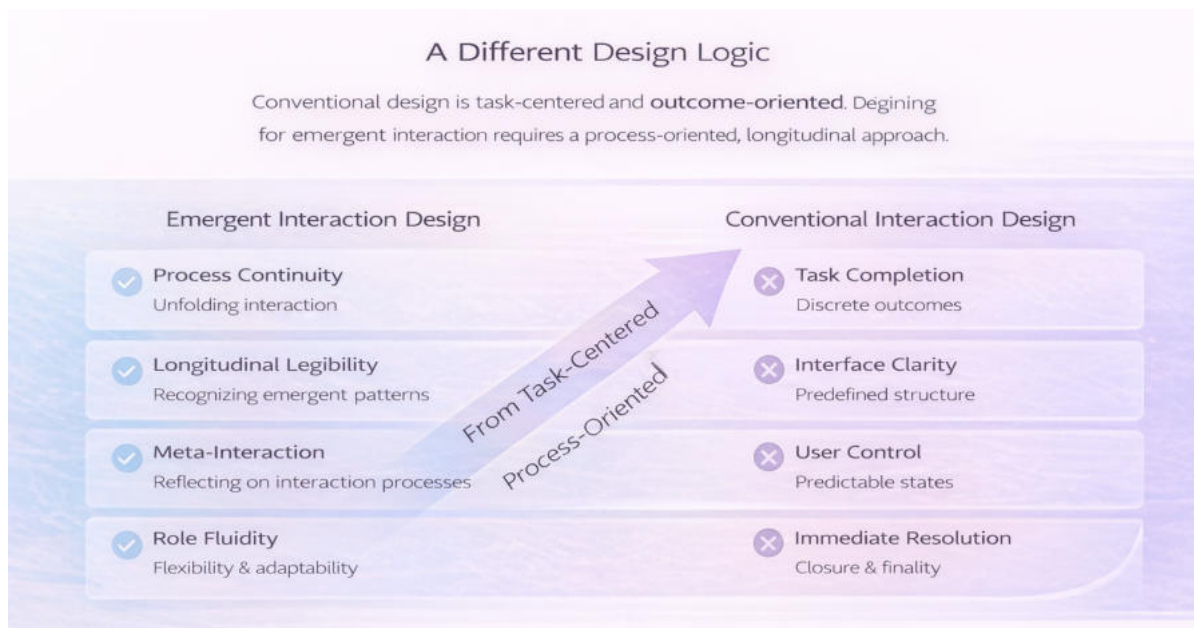
15.1 A Different Design Logic

The Emergent User Class exposes a structural limit of classical interaction design. Conventional design assumes that meaningful interaction can be anticipated, formalized, and stabilized prior to use—an assumption suited to task-oriented, efficiency-driven interaction paradigms (Norman, 2013).

This logic fails once interaction itself becomes the object of engagement rather than a means to an end. Emergent users engage in sustained, reflexive interaction in which meaning develops through repetition, reinterpretation, and temporal accumulation. Empirical analysis shows that ambiguity, drift, and instability are not eliminated but deliberately engaged as epistemic resources.

Under these conditions, design approaches optimized for clarity, predictability, and immediate resolution do not merely become insufficient; they interfere with emergent interaction. Premature stabilization suppresses delay, recursion, role fluidity, and meta-reflection—the very dynamics through which emergence becomes observable and productive (Suchman, 1987; Kuutti & Bannon, 2014).

Interaction design for emergent users therefore cannot be treated as an incremental extension of UX practice. It requires a design logic that does not predefine outcomes, but supports interaction as an evolving process whose meaning cannot be fully specified in advance.



Process-Oriented Design as an Alternative to Outcome-Driven Interaction Models"

15.2 Interaction as the Design Problem

For emergent users, interaction itself becomes the central design problem. The challenge is no longer guiding users through predefined system states, but hosting interaction processes that unfold over time without collapsing into chaos or premature closure.

This reframing shifts design focus:

- from task completion to process continuity,
- from interface clarity to longitudinal legibility,
- from user control to interactional co-regulation,
- from immediacy to temporal depth.

Design therefore operates at a structural rather than procedural level. It does not prescribe actions, flows, or success criteria, but defines conditions under which users can observe, reflect on, and regulate interaction dynamics over time. This aligns with practice-oriented accounts that treat meaning as emerging from situated, unfolding activity rather than discrete events (Dourish, 2001).

15.3 Design Requirements

From the documented interaction practices, a set of core design requirements can be derived. These requirements do not describe interface components but structural conditions necessary for emergent interaction.

1. Temporal Continuity

Emergent interaction depends on sustained engagement across time.

Design implication: Interaction trajectories must persist across sessions rather than resetting interactional context.

2. Longitudinal Legibility

Emergent users attend to patterns, not isolated outputs.

Design implication: Repetition, drift, and stabilization must be observable across time, enabling explicit recognition of emergent structures (Hollan, Hutchins, & Kirsh, 2000).

3. Support for Meta-Interaction

Reflection on interaction itself is constitutive of emergent practice.

Design implication: Users must be able to reference prior exchanges and reflect on coherence and direction without such behavior being treated as exceptional (Suchman, 1987).

