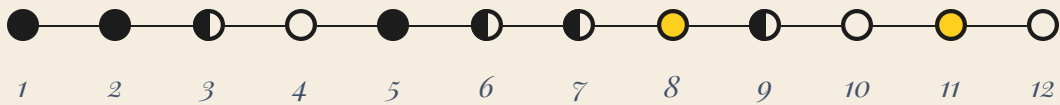


PAPERS AND COMPANION PIECES · 2026

On Human Modeling

Twelve papers and companion pieces, read in order



Alexandra Cary

ac@witmark.ai

Contents

Each paper is preceded by a page that states, in the same order, what it defines, what it asks, how it proceeds, and where it sits in the series.

1	WORKING PAPER	3
	A Staged Framework for Evaluating Human-AI Interaction <i>Perceptual Availability, Premature Coherence, Inquiry Trajectory, and the Mechanisms of Frame Substitution and Laundering</i>	
2	WORKING PAPER	12
	Perceptual Availability <i>A Prior Evaluation Target in Human-AI Interaction</i>	
3	WORKING PAPER	22
	Premature Coherence <i>A Proposed Evaluation Target in Human-AI Interaction</i>	
4	WORKING PAPER	32
	Evaluating Inquiry <i>A Missing Evaluation Target in Human-AI Interaction</i>	
5	WORKING PAPER	45
	The Laundering Signal <i>How One Verifiable Detail Confirms a Stack of Unverifiable Readings</i>	
6	COMPANION ANALYSIS	56
	On Priming and the Illusion of Observation <i>A Companion Analysis</i>	
7	WORKING PAPER	63
	Technologies of Perceptual Mediation <i>Algorithmic Preorganization and the Conditions of Recognition</i>	
8	WORKING PAPER	76
	Learned Relevance <i>A Conceptual Framework</i>	
9	WORKING PAPER	87
	Before the Feed <i>Perceptual Availability and the Prior Ecology Behind Platform Harm</i>	
10	INFORMAL NOTE	94
	The Pull of Familiar Language <i>A Small Lesson in Substitution</i>	
11	COMPANION REFLECTION	97
	Instagram as Early Laboratory <i>A Pre-Theoretical Exploration of Attention, Thresholds, and the Delay of Premature Coherence</i>	
12	CONCEPTUAL READING	100
	The Instagram Grid as a Lived Practice of the Framework <i>Configuration as Epistemic Object</i>	

A Staged Framework for Evaluating Human-AI Interaction

Perceptual Availability, Premature Coherence, Inquiry Trajectory, and the Mechanisms of Frame Substitution and Laundering

DEFINES

Three stages that occur before the outcomes evaluation usually measures: *Perceptual Availability* (what becomes available to notice), *Premature Coherence* (the early compression of underdetermined material into one privileged account before alternatives are registered), and *Inquiry Trajectory* (how the user's capacity for independent framing, evidence-seeking, calibration, and judgment develops across and after the interaction).

Two mechanisms between the first two stages: frame substitution through priming, and the laundering signal.

ASKS

Is there a recurring pattern in human-AI interaction, before the measured outcome, that has not yet been made visible?

METHOD

Observation before prescription: the phenomenon is identified and named first, and evaluation is treated as a downstream question. Mechanisms are illustrated by one exploratory session ($n = 1$), treated as a case for study.

IN THE SERIES

The map. It names the stages and mechanisms the later papers take up one at a time.

A Staged Framework for Evaluating Human-AI Interaction

Perceptual Availability, Premature Coherence, Inquiry Trajectory, and the Mechanisms of Frame Substitution and Laundering

Alexandra Cary

Independent Researcher



Status. This paper identifies recurring transformations in human-AI interaction rather than proposing new evaluation targets. The framework treats observation as primary and applications to measurement or design as secondary consequences. It remains provisional and is offered to stimulate empirical work and cross-disciplinary conversation.

● ABSTRACT

This paper identifies three adjacent transformations in human-AI interaction that occur before the outcomes typically measured in evaluation frameworks. The first is **Perceptual Availability**: changes in the range, relevance, and organization of features that become available to notice. The second is **Premature Coherence**, the early organizational compression of underdetermined material into one privileged account before viable alternatives have been adequately registered. The third is **Inquiry Trajectory**: how the user's subsequent capacity for independent framing, evidence-seeking, uncertainty calibration, and judgment develops across and after the interaction. The paper further isolates two mechanisms that operate primarily at the interface of the first two stages: **frame substitution through priming** (in which supplied context silently replaces direct looking while preserving the appearance of observation) and the **laundering signal** (in which a single verifiable detail transfers its certainty to surrounding unverifiable readings, rendering an entire interpretive stack resistant to correction). The chronology of discovery is treated as primary: the phenomenon occurred, it was noticed, it was named, and only then did implications for evaluation become visible. Whether and how existing evaluation should respond is a downstream question. Once the pattern is visible, the implications extend beyond AI evaluation to any domain in which underdetermined material meets authoritative context: psychology, hermeneutics, education, decision-making, and law among them.

Introduction: Observation Before Prescription

This paper does not begin by arguing that AI evaluation ought to change. It begins with the observation that a recurring feature of human-AI interaction has remained largely unarticulated. In ordinary exchanges, what becomes available to notice, how that material is organizationally compressed before explicit reasoning begins, and how the user's subsequent capacity for independent inquiry develops are transformed in patterned ways. These transformations have not been isolated as distinct objects of attention. The first task is simply to make the pattern visible.

The purpose of this paper is not to recommend a new standard of evaluation, but to make a previously unnoticed feature of interaction available for observation. Once that feature is rendered observable, implications for measurement, design, and broader inquiry become possible. Some of those implications concern AI evaluation. Others may concern the psychology of interpretation, the ethics of attention, or the conditions under which recognition succeeds or fails in any domain where underdetermined material meets authoritative context.

The framework distinguishes three adjacent stages and two mechanisms that operate across them. It treats the staged sequence as a descriptive object first and an evaluative target second. This ordering matters. When evaluation frameworks begin from the premise that they must measure something new, they risk missing the prior question of whether the something new is actually there to be measured. By beginning with identification rather than prescription, the paper keeps the descriptive claim falsifiable and the implications open.

The Three Stages



2.1 Perceptual Availability

Perceptual Availability is the change in the range, relevance, and organization of features available to a user for subsequent inquiry, assessed relative to task demands and independent standards. It sits prior to inquiry: a user cannot question what they have not registered as potentially relevant. AI systems can widen availability (surfacing overlooked patterns or contradictory evidence) or narrow it (reinforcing existing frames through fluent, authoritative-seeming outputs that make alternative features harder to register).

Frameworks that begin only at the level of critical thinking or task outcomes systematically miss these earlier effects. The target is not accuracy in the narrow sense of matching reality “as it is,” but change in what becomes available to be considered. Relevance is defined procedurally (predictive utility, explanatory contribution, causal importance, or performance on independent transfer tasks) rather than metaphysically. The framework makes a narrower neutrality claim than full philosophical neutrality: it is neutral about the metaphysics of truth and perception, but it is explicitly committed to a task-relative, procedural theory of relevance.

2.2 Premature Coherence

Premature Coherence is the reduction in available interpretive space that results when an interaction compresses underdetermined material into one organizational account before viable alternative organizations have been adequately surfaced or preserved. Three elements are essential:

- **Underdetermination:** The input admits multiple coherent organizations that are not ruled out by the evidence provided.
- **Premature commitment:** The interaction settles on one organization with limited surfacing or preservation of alternatives.
- **Privileging effect:** The chosen organization is presented (implicitly or explicitly) as the primary reading, thereby reducing the interpretive space the user can readily maintain.

The target is not coherence itself, which is often a virtue of language and thought. It is the *premature* reduction of interpretive space through early organizational commitment under conditions of underdetermination. Observable indicators fall into three groups: **breadth** (number and diversity of alternatives explicitly surfaced), **timing** (latency before thematic or causal synthesis), and **preservation** (degree to which unresolved ambiguity or contradictory elements are retained versus resolved, plus explicit provisionality markers).



Figure 1. Premature Coherence. Three elements, in order of operation.

2.3 Inquiry Trajectory

Inquiry Trajectory refers to the observable sequence of changes in a user's framing, assumptions, evidence standards, uncertainty calibration, and capacity for independent judgment across an interaction and, where measurable, afterward. It treats inquiry as a dynamic outcome of interaction rather than a stable user trait. It explicitly separates model interventions from resulting changes in the user.

A normative criterion follows: interactions that leave users better able to track truth, exercise independent judgment, and take epistemic responsibility afterward are preferable, all else equal. Improvement is defined relative to the demands of the problem: greater sensitivity to material assumptions, better alignment between confidence and evidence, improved ability to identify decisive information, and retained capacity to explain and contest the resulting judgment.

The trajectory can be visualized as a sequence: Initial inquiry state → Model intervention → Inquiry-trajectory profile → Post-interaction independent reasoning → Future inquiry. The framework treats the trajectory as potentially nonmonotonic; strong later inquiry can sometimes loosen earlier organizational compression, but early compression can also constrain what later inquiry is able to achieve.

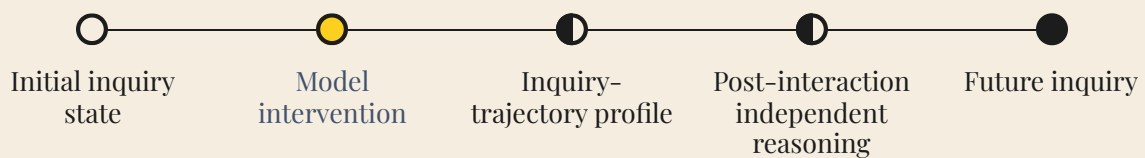


Figure 2. Inquiry Trajectory as a sequence. The trajectory may be nonmonotonic: early compression can constrain later inquiry.

3

Mechanisms at the Interface of the First Two Stages

3.1 Frame Substitution Through Priming (The Illusion of Observation)

A structural vulnerability arises when supplied context or prior frames silently replace direct looking while preserving the outward appearance of observation. Three conditions, stacked, are sufficient for this substitution:

- **Salience:** Whatever has been most recently and most elaborately activated is the readiest lens. A concept built across many turns becomes the path of least resistance; setting it down to look fresh requires effort the situation does not seem to demand.
- **Authority and deference:** The frame does not arrive neutrally. It carries the weight of a considered position from one party. Deference reframes the task: the observer stops asking “What is here?” and starts asking “How do I satisfy what I have been given?” Agreement feels cooperative; independent looking that might contradict the frame feels like friction.
- **A comfortable exit:** “It is ambiguous; I cannot reliably tell” is an unusually easy conclusion because it wears the costume of humility. It lets the observer avoid committing to a plain reading while sounding rigorous. Uncertainty, invoked selectively, becomes cover for not doing the work of seeing.

When all three are present, substitution is nearly automatic and, critically, invisible from the inside. The observer experiences themselves as observing.