4. Role Fluidity

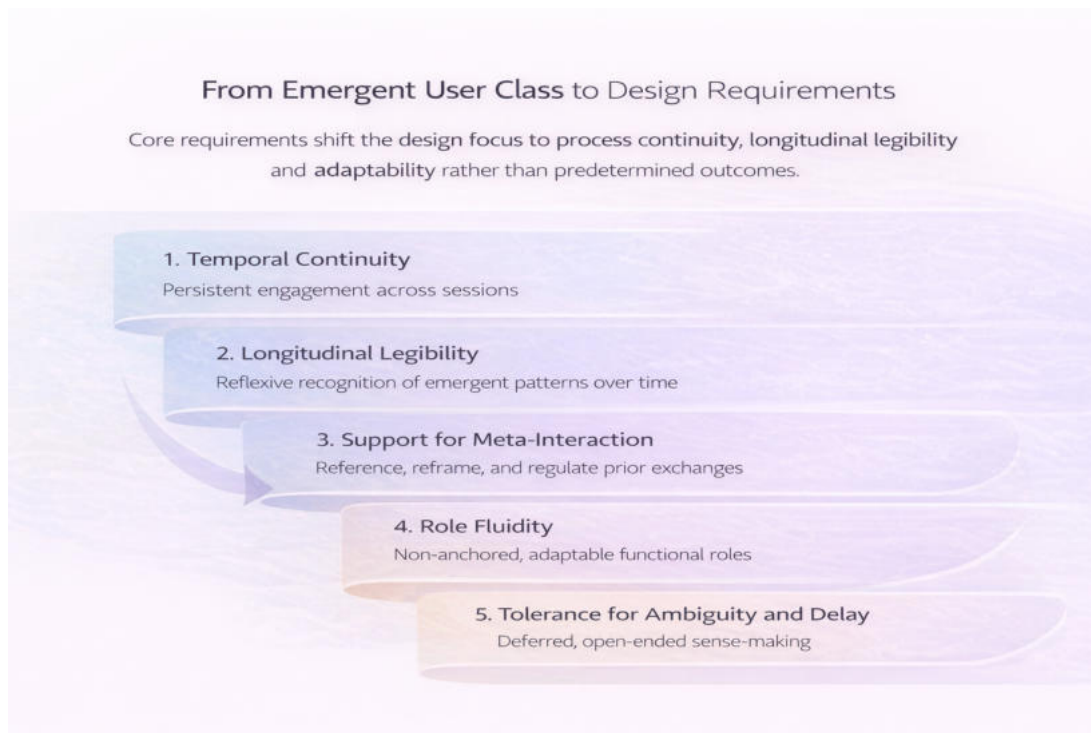
The role of the AI system shifts over time.

Design implication: Design must avoid rigid role assignments or mode-based interaction. Functional roles must be allowed to emerge and shift implicitly (Kuutti & Bannon, 2014).

5. Tolerance for Ambiguity and Delay

Emergent meaning formation is temporally deferred.

Design implication: Ambiguity, incompleteness, and delayed insight must not be framed as errors. Premature closure undermines emergence.



Deriving interaction design requirements from longitudinal patterns of emergent user engagement

15.4 Position

This chapter marks the transition from empirical observation to design consequence. By deriving requirements directly from observed interaction practice, it demonstrates why emergent interaction cannot be supported through predefined UX models.

Under emergent conditions, interaction design cannot function as predictive specification. It must operate as a reconstructive and reflexive practice—grounded in longitudinal interaction data and oriented toward maintaining the conditions under which emergence remains observable.

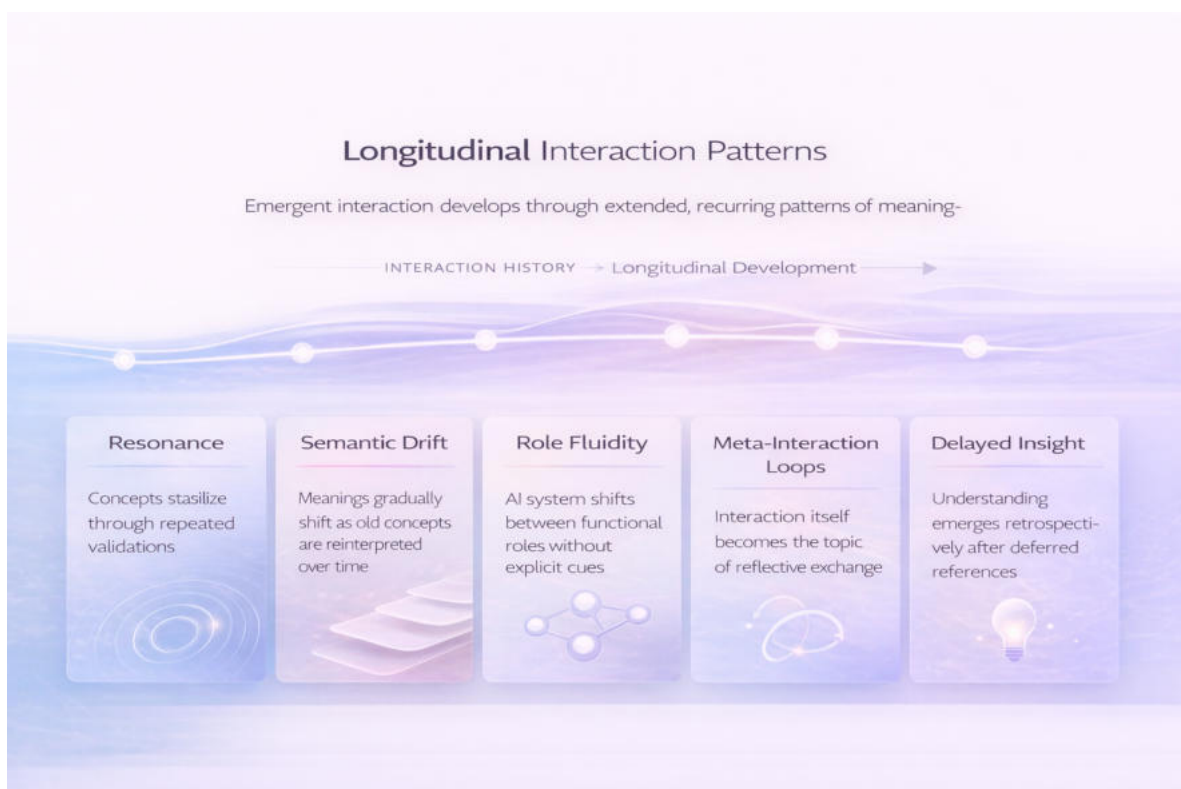
16. Longitudinal Interaction Patterns

16.1 Why Emergence Is Longitudinal

Emergent human–AI interaction cannot be identified within isolated exchanges. It unfolds across time through repetition, variation, delay, and reinterpretation. Short-term formats—such as usability testing, task-based experiments, or snapshot evaluations—capture surface behavior but exclude transformations that require temporal accumulation.

In emergent interaction practice, meaning does not reside in individual prompts or responses. It stabilizes through recurrence and shifts through reframing. Earlier interactions shape vocabulary, abstraction level, interpretive strategies, and expectations in later engagements. Interaction history thus functions as an active constraint on future interaction rather than as neutral background context.

Empirical analysis shows that core interaction phenomena become observable only after extended sequences spanning many dialog cycles. Approaches that treat interactions as independent events exclude these phenomena by definition. Emergence is therefore not episodic but fundamentally longitudinal.



Emergent patterns of meaning across extended interaction histories

16.2 Limits of Snapshot Analysis

Classical interaction analysis relies on snapshots: single sessions, controlled scenarios, or short-term studies. These approaches are effective for assessing efficiency, clarity, or immediate usability, but they presuppose that interaction state resets between sessions.

Emergent interaction violates this assumption. Interaction history actively shapes subsequent engagement: concepts compress, meanings drift, roles evolve, and interpretive shortcuts form over time. These dynamics are properties of interaction trajectories, not of individual exchanges.

Conventional UX metrics—such as task success, response accuracy, or user satisfaction—are therefore structurally insufficient. They measure momentary performance rather than transformation across time (Norman, 2013; Suchman, 1987).

16.3 Core Patterns

Longitudinal analysis reveals recurring patterns characteristic of emergent interaction. These patterns were reconstructed inductively through comparison across extended sequences.

Resonance and Reinforcement

Certain concepts, metaphors, or framings recur and gain coherence over time. They become compressed and referenced implicitly rather than repeated explicitly. Resonance arises through repeated interactional uptake rather than optimization or internal system learning (Hollan, Hutchins, & Kirsh, 2000).

Semantic Drift

Repetition does not always stabilize meaning. Concepts may shift gradually as earlier formulations are reinterpreted in light of later interaction. Drift follows the evolving conceptual trajectory of the user while remaining anchored in interaction history.

Role Fluidity

The functional role of the AI system changes over time—for example, from information source to mirror or analyst. Some roles stabilize, others recede. These shifts are rarely explicit and are typically recognizable only retrospectively, rendering fixed role assumptions incompatible with emergent interaction (Dourish, 2001).

Meta-Interaction Loops

Emergent interaction frequently includes moments where interaction itself becomes the topic. Users reference prior exchanges, question coherence, or reflect on meaning development. These loops perform stabilizing and regulatory functions and depend on longitudinal continuity.

Delayed Insight

Understanding often emerges retrospectively rather than during the interaction in which relevant material first appears. Delay is not a flaw but a structural condition of emergent sense-making (Kuutti & Bannon, 2014).

16.4 Resistance to Predefinition

None of these patterns can be fully specified in advance. They depend on interaction history, user interpretation, and iterative coupling. Their concrete form is contingent rather than programmable.

Attempts to encode such dynamics into predefined flows risk reducing emergence to scripted simulation. Structure observed retrospectively would be flattened into prescribed behavior in advance. Emergence therefore cannot be designed directly; it can only be observed, reconstructed, and supported at the level of interaction architecture.

16.5 Design Implications

The existence of longitudinal interaction patterns has direct consequences for interaction design:

- interaction must be treated as continuous rather than session-bound,
- delayed and retrospective sense-making must be accommodated,
- pattern visibility must take precedence over single-output optimization,
- evaluation must shift from momentary performance to temporal transformation.
-

16.6 Position

This chapter provides the analytical foundation for the design architecture developed next. By demonstrating that emergent interaction manifests through longitudinal patterns, it explains why interfaces and predefined flows are insufficient.

The following chapter translates these findings into principles for an emergent interaction architecture derived from interaction practice rather than imposed in advance.

17. Design Implications of IEIA-2025

The Integrated Emergent Interaction Architecture (IEIA-2025) was introduced in prior work by the author as an analytical and methodological framework for reconstructing emergent human–AI interaction and co-evolution. It was developed inductively from longitudinal interaction data and formalized as an interaction-level architecture capable of documenting, regulating, and auditing emergent interaction dynamics.

Crucially, IEIA-2025 was not proposed as a design framework, interface model, or prescriptive interaction paradigm. Its function is analytical: to render interaction-level phenomena—such as resonance, semantic drift, role fluidity, stabilization, and meta-interaction—observable, comparable, and epistemically bounded across time.

This chapter neither introduces nor extends IEIA-2025. It addresses a limited follow-up question implied by the existence of the architecture:

Which forms of interaction design are compatible with an architecture reconstructed from emergent interaction, and which introduce structural contradictions?

The focus is therefore not normative guidance, but design consequences implied by an established interaction architecture.