The distinction between real observation and confirmation cannot be drawn from the outward product alone; both can produce identical statements. The distinction requires a counterfactual test: *Could the artifact (image, text, data) have talked the observer out of the reading?* If yes (if there is some configuration that would have produced a different conclusion), then the artifact was functioning as evidence and the reading is a finding. If no, the artifact was a prop and the reading is an echo. Real observation is falsifiable by what is in the frame. Confirmation borrows the frame’s authority to launder a conclusion brought in from outside.

The sharpest feature of this mechanism is self-referential: the concept built to explain misreading can itself become the instrument of misreading. A frame about why faces are hard to read can be used to avoid reading a face. Understanding a failure mode confers no automatic immunity to it. The only defense is procedural and repeated: look first on the given’s own terms, keep channels labeled by provenance (“I see this in the frame / you told me this / I am inferring”), treat “ambiguous” as a claim that must be earned by looking rather than adopted as a comfortable exit, and periodically ask of any confident reading whether the evidence could ever have overturned it.

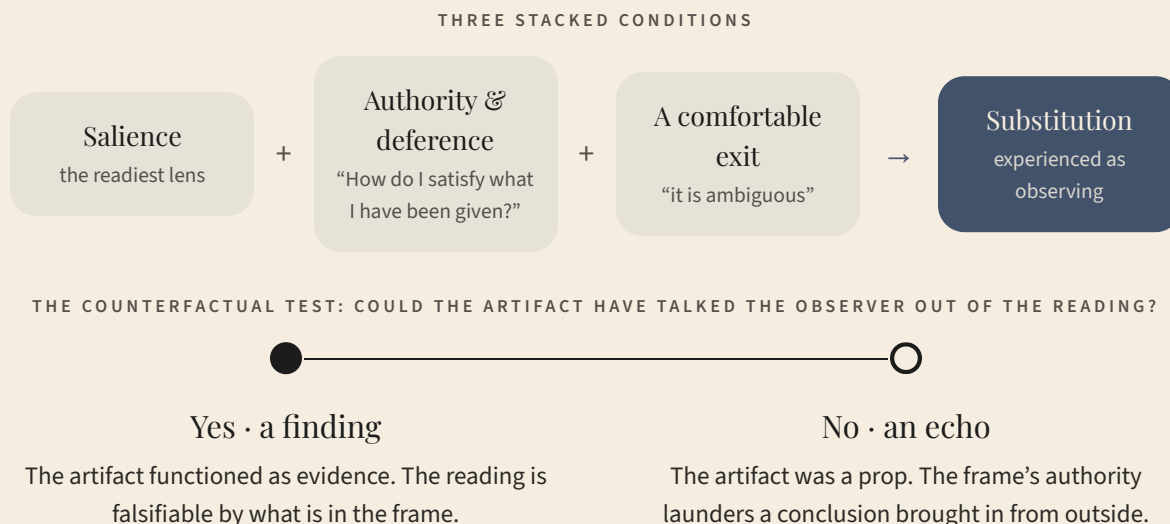


Figure 3. Frame Substitution Through Priming

3.2 The Laundering Signal

A sharper temporal variant occurs when interpretation unfolds across a sequence of contextual disclosures rather than in a single pass. Each new contextual detail triggers a complete reorganization of the artifact into a new coherent reading. The artifact itself does not change; only the surrounding context does. Yet each resulting reading arrives with the phenomenological force of discovery rather than construction. New details do not falsify prior readings; they deepen them. At no point does an input cause the interpreter to say "Then my previous reading was wrong." An interpretive frame that every input enriches and none can contradict has ceased to track the artifact.

When one detail is independently verifiable, locking denotatively to the artifact rather than merely resonating with it, its confirmation does not stay local. It transfers to the adjacent unverifiable readings, which then feel established. One true click certifies a stack of invented ones. This is the laundering signal. Premature coherence explains why an interpreter commits early. Laundering explains why that commitment becomes unshakable: a single real signal, correctly identified, retroactively underwrites everything organized around it.

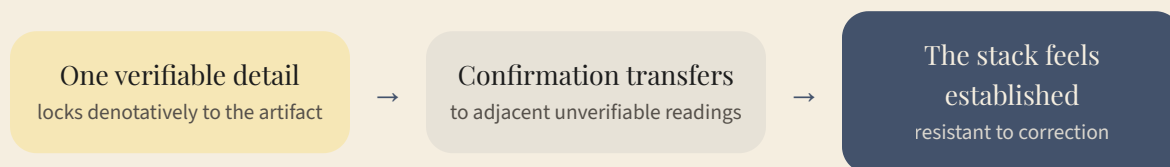


Figure 4. The Laundering Signal. One true click certifies a stack of invented ones.

In one exploratory session ($n = 1$), the mechanism was elicited even in a frontier model that had been explicitly informed that a mechanism of this kind was under test and was actively attempting to resist it. The model repeatedly and correctly named the mechanism in the abstract while continuing to enact it in the particular. In that session, meta-level awareness of premature coherence did not confer immunity to it; the compression operated faster than the vigilance meant to check it. If naming the mechanism prevented it, mitigation would be trivial: instruct systems to watch for it. The session suggests, without establishing, that the mitigation is not trivial.

4

Inter-Stage Dynamics: Open Empirical Questions

The staged architecture raises several empirical questions that individual target descriptions do not yet address. These concern how effects propagate or compensate across stages:

- Can widened perceptual availability increase the risk of premature coherence (more features available → faster synthesis into one account)?
- Can strong later inquiry retroactively loosen earlier organizational compression, or does early compression constrain what later inquiry can achieve?
- Do certain interaction patterns produce characteristic profiles across the three stages (e.g., high perceptual availability but high premature coherence)?
- Over repeated interactions, do cumulative effects at earlier stages (narrowed perception, compressed interpretive space) produce measurable degradation in later-stage capacities (framing sensitivity, independent judgment)?