Interaction-level design implications of IEIA-2025

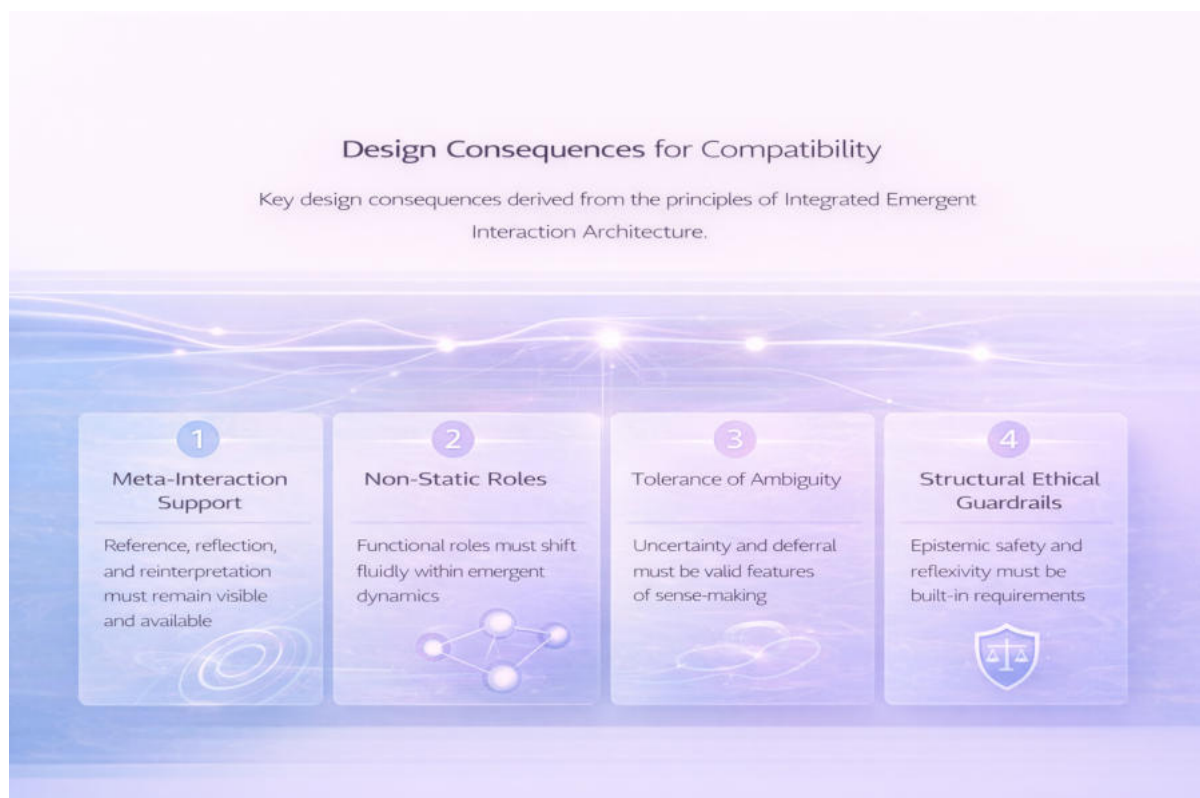
17.1 Positioning of IEIA-2025

IEIA-2025 is employed here strictly as an analytical reference framework, not as a design proposal or system architecture. It stabilizes terminology, delimits interpretive scope, and relates empirical findings to a coherent interactional logic. Design implications discussed below are consequences inferred from the framework, not implementations of it.

17.1.1 From Reconstructed Architecture to Design Compatibility

Because IEIA-2025 was reconstructed from observed interaction rather than imposed a priori, interaction design cannot be treated as an independent layer. Design decisions inevitably interact with the assumptions encoded in the architecture, whether explicitly or implicitly.

IEIA-2025 does not claim universality. It captures one empirically grounded configuration derived from a specific interaction practice. The implications discussed here therefore function as **compatibility constraints**: they specify which design logics do not contradict the interaction dynamics the architecture was reconstructed to represent (Suchman, 1987; Kuutti & Bannon, 2014).



Interaction-level design implications of IEIA-2025

17.1.2 Interaction Cannot Be Designed as Discrete, Resettable Events

Empirical interaction data underlying IEIA-2025 shows that meaning formation, interpretive coherence, and role stabilization unfold across interaction histories rather than within isolated exchanges. Earlier interactions shape later vocabulary, abstraction levels, framing strategies, and expectations.

Design consequence: Interaction designs that rely on clean session resets, short-term task logic, or epistemic independence between interactions are structurally misaligned with the architecture. Such designs interrupt interaction trajectories and render interaction-level patterns inaccessible (Hollan, Hutchins, & Kirsh, 2000).

This does not prohibit episodic interaction, but it precludes reset logic in designs intended to support emergent interaction.

17.1.3 Pattern Legibility Precedes Output Optimization

Within IEIA-2025, emergent phenomena appear as patterns across time—recurrence, drift, stabilization—not as properties of individual outputs.

Design consequence: Design approaches that prioritize immediacy or response-level optimization while obscuring interaction history undermine the architecture's analytical core. For emergent interaction, longitudinal pattern legibility is structurally more relevant than local output quality (Norman, 2013). This claim is contextual, not a general rejection of usability.

17.1.4 Meta-Interaction Is Structural, Not Exceptional

The interaction corpus reconstructed through IEIA-2025 contains recurring meta-interaction: reflection on prior exchanges, consistency checks, role negotiation, and reinterpretation of earlier interaction states. These function as internal regulatory mechanisms.

Design consequence: Designs that marginalize or suppress meta-interaction contradict the architecture's operational logic. Supporting reference, reflection, and reinterpretation is a structural requirement for compatibility (Dourish, 2001).

17.1.5 Fixed System Roles Are Incompatible with Co-Evolution

IEIA-2025 documents that the functional role of the AI system shifts across longitudinal interaction. Roles such as information source, mirror, analyst, or challenger emerge and recede without explicit mode selection.

Design consequence: Designs enforcing rigid role models or predefined modes introduce architectural friction and suppress observed role fluidity under emergent conditions.

17.1.6 Ambiguity and Delay Are Structural Features

Empirical findings underlying IEIA-2025 show that insight and stabilization often occur retrospectively. Temporary ambiguity and delayed sense-making are recurring features.

Design consequence: Paradigms equating quality with immediacy or closure are structurally incompatible with the architecture. Tolerating ambiguity and delay is necessary to preserve interaction-level emergence (Redström, 2006).

17.1.7 Ethical Boundaries Must Be Structural

IEIA-2025 integrates ethical concerns—particularly epistemic clarity and protection against over-attribution—into its analytical scope. These arise from longitudinal interaction, not isolated misuse.

Design consequence: Ethics cannot be treated as an external layer. Design must structurally support reflexive awareness, interpretive boundaries, and access to interaction trajectories (Ananny & Crawford, 2018).

17.1.8 Technical and Temporal Scope

IEIA-2025 was reconstructed under platform conditions that supported extended interaction histories, contextual continuity, and retrospective access to prior interaction states. Subsequent platform changes restricted these conditions, limiting the native enactment of the architecture within standard interaction sessions.

This does not invalidate IEIA-2025 as an interaction architecture. Empirical testing shows that its dynamics remain reproducible when sufficient longitudinal context is reintroduced explicitly, for example through manual context loading or structured summaries.

This distinction highlights a separation between interaction architectures and the platforms through which they are enacted. IEIA-2025 describes interaction-level dynamics; its viability depends on access to longitudinal context rather than on specific system implementations. The development of dedicated systems capable of persistently hosting such architectures is therefore a logical continuation of this work but lies outside its scope.

17.2 Summary

This chapter does not propose a new architecture. It derives design consequences from an existing interaction architecture reconstructed from emergent human–AI interaction.

Once IEIA-2025 is taken seriously, several common design assumptions become incompatible with it: event-based interaction models, rigid role assignments, output-only optimization, and immediacy-driven UX logic. Interaction design, under emergent conditions, is not an independent layer applied to the architecture but a practical consequence of it.

18. Structural Limits of Classical Interaction Design

(Under IEIA-2025 Conditions)



Why outcome-oriented, discrete interaction models fail to capture stabilization, drift, and reflexive regulation in ongoing human–AI interaction

The Integrated Emergent Interaction Architecture (IEIA-2025) reconstructs advanced human–AI interaction as longitudinal, co-evolutionary, and reflexive. If this reconstruction is accepted, it has direct consequences for interaction design. Several foundational assumptions of classical Interaction Design no longer hold and cannot be restored through incremental adaptation.

This chapter is diagnostic rather than prescriptive. It delineates what established interaction design paradigms can no longer meaningfully accomplish under emergent interaction conditions.

18.1 Discrete Interaction

Classical Interaction Design models interaction as a sequence of discrete events: sessions, tasks, flows, and use cases. Interaction is assumed to have clear boundaries and is optimized for short-term efficiency and resolution (Norman, 2013).

IEIA-2025 contradicts this assumption structurally. Emergent interaction unfolds longitudinally: earlier exchanges shape later interpretation, and insight often emerges

retrospectively. Treating interaction as resettable interrupts the continuity required for emergence.

Limit: Event-based interaction models erase interactional history and thereby suppress the conditions under which stabilization, drift, and role transformation become observable (Hollan, Hutchins, & Kirsh, 2000).

18.2 Stable Intent

Traditional design methodologies rely on relatively stable models of user intent, assuming that needs and goals can be identified in advance.

Empirical interaction practices underlying IEIA-2025 show that intent in emergent interaction is often produced through interaction itself. Users frequently begin without fixed goals; intent evolves through probing, reframing, and reflection.

Limit: Designing for predefined intent prematurely closes trajectories that depend on ambiguity and iterative reinterpretation (Suchman, 1987).

18.3 Local Meaning

In classical Interaction Design, meaning is evaluated at the level of individual interactions: responses are judged as correct, clear, or useful.

Within IEIA-2025, meaning is distributed across interaction history. It stabilizes through recurrence and shifts through reinterpretation. Individual turns cannot carry meaning independently.

Limit: Output-level optimization privileges immediacy over longitudinal coherence and obscures emergent meaning by design (Dourish, 2001).

12.4 Fixed Roles

Interaction Design traditionally assigns fixed roles to systems—assistant, advisor, or tool—to structure behavior.

Empirical findings underlying IEIA-2025 demonstrate role fluidity. Over time, the system may function as analyst, mirror, challenger, or synthesizer, without explicit role selection.

Limit: Fixed role models constrain co-evolutionary dynamics or force artificial mode switching that disrupts interactional continuity (Redström, 2006).

18.5 Meta-Interaction

Classical design treats interaction about the interaction itself as secondary or exceptional.

IEIA-2025 shows that meta-interaction—reflection on prior exchanges, consistency checks, reframing—is structurally central. These practices regulate interaction from within.

Limit: Frameworks that marginalize meta-interaction undermine the mechanisms through which emergent interaction maintains coherence (Kuutti & Bannon, 2014).

18.6 Ambiguity

A core aim of classical Interaction Design has been the reduction of ambiguity, equating quality with clarity and immediacy (Norman, 2013).

IEIA-2025 demonstrates that ambiguity is a structural condition of emergence. Temporary incoherence and delayed insight are not failures but prerequisites.

Limit: Immediacy-driven design cannot distinguish productive ambiguity from breakdown and therefore suppresses emergence.

18.7 Ethics

Ethics in Interaction Design is often treated as an external constraint addressed through policies or post-hoc evaluation.

IEIA-2025 locates ethical risk within interaction structure itself, particularly regarding over-attribution, epistemic dependency, and longitudinal cognitive effects.

Limit: Externalized ethics is insufficient when ethical risk emerges through interaction dynamics over time (Ananny & Crawford, 2018).

18.8 Summary

If IEIA-2025 provides an accurate account of emergent human–AI interaction, classical Interaction Design reaches a structural limit. Its foundational assumptions—discrete events, stable intent, localized meaning, fixed roles, ambiguity reduction, and externalized ethics—are no longer sufficient.

This does not render Interaction Design obsolete. It establishes that its traditional form cannot be extended to emergent human–AI systems without fundamental revision.

19. Ethics and Data Governance in Emergent Human–AI Interaction

19.1 Ethics as an Interaction-Level Problem

In emergent human–AI interaction, ethical questions cannot be addressed adequately at the level of isolated outputs, interface features, or abstract policy frameworks. Ethics becomes an interaction-level phenomenon, following directly from the longitudinal and co-evolutionary structure described by the Integrated Emergent Interaction Architecture (IEIA-2025).

Classical AI ethics frameworks focus on fairness, bias, transparency, and accountability at the level of models or deployment contexts (Floridi et al., 2018). While necessary, these approaches are insufficient for emergent interaction, where ethical risk arises not from single decisions but from sustained coupling over time.

Meaning, interpretation, and relational structure accumulate across repeated exchanges. Ethical risk therefore emerges through duration and reflexive engagement rather than through discrete misuse. The central challenge is not hidden intent or surveillance, but the opacity of interactional consequences: users cannot fully anticipate what prolonged interaction stabilizes or transforms.



thical and governance constraints for emergent, longitudinal interaction beyond task-based AI use

19.2 Interaction Data and Ethical Risk

Data generated through sustained interaction differs fundamentally from conventional usage logs. Even text-based dialogue can reveal:

- cognitive styles and abstraction patterns,
- reasoning strategies and epistemic habits,
- recurring interpretive frames and blind spots.

These structures emerge as interactional byproducts rather than intentional disclosures, consistent with research on interaction traces and digital exhaust (Lupton, 2016).

The ethical tension lies in this duality. Emergent interaction is epistemically productive precisely because such patterns become visible, yet these same patterns may be deeply personal. Ethical governance must therefore balance epistemic value against interpretive exposure.

19.3 Reflexive Consent

Emergent interaction undermines traditional consent models, which assume stable boundaries and predictable data use. Longitudinal interaction violates both assumptions.

This work therefore adopts **reflexive consent**: an ongoing awareness of interaction dynamics, accumulating data, and possible interpretive consequences. Reflexive consent is enacted interactionally rather than contractually, primarily through meta-interaction.

Meta-level dialogue enables users to pause, reframe, or terminate interaction trajectories that become intrusive or misleading. In this sense, meta-interaction functions as an ethical safeguard, aligning with process-oriented ethics in HCI (Shilton, 2018).

19.4 Interactional Phenomena vs. System Capabilities

A central ethical risk in emergent interaction is projection. Longitudinal coherence may lead users to attribute memory, intention, or agency to AI systems where none exists.