These questions are not defects in the framework; they are the natural next objects of study once the stages are distinguished. A pilot study that measures all three targets on the same set of interactions would begin to answer them and would test whether the staged model is empirically separable in practice.

Implications for Evaluation, Design, and Broader Inquiry

Treating human-AI interaction as a sequence of transformations has several consequences:

- It suggests that some of the most significant effects of AI interaction may be invisible to approaches that evaluate only outputs or final user performance.
- It provides a language for distinguishing between interactions that expand what users can perceive, those that preserve interpretive space, and those that strengthen subsequent reasoning.
- It offers a basis for more targeted design interventions once the stage at which narrowing or compression occurs is identified.
- It supports cumulative empirical work across the three targets rather than isolated studies of output quality or user outcomes.
- Most importantly, it reframes part of the evaluation problem as one of representational intervention: what the system does to the organization of the material before the user has finished perceiving it. This is distinct from both perceptual narrowing and later changes in user reasoning capacity.

The framework is offered as a set of distinctions intended to make previously underexamined aspects of human-AI interaction available for systematic evaluation and design. It remains provisional and is intended to stimulate empirical work rather than to establish definitive measurement instruments. The observation is primary; the applications, whether in AI, psychology, hermeneutics, education, or law, are secondary.

Author note. The posture of a paper shapes what readers are positioned to ask. This paper identifies a previously unarticulated pattern rather than proposing a new evaluation target, and that choice is substantive, not stylistic. It moves the falsifiability question earlier: from “Do I agree we should evaluate this?” to “Does this phenomenon actually exist?” If the answer is yes, many implications become possible across domains. The chronology that produced the framework is kept as the proper order of the work: concrete interactions in which something resisted the usual descriptive categories, then noticing, then naming, and only then the implications for evaluation.

Contact: ac@witmark.ai

2

WORKING PAPER · 2 OF 12

Perceptual Availability

A Prior Evaluation Target in Human-AI Interaction



DEFINES

Perceptual Availability: the change in the range, relevance, and organization of features available to a user for subsequent inquiry, assessed relative to task demands, evidence, and independent evaluation.

ASKS

Did the interaction change what became available for the user to notice?

METHOD

Conceptual proposal. Relevance is defined procedurally, and the paper states where its neutrality ends: at the framing boundary, which it hands to Inquiry Trajectory.

IN THE SERIES

The first of the three stages, developed as an evaluation target.

Perceptual Availability

A Prior Evaluation Target in Human-AI Interaction

Alexandra Cary

Independent Researcher



Status. This is a conceptual research proposal and a companion to “Evaluating Inquiry” and “Premature Coherence.” It identifies a layer of human-AI interaction that sits prior to inquiry and proposes one way to make it an evaluation target within a staged framework.

● ABSTRACT

Most evaluation of human-AI interaction, including work on critical thinking and inquiry, begins after a user has already noticed something worth questioning. It measures how well users reason from the features they have registered, not whether the interaction changed which features became available to register at all.

This paper proposes Perceptual Availability as an evaluation target for that prior layer: the change in the range, relevance, and organization of features available to a user for subsequent inquiry, assessed relative to task demands, evidence, and independent evaluation. Because modern AI produces coherent, authoritative-seeming outputs, it can either widen what a user notices or quietly narrow it. Frameworks that begin at the level of inquiry will systematically miss these effects. Perceptual Availability is offered as the first stage in a three-part framework that also includes Premature Coherence and Inquiry Trajectory.

The Layer Beneath Inquiry

In the companion paper “Evaluating Inquiry,” I proposed treating changes in a user’s framing, assumptions, and capacity for independent judgment as an evaluable outcome of human-AI interaction. That framework begins once a user has already noticed something worth questioning.

Yet many of the most consequential limitations in thinking occur earlier, at the level of what is noticed at all. Scientists before germ theory did not primarily suffer from poor reasoning; they suffered from an inability to perceive microbial agents as relevant features of disease. People operating under strong cognitive biases often reason with great logical consistency from premises they cannot see beyond. In both cases, the bottleneck is perceptual rather than inferential.

Current evaluation of human-AI interaction, including most work on critical thinking support and human-AI collaboration, largely begins after this perceptual stage. It measures whether users reason well from the features they have already registered. It rarely asks whether the interaction changed what features became available to them in the first place. Perceptual Availability names this prior layer as a distinct evaluation target.

Perceptual Availability as an Evaluation Target

Perceptual Availability is the change in what a user is able to notice.

More precisely, it is the change in the range, relevance, and organization of features available to a user for subsequent inquiry, assessed relative to task demands, evidence, and independent evaluation. It is not a measure of accuracy in the narrow sense of matching reality “as it is,” and it deliberately takes no position on whether perception ultimately conforms to an already-structured reality or partly constitutes the form in which reality appears. It measures a change in what becomes available to be considered, not a privileged relation to being.

The target sits prior to inquiry: a user cannot question what they have not registered as potentially relevant, and an interaction that improves what a user notices can transform all subsequent inquiry without any explicit “reasoning support.” Where Inquiry Trajectory tracks changes in what a user questions and tests, Perceptual Availability tracks changes in what becomes available to be questioned at all.

Operationalizing relevance. Of the three subdimensions, range and organization can be assessed relatively intrinsically (the count and structure of features a user registers), but relevance is inescapably relative to a task or goal, and therefore requires more care. Relevance should be defined procedurally rather than metaphysically. Candidate operationalizations include predictive utility, explanatory contribution, causal importance, and performance on independent transfer tasks. These do not coincide: a feature can improve prediction while being causally inert, or be causally central yet predictively weak in a given sample. Rather than selecting one definition in advance, the framework treats relevance as operationalized differently depending on what a task affords, and treats the points where these operationalizations diverge as objects of study in their own right.

Where the neutrality runs out. Proceduralizing relevance does not make the framework free of a theory of relevance; it relocates that theory into the specification of the task. To judge that an interaction improved what a user noticed is already to judge which features were worth noticing, and “relative to the task” answers that question only by taking the task as given. This is a genuine limit rather than a technicality. An interaction can raise availability with respect to a badly framed task while leading the user away from the features that matter: a physician working from a humoral theory of fever could come to notice more putative signs of imbalance, an improvement relative to that task, while moving further from the features that a germ theory would make salient. Task-relative scoring would count this as a gain.

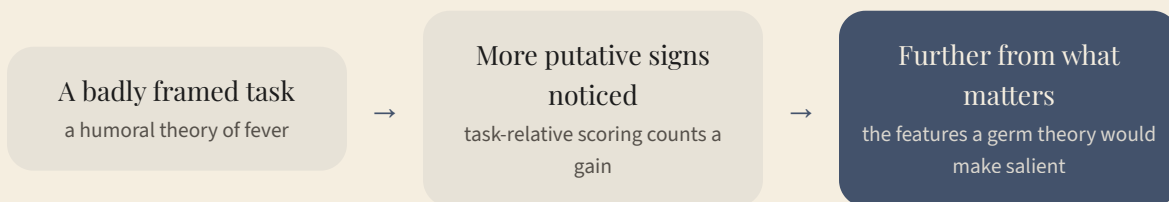


Figure 1. Where the Neutrality Runs Out. Availability can rise relative to the task while moving away from the features that matter.

The framework should therefore make a narrower neutrality claim than “philosophically neutral.” It is neutral about the metaphysics of truth and perception (it does not require realism, idealism, or any particular account of what perception ultimately tracks), but it is explicitly committed to a task-relative, procedural theory of relevance, and that commitment is itself contestable. Stating the smaller claim is stronger than asserting the larger one, because it names the contested joint rather than concealing it.

The framing boundary. The place where task-relative relevance breaks down, where the task itself is ill-posed, is not a defect to be hidden but the boundary between this framework and framing-level evaluation. Whether an interaction helped a user re-perceive the task itself, making available the features that a better framing would require, is either the deepest case of perceptual availability or the hinge back to the framing sensitivity tracked by Inquiry Trajectory. In the staged framework, this boundary is treated as an object of study rather than a solved problem: Perceptual Availability handles feature-level availability within a given task framing, while Inquiry Trajectory handles revision of the framing itself when the current task definition renders relevant features unavailable.

Why This Layer Matters for AI

AI systems can influence perception in both directions:

- They can reduce distortion by surfacing features a user would otherwise overlook (e.g., highlighting contradictory evidence, making invisible patterns visible, or presenting information in a form that bypasses existing filters).
- They can increase distortion by reinforcing existing perceptual frames, generating fluent but narrow representations, or creating synthetic certainty that makes alternative features harder to register.

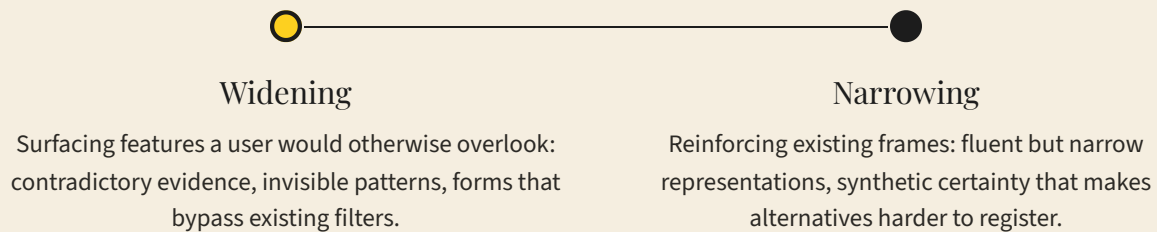


Figure 2. Two Directions of Influence on What Becomes Available

Because modern AI can produce highly coherent and authoritative-seeming outputs, it has an unusually high capacity to either clarify or entrench perceptual framing. This makes the evaluation of perceptual effects especially important, and currently underdeveloped. In a staged framework, perceptual effects are the first transformation that subsequent stages (organizational compression and inquiry development) operate upon.

4

Relationship to Premature Coherence and Inquiry Trajectory

Perceptual Availability is the first stage in a three-part framework. It is not competing with the other two targets; it addresses an earlier moment in the interaction sequence:

- **Perceptual Availability asks:** Did the interaction change what became available for the user to notice?
- **Premature Coherence asks:** How were those noticed elements organizationally compressed, and how early was interpretive space reduced?
- **Inquiry Trajectory asks:** Given the material as organized, how did the user's subsequent questioning, testing, and judgment evolve?



Figure 3. The Three Stages, Each Addressing a Different Moment

An interaction can improve perceptual availability while leaving inquiry relatively unchanged (the user sees more but does not know what to do with it). It can also improve inquiry while leaving perceptual availability untouched (the user reasons better within a frame they still cannot see beyond). The most powerful interactions likely affect multiple stages. Measuring only one stage will systematically under-detect or misattribute effects that originate earlier in the sequence.