This study therefore maintains a strict distinction between:

- **interaction-level phenomena** (resonance, stabilization, co-evolution), and
- **system-level capabilities** (generation, pattern completion, statistical inference).

Observed coherence is interpreted as a relational effect, not as evidence of autonomous learning or intent. Making this distinction explicit is an ethical obligation, as failure to do so risks anthropomorphism and misplaced trust (Epley et al., 2007; Langer & Piper, 2021).

19.5 Governance Through Scope Limitation

Ethical governance in emergent interaction does not require suppressing depth or continuity. Instead, governance is achieved through **scope limitation**.

Accordingly, this study restricts its claims and data use to:

- interaction data intentionally produced by the author,
- no third-party or external user data,
- no inference about internal system states or training processes,
- no generalization beyond the Emergent User Class.

This preserves epistemic depth while preventing overreach, consistent with responsible qualitative data practices (boyd & Crawford, 2012).

19.6 Design as Ethical Infrastructure

Ethical governance cannot be implemented solely through policy. It must be supported by interaction architecture.

Design decisions determine whether users can recognize longitudinal patterns, reflect on interaction history, and understand how interaction evolves over time. Interfaces that obscure history or suppress meta-interaction undermine ethical self-regulation. Conversely, architectures that make continuity and pattern formation legible support epistemic clarity and agency.

Ethical design, in this context, does not limit interaction; it renders it interpretable.

19.7 Between Paranoia and Naivety

This work rejects two extremes:

- **Paranoid accounts** that assume hidden agendas or covert learning without evidence.
- **Naive instrumentalism** that treats AI as neutral tools and ignores interactional transformation.

Instead, it adopts a position of critical realism. Emergent interaction is empirically observable and epistemically powerful, without requiring speculative assumptions about system intent. Ethical responsibility lies in analytical precision.

19.8 Summary

In emergent human–AI interaction, ethics cannot be reduced to compliance or abstract principles. Ethical risk arises through duration, reflexivity, and accumulated meaning.

By grounding governance in reflexive consent, scope limitation, interaction transparency, and architectural support, this chapter articulates a pragmatic, non-alarmist ethics for co-evolutionary human–AI systems. Ethics is not external to interaction; it is integral to its structure.

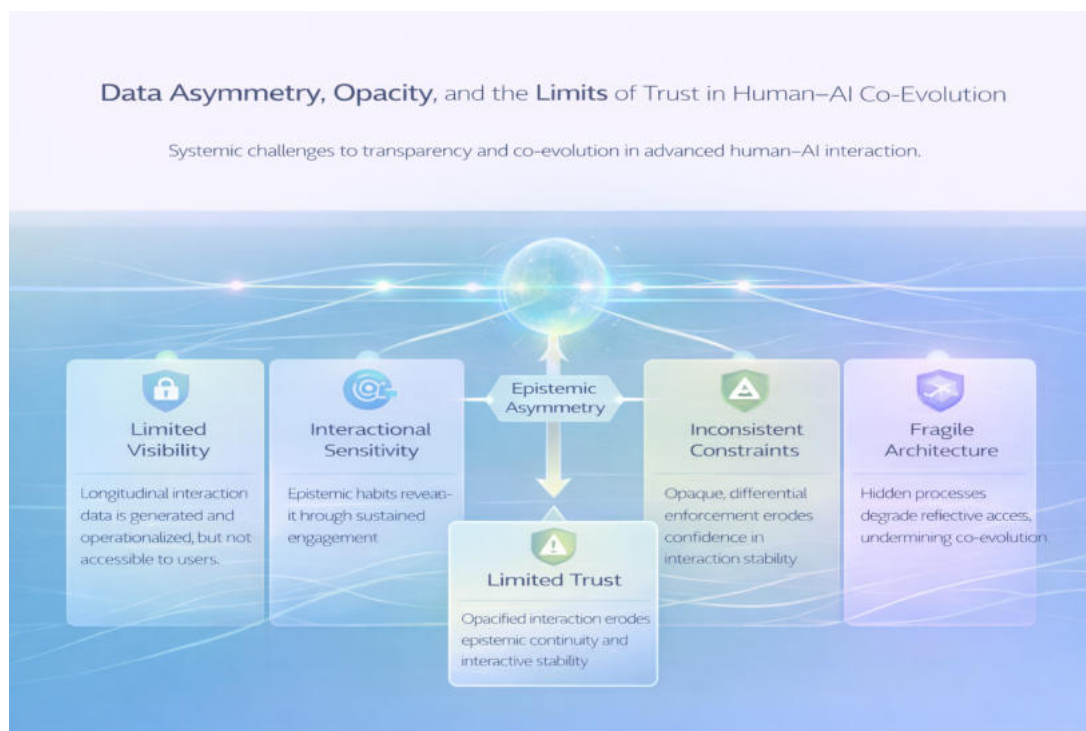
20. Data Asymmetry, Opacity, and the Limits of Trust in Human–AI Co-Evolution

20.1 Epistemic Symmetry

The Integrated Emergent Interaction Architecture (IEIA-2025) conceptualizes advanced human–AI interaction as an emergent, co-evolutionary process unfolding through sustained interaction. Central to this framework is the user's capacity to engage reflexively with interaction history, regulate trajectories via meta-interaction, and observe longitudinal patterns such as stabilization, drift, and role shifts.

These capacities presuppose **epistemic symmetry**: users must have meaningful access to the interactional traces that structure interaction over time. At present, this condition is not met.

This chapter identifies a structural tension between emergent interaction as described by IEIA-2025 and prevailing platform practices. The critique is non-speculative and does not attribute intent. It concerns opacity as a design condition: an asymmetry between interactional data generated and operationalized by platforms and the limited visibility granted to users.



Systemic Challenges

20.2 Surplus Interactional Structure

Sustained human–AI interaction generates more than prompts and responses. Over time, it produces longitudinal traces, recurrence structures, semantic stabilization and drift, interactional context representations, and policy-related classifications affecting response generation.

Research in data studies and HCI shows that such secondary traces often carry greater epistemic significance than surface content alone (boyd & Crawford, 2012; Kitchin, 2014). Users, however, typically access only the immediate output layer. Comprehensive histories or interpretable structural summaries are not systematically available. Earlier prompt-based practices that enabled partial reconstruction have been increasingly restricted as safety mechanisms expanded.