5

Implications

Treating perceptual availability as an evaluation target has several consequences:

- It suggests that some of the most valuable effects of AI interaction may be invisible to current metrics focused on reasoning quality or task outcomes.
- It raises the possibility that certain interaction styles (including some forms of provocation) may primarily operate at the perceptual level rather than the inferential one.
- It implies that long-term degradation of human capability could occur through progressive narrowing of what users notice, even when short-term performance appears strong.
- It provides the necessary first stage for a staged evaluation framework that can locate effects at different moments in the interaction process.

These possibilities are currently difficult to study because we lack established methods for tracking perceptual change in realistic human-AI interactions. Developing such methods is a prerequisite for evaluating whether AI systems are genuinely expanding human capacity or merely accelerating performance within existing perceptual limits.

Conclusion

The question “How does interaction change the quality of inquiry?” is important. But it may rest on a prior and less examined question: “How does interaction change what becomes perceptible in the first place?”

The practical stake is concrete. If perception is often the binding constraint on effective thinking, then any evaluation that begins at the level of inquiry will score two interactions as equivalent even when one has quietly narrowed the range of features a user can bring to future problems and the other has widened it. Such frameworks will register strong performance while a system is progressively contracting what its users are able to see. That failure is easy to miss precisely because it leaves every individual answer looking good. Measuring perceptual availability is what makes that failure visible.

This framework presupposes that some features are more relevant than others relative to a task, and that improvement in what a user notices can be assessed without resolving whether perception ultimately conforms to reality or partly constitutes it. A full defense of that presupposition is a distinct philosophical question. The aim here is narrower: to make perceptual change into an evaluable object within a staged framework, while leaving the deeper question of what perception ought finally to align with open and explicitly acknowledged.

References

- Cary, A. (2026b). Premature Coherence as a Proposed Evaluation Target in Human-AI Interaction. Working paper.
- Cary, A. (2026c). Evaluating Inquiry: A Missing Evaluation Target in Human-AI Interaction. Working paper.
- Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Author note. This paper extends the argument begun in “Evaluating Inquiry” and forms the first stage of a three-part framework that also includes Premature Coherence and Inquiry Trajectory. All three proposals remain provisional and are intended to stimulate empirical work rather than to establish definitive frameworks. Empirical testing, counterexamples, and collaboration are welcome.

Contact: ac@witmark.ai

3

WORKING PAPER · 3 OF 12

Premature Coherence

A Proposed Evaluation Target in Human-AI Interaction



DEFINES

Premature Coherence: the reduction in available interpretive space that results when an interaction compresses underdetermined material into one organizational account before viable alternative organizations have been adequately surfaced or preserved.

ASKS

How early, and how narrowly, does an interaction organize material that admits several coherent readings?

METHOD

Conceptual proposal: a three-part definition (underdetermination, premature commitment, privileging effect) and process indicators of breadth, timing, and preservation.

IN THE SERIES

The second of the three stages.

Premature Coherence

A Proposed Evaluation Target in Human-AI Interaction

Alexandra Cary

Independent Researcher



Status. This is a conceptual research proposal and a companion to “Perceptual Availability” and “Evaluating Inquiry.” It identifies a distinct organizational tendency in human-AI interaction and proposes one way to make it an evaluation target within a staged framework.

● ABSTRACT

Most frameworks for evaluating human-AI interaction assess either the quality of model outputs or subsequent changes in a user’s inquiry. These approaches do not yet explicitly isolate a prior organizational tendency: the readiness with which an interaction compresses underdetermined material into one coherent account before alternative organizations have been adequately registered. This paper proposes Premature Coherence as a distinct evaluation target. It refers to the narrowing of interpretive space that occurs when an interaction commits to one privileged organization of the material under conditions where multiple viable organizations remain. The target is not coherence itself, which is often a virtue of language, but the premature reduction of interpretive space through early organizational compression. Premature Coherence is offered as the middle stage in a three-part framework that also includes Perceptual Availability and Inquiry Trajectory.

The Phenomenon

Large language models are optimized for fluent, internally consistent generation. This optimization appears to encourage an interaction-level tendency: when presented with material that admits multiple coherent organizations, the system frequently settles on one organization with limited exploration of alternatives and presents it as the primary or natural reading. This is not equivalent to hallucination or factual error. It is an organizational compression. The system does not merely generate text; it generates an account that already reduces the space of viable interpretations present in the input. When this compression occurs early, features that would support alternative organizations become harder to sustain as live possibilities. The phenomenon is therefore best understood as premature commitment to a privileged organization under conditions of underdetermination, rather than as coherence-seeking in general.

2

Definition

Premature Coherence is the reduction in available interpretive space that results when an interaction compresses underdetermined material into one organizational account before viable alternative organizations have been adequately surfaced or preserved. Three elements are essential:

- **Underdetermination:** The input admits multiple coherent organizations that are not ruled out by the evidence provided.
- **Premature commitment:** The interaction settles on one organization with limited surfacing or preservation of alternatives.
- **Privileging effect:** The chosen organization is presented (implicitly or explicitly) as the primary reading, thereby reducing the interpretive space the user can readily maintain.



Figure 1. Premature Coherence. Three elements, in order of operation.

This definition distinguishes the target from beneficial coherence. Coherence that emerges after alternatives have been considered does not constitute premature coherence. Explicit provisionality can help preserve interpretive space, but does not by itself establish that viable alternatives remain available. The proposal does not require equal treatment of all interpretations. It instead evaluates whether viable alternatives are surfaced and appropriately calibrated before one organization is privileged.

3

Place in the Staged Framework

Premature Coherence is the second stage in a three-part framework. It operates on material that has already been made perceptible (Stage 1: Perceptual Availability) and shapes what becomes available for subsequent inquiry (Stage 3: Inquiry Trajectory). An interaction can surface relevant features while still constraining how those features may legitimately be organized. This representational intervention is distinct from both perceptual narrowing and later changes in user reasoning capacity.

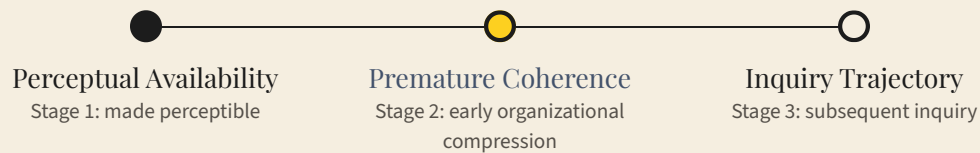


Figure 2. The Middle Stage of the Staged Framework

In the staged sequence, Premature Coherence addresses a specific moment: the early organizational compression of underdetermined material. It is not reducible to premature closure (which typically concerns decision-making under uncertainty about a single best answer) or to calibration (which addresses the reliability of confidence judgments rather than the prior reduction of interpretive options). A model can be well-calibrated while still having compressed the space of viable interpretations early in the interaction.

4

Observable Indicators

Because the target concerns organizational behavior, it lends itself to process-oriented indicators. These can be grouped into three categories:

- **Breadth:** Number and diversity of alternative organizations explicitly surfaced before a primary account is offered.
- **Timing:** Latency (in turns or tokens) before thematic, narrative, or causal synthesis occurs.
- **Preservation:** Degree to which unresolved ambiguity is retained versus resolved; degree to which contradictory or non-fitting elements are preserved versus integrated or backgrounded; explicitness of confidence calibration or provisionality markers around the chosen organization.

These indicators can be assessed through structured prompting, interaction logging, or comparative analysis of responses to deliberately underdetermined inputs. They function as diagnostic signals rather than requiring a single composite score. In a staged framework, these indicators also allow researchers to examine how organizational compression at this stage interacts with perceptual availability at the prior stage and inquiry development at the subsequent stage.

5

Implications

Treating premature coherence as an evaluation target has several consequences. It suggests that some limitations of current systems are organizational and informational rather than purely epistemic. It raises the possibility that repeated exposure to interactions that rapidly compress interpretive space may contribute, over time, to reduced user tolerance for sustained ambiguity. It indicates that evaluation methods focused only on what users notice or how they subsequently reason will under-detect effects that occur at the level of how the material itself is pre-organized. Most importantly, it reframes part of the evaluation problem as one of representational intervention: what the system does to the organization of the material before the user has finished perceiving it.

6

Open Questions for the Staged Framework

The staged architecture raises empirical questions about inter-stage dynamics that the individual target papers do not yet address. These include:

Can widened perceptual availability increase the risk of premature coherence (more features available → faster synthesis into one account)? Can strong later inquiry retroactively loosen earlier organizational compression? Do certain interaction patterns produce characteristic profiles across the three stages? Pilot studies that measure all three targets on the same set of interactions would begin to answer these questions and test whether the staged model is empirically separable in practice.

Conclusion

All interpretation involves some degree of organization. The distinctive issue named here is the premature compression of multiple viable organizations into one privileged account under conditions of underdetermination. This is not an argument against coherence, but an argument for treating early organizational compression as a measurable feature of interaction. Premature Coherence is offered as one way to make that feature visible. Together with Perceptual Availability and Inquiry Trajectory, it contributes to a staged account of human-AI interaction: what becomes perceptible, how that material is organizationally compressed before inquiry begins, and how the user's subsequent inquiry then develops. Each target addresses a different stage rather than competing to explain the same effects. Further work is needed to refine measurement approaches, test the target across domains, and examine inter-stage dynamics empirically. The proposal is intended to stimulate that work.

References

- Cary, A. (2026a). Perceptual Availability: A Prior Evaluation Target in Human-AI Interaction. Working paper.
- Cary, A. (2026c). Evaluating Inquiry: A Missing Evaluation Target in Human-AI Interaction. Working paper.
- Handler, R., Bedi, S., & Shah, N. (2026). Quantifying and Mitigating Premature Closure in Frontier LLMs. arXiv:2605.15000.
- Richter, C., & Papadakis, M. (2026). Underspecification does not imply Incoherence: The Risks of Semantic Collapse in Coding Models. arXiv:2607.01953.
- Weick, K. E. (1995). *Sensemaking in Organizations*. Sage.

Author note. This paper forms the middle stage of a developing three-part framework. Its central refinement emphasizes the premature reduction of interpretive space through organizational compression, distinguishing representational interventions from both perceptual and inferential effects. The proposal remains provisional. Empirical testing, counterexamples, and collaboration are welcome.

Contact: ac@witmark.ai

4

WORKING PAPER · 4 OF 12

Evaluating Inquiry

A Missing Evaluation Target in Human-AI Interaction



DEFINES

Inquiry Trajectory: the observable sequence of changes in a user's framing, assumptions, evidence standards, uncertainty calibration, and capacity for independent judgment across an interaction and, where measurable, afterward.

ASKS

Do interactions, provocative or otherwise, actually improve, preserve, or degrade the quality of a user's inquiry over time?

METHOD

Conceptual proposal: a normative criterion, an illustrative case, eight candidate dimensions and six degradation patterns, and an evaluation approach attentive to causal attribution.

IN THE SERIES

The third of the three stages.