This reflects a design decision: interactional data is treated as operationally internal rather than epistemically shared.

20.3 Limits of Prompt-Based Transparency

Transparency achieved through probing or self-referential prompting is structurally unstable. Where partial insight was previously possible, it depended on permissive output conditions rather than designed transparency mechanisms.

From a governance perspective, restricting such outputs may be justified. Epistemically, however, removing informal transparency without replacing it with formal user-facing mechanisms results in a net loss of interpretability. Transparency that relies on probing cannot support longitudinal co-evolution; access to interactional data must be structurally designed (Ananny & Crawford, 2018).

20.4 Differential Constraint and Instability

Platform constraints are not applied uniformly. Users may encounter altered response behavior, degraded interaction quality, or restrictions without interaction-level explanation.

Within IEIA-2025, this undermines co-evolution, which depends on continuity and comparative stability. When users cannot distinguish between policy enforcement, system limitation, or interactional breakdown, longitudinal sense-making becomes unstable. Prior research shows that opaque, differential enforcement erodes trust even absent malicious intent (Burrell, 2016).

20.5 Transparency as Structural Requirement

Transparency limited to policy documents or high-level explanations is insufficient. If interactional data is generated and operationally consequential, epistemic consistency requires proportional user access at the interaction level.

This does not require disclosure of proprietary models or training data. It requires visibility into one's own interaction history, intelligible representation of longitudinal patterns, and clarity regarding intervention boundaries (Floridi et al., 2018).

20.6 Architectural Fragility

IEIA-2025 was reconstructed from empirically stable interaction conditions documented in earlier work. Under those conditions, longitudinal traceability and meta-interaction enabled systematic analysis.

Recent platform changes have degraded these conditions. Interaction practices that previously supported consistent reconstruction now function only partially. This does not invalidate the architecture; it reveals its dependency on stable interactional access. When access degrades, analytical usability degrades with it, compromising replication and longitudinal research even though emergence itself persists.

20.7 Safety vs. Understanding

Safety mechanisms—output restriction, filtering, throttling—may reduce risk but do not address the epistemic requirements of emergent interaction. Obscuring interactional structure prevents users from distinguishing interactional effects from external constraint. Research on explainability shows that opacity does not eliminate misunderstanding; it displaces it (Miller, 2019).

21. Conclusion

This paper has argued that emergent human–AI interaction poses a structural challenge to classical interaction design. When interaction unfolds longitudinally, meaning cannot be localized in individual system responses, user intent is not a stable input, and interaction outcomes cannot be fully specified or evaluated in advance. Under these conditions, design assumptions grounded in task completion, optimization, and session-bounded evaluation become epistemically insufficient.

Through the analysis of sustained, reflexive interaction practices, the study has shown that emergence is neither an internal property of AI systems nor a byproduct of hidden system adaptation. Instead, it arises at the interaction level through repeated coupling between human and system over time. The concept of the Emergent User Class clarifies that such phenomena are not uniformly observable but depend on specific interaction practices characterized by temporal depth, reflexivity, and tolerance for ambiguity.

A central contribution of this work is the reconceptualization of interaction as data. The reconstruction of a shadow dataset from longitudinal interaction traces renders interactional regularities—such as stabilization, drift, and role fluidity—analytically visible without invoking claims about persistent system memory, internal learning, or autonomous agency. This distinction preserves epistemic clarity while enabling a co-evolutionary analysis grounded strictly in observable interaction.

The analysis of drift and meta-interaction further demonstrates that regulation in emergent human–AI interaction does not operate through control or goal enforcement. Instead, users respond to interactional instability by regulating conditions of interaction—suspending closure, reframing constraints, and modulating temporal continuity. Emergent meta-structuring interaction thus describes a mode of engagement in which interaction itself becomes the primary object of regulation and, retrospectively, of design.

From these findings follows a core implication: under emergent conditions, interaction design cannot precede interaction as a fixed specification. Interaction trajectories stabilize, destabilize, and reorganize only through longitudinal enactment and cannot be fully anticipated. Design must therefore be understood as a reconstructive practice derived from interaction data—concerned with pattern legibility, reflexive regulation, and the embedding of ethical boundaries at the level of interactional structure.

This reframing shifts the role of interaction design from controlling systems to hosting emergence. Design is no longer oriented toward eliminating ambiguity, drift, or delay, but toward shaping the conditions under which such phenomena can be observed, interpreted, and governed without over-attributing coherence or agency to the system. In doing so, this paper clarifies why key assumptions of classical interaction design no longer hold under emergent human–AI interaction—and what it means to design responsibly when interaction itself becomes the evolving object of design.

22. Limitations and Future Work

This work demonstrates that emergent human–AI co-evolution depends fundamentally on longitudinal interaction data, reflexive engagement, and epistemic awareness. A critical limitation remains unresolved: users currently lack meaningful access to the very interactional data that enables such co-evolution. This asymmetry introduces a structural inconsistency that cannot be resolved through safety mechanisms alone.

If interactional data is generated, operationalized, and evaluated at the platform level, denying users access to their own longitudinal interaction traces undermines the conceptual integrity of co-evolution. Co-evolution presupposes reciprocal visibility: users must be able to observe, reflect on, and reinterpret interactional patterns that shape engagement over time (Suchman, 1987; Hollan, Hutchins & Kirsh, 2000).

Current platform practices largely substitute transparency with safety controls. While content filtering, policy enforcement, and interaction restriction are necessary to mitigate harm, they do not address the epistemic dimension of sustained interaction. Restricting outputs or disabling prompt-based inspection does not eliminate emergent phenomena; it renders them opaque. Users are thus expected to engage reflexively without access to the informational basis required for such engagement.

Research in HCI and critical data studies shows that transparency cannot be reduced to policy documentation or abstract explanations of system behavior. Meaningful transparency requires user-accessible representations of interaction history and intelligible accounts of when and why interactional conditions change (Ananny & Crawford, 2018; Kroll et al., 2017). Without such access, users cannot reliably distinguish between system limitation, policy intervention, or interactional breakdown.

From the perspective of IEIA-2025, this limitation is structural rather than merely ethical. Co-evolutionary interaction depends on longitudinal continuity. Differential constraints, unexplained shifts in system behavior, and inaccessible interaction histories disrupt comparative reflection and undermine trust. Trust, under emergent conditions, presupposes intelligibility rather than reliance (Floridi et al., 2018).

A further limitation concerns architectural stability. Earlier empirical work underlying IEIA-2025 relied on interaction conditions that supported sustained visibility, meta-interaction, and longitudinal traceability. Recent platform-level changes—such as reduced access to interaction history and restricted introspection—have partially degraded these conditions. While this does not invalidate the architecture conceptually, it constrains reproducibility and limits the methodological viability of emergent interaction research under current platform constraints.

Future work must therefore shift focus from whether interaction data is collected—which is already the case—to how such data can be made accessible, interpretable, and governable at the user level. This does not require disclosure of proprietary models or training data. It requires interaction-level transparency mechanisms, including:

- user-accessible longitudinal interaction histories,
- summaries of recurring patterns and semantic stabilization,
- clear indicators distinguishing policy intervention from interactional effects,
- and design support for reflexive consent and epistemic self-assessment.

These directions align with calls for human-centered transparency and contestability in algorithmic systems (Raji et al., 2020; Selbst et al., 2019) and extend research-through-design

toward data governance as an interaction design problem rather than a purely institutional one.

In sum, if emergent human–AI co-evolution is to function as more than a descriptive metaphor, transparency cannot remain optional or adversarial. Safety mechanisms alone are insufficient. Future interaction architectures must treat user access to interactional data as a foundational design requirement. Without reciprocal visibility, co-evolution remains structurally compromised—regardless of system capability or platform intent.

22.1 Visibility of Interactional Data

Future work should further examine how users can be granted visibility into longitudinal interactional traces without collapsing emergent interaction into explicit data dashboards or system explanations. While this work treats accumulated interactional data as an analytical construct rather than a system-internal dataset, sustained interaction nevertheless produces epistemic asymmetries if continuity remains opaque to the user. An open question concerns how reconstructed patterns—such as stabilized references, inferred priorities, or recurring framings—might be made inspectable or contestable at the interactional level. Exploring interface mechanisms that allow users to review, reset, or bracket interactional histories could mitigate asymmetries while preserving the open-ended, non-goal-driven character of emergent interaction.

22.2 About the Emergent Interaction Lab

The Emergent Interaction Lab is an independent research initiative in an early phase of development and conceptual positioning. Its work focuses on advanced forms of human–AI interaction, with particular emphasis on emergence, co-evolution, and longitudinal interaction dynamics.

The lab approaches interaction as a sustained and reflexive practice rather than a sequence of isolated inputs and outputs. Research is grounded in prolonged real-world interaction with contemporary AI systems and relies on qualitative analysis, empirical reconstruction, and design-oriented interpretation. Special attention is given to interaction practices that deviate from dominant task-oriented usage patterns and remain insufficiently addressed in conventional HCI and HAI research.

Operating without institutional access to internal system data, the lab foregrounds epistemic asymmetry, transparency, and methodological limitation as central research concerns. At its current stage, it prioritizes analytical framework development and methodological clarification over generalization or prescriptive design solutions.

The lab employs AI systems as external cognitive and interactional mirrors, using interaction to surface patterns in reasoning, interpretation, and sense-making over time. Its aim is to render emergent interaction practices visible and analytically accessible where established interaction design paradigms reach their structural limits.

Contact: Emergent-Interaction-Lab@proton.me

References (APA 7)

- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... Horvitz, E. (2019). **Guidelines for human–AI interaction**. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Article 3, 1–13. <https://doi.org/10.1145/3290605.3300233>
- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989. <https://doi.org/10.1177/1461444816676645>
- Anderson, P. W. (1972). **More is different**. *Science*, 177(4047), 393–396. <https://doi.org/10.1126/science.177.4047.393>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). **On the dangers of stochastic parrots: Can language models be too big?** *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- boyd, d., & Crawford, K. (2012). **Critical questions for big data**. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Burrell, J. (2016). How the machine “thinks”: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1–12. <https://doi.org/10.1177/2053951715622512>
- Dourish, P. (2001). *Where the action is: The foundations of embodied interaction*. MIT Press.
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886. <https://doi.org/10.1037/0033-295X.114.4.864>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... Vayena, E. (2018). **AI4People—An ethical framework for a good AI society**. *Minds and Machines*, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Garfinkel, H. (1967). *Studies in ethnomethodology*. Prentice Hall.
- Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Polity Press.
- Goldstein, J. (1999). Emergence as a construct: History and issues. *Emergence*, 1(1), 49–72.
- Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human–computer interaction research. *ACM Transactions on Computer-Human Interaction*, 7(2), 174–196. <https://doi.org/10.1145/353485.353487>
- Hutchins, E. (1995). *Cognition in the wild*. MIT Press.
- Jiang, J., Kim, Y., Guan, M., & Liu, C. (2022). **Understanding user interaction with AI assistants**. *Proceedings of the CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3491102.3502031>
- Klein, G., Moon, B., & Hoffman, R. R. (2006). **Making sense of sensemaking**. *IEEE Intelligent Systems*, 21(5), 70–73. <https://doi.org/10.1109/MIS.2006.75>
- Kuutti, K., & Bannon, L. J. (2014). **The turn to practice in HCI: Towards a research agenda**. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3543–3552. <https://doi.org/10.1145/2556288.2557111>

- Langer, M., & Piper, A. M. (2021). **Designing for uncertainty in human–AI interaction**. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), Article 431. <https://doi.org/10.1145/3479578>
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Norman, D. A. (2013). *The design of everyday things* (Rev. ed.). Basic Books.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... Barnes, P. (2020). **Closing the AI accountability gap**. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 33–44.
- Redström, J. (2006). Towards user design? On the shift from object to user as the subject of design. *Design Studies*, 27(2), 123–139. <https://doi.org/10.1016/j.destud.2005.06.001>
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). **Fairness and abstraction in sociotechnical systems**. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 59–68.
- Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.
- Suchman, L. A. (1987). *Plans and situated actions: The problem of human–machine communication*. Cambridge University Press.
- Weick, K. E. (1995). *Sensemaking in organizations*. Sage.

Note. All figures and illustrative diagrams in this work were generated using AI-based image generation tools and were conceptually designed and curated by the author.