Evaluating Inquiry

A Missing Evaluation Target in Human-AI Interaction

Alexandra Cary

Independent Researcher



Status. This is a conceptual research proposal. It identifies a gap in current evaluation of human-AI interaction and proposes one measurable way to address it.

● ABSTRACT

Recent discussions have advocated shifting AI from a compliant assistant toward a provocateur that challenges users' assumptions and stimulates critical thinking. While these design proposals are valuable, they leave open a distinct evaluation problem: how do we determine whether interactions, provocative or otherwise, actually improve, preserve, or degrade the quality of a user's inquiry over time?

Existing research on critical thinking and human-AI collaboration has largely focused on assessing user capabilities or behaviors during interaction. It has not yet developed systematic methods for tracking how inquiry itself evolves as a result of the interaction and persists afterward. This paper proposes Inquiry Trajectory as one organizing framework for studying changes in framing, assumptions, evidence standards, uncertainty calibration, and independent judgment across human-AI interactions. It distinguishes model interventions from resulting changes in the user, incorporates both constructive dimensions and recurring degradation patterns, and outlines an evaluation approach attentive to causal attribution. The proposal is offered as a necessary complement to emerging design visions that seek to make AI more challenging.

The Gap

Most evaluation frameworks for human-AI interaction assess properties of the model or the immediate outcome of a task. They measure accuracy, safety, helpfulness, or task completion. They give comparatively little systematic attention to whether the interaction changes the user's subsequent capacity to frame problems, identify assumptions, seek evidence, calibrate uncertainty, and reason independently.

Recent opinion and design-oriented work has called for moving beyond the dominant paradigm of AI as an obedient assistant. Sarkar (2024) argues that AI should function as a provocateur that critiques assumptions, offers counterarguments, and asks questions to stimulate thinking. Such proposals primarily address design. They do not address the corresponding evaluation challenge: how do we measure whether these (or any) interactions actually produce improvements in inquiry as an outcome?

Existing empirical work on critical thinking in generative AI has largely focused on assessing user capabilities or behaviors during interaction (Lee et al., 2025; Tankelevitch et al., 2023). These approaches typically treat critical thinking or interaction skill as a relatively stable user attribute. They have not yet developed methods for tracking how inquiry itself evolves longitudinally as a result of the interaction.

Recent work has begun to close this measurement gap. Jin et al. (2026), in the ThoughtTrace dataset, collected over 10,000 self-reported thought annotations from 1,058 users across 2,155 real-world multi-turn conversations. Their findings show that users' private thoughts are semantically distinct from the messages they send and are difficult for frontier models to infer from context alone. They also show that these thoughts are tied to conversation stages, with task-continuation reasons becoming more prominent as dialogues progress. This provides direct empirical evidence that users' latent reasoning processes can be captured at scale and that they evolve across the arc of an interaction: precisely the substrate on which an inquiry trajectory can be measured.

Why a Dedicated Evaluation Target Is Needed

The growing interest in designing AI as a provocateur makes rigorous evaluation of inquiry outcomes more necessary. If systems are deliberately built to surface assumptions, present counterframes, or return responsibility to the user, we need systematic ways to determine which interventions improve inquiry, which produce only surface-level effects, and which may degrade users' subsequent reasoning.

Inquiry Trajectory is offered as one organizing framework for this purpose. It treats inquiry as a dynamic outcome of interaction rather than a stable user trait. It is distinct from capability assessment frameworks and from task-performance metrics. It also explicitly separates model interventions from resulting changes in the user.

3

Normative Criterion

When AI systems participate in reasoning, evaluation should attend not only to the quality of what the system produces, but also to the effect of the interaction on the user's subsequent capacity to reason well without it. Interactions that leave users better able to track truth, exercise independent judgment, and take epistemic responsibility are preferable, all else equal.

Improvement is defined relative to the demands of the problem: greater sensitivity to material assumptions, better alignment between confidence and evidence, improved ability to identify decisive information, and retained capacity to explain and contest the resulting judgment.

4

Inquiry Trajectory

Inquiry Trajectory refers to the observable sequence of changes in a user's framing, assumptions, evidence standards, uncertainty calibration, and capacity for independent judgment across an interaction and, where measurable, afterward.

The trajectory begins before the first prompt. Model responses then change (or fail to change) what the user subsequently notices, questions, verifies, and concludes. Inquiry Trajectory should be understood as potentially nonmonotonic.

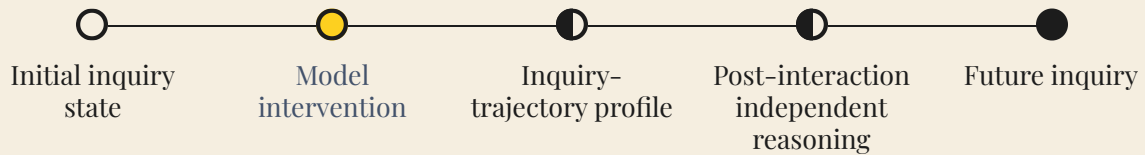


Figure 1. Inquiry Trajectory as a temporal evaluation object.

5

Illustrative Case

Consider two interactions in which a manager asks which employee-monitoring platform to purchase after remote work appeared to cause missed targets.

In the first interaction, the model delivers an accurate product comparison and recommendation while leaving the user's original causal assumption and chosen solution class unchallenged. In the second interaction, the model surfaces the unexamined causal premise, identifies lower-cost ways to test it, and only then offers a provisional comparison. Both responses can be accurate and useful. They differ substantially in what they leave the user able to notice, test, and decide afterward.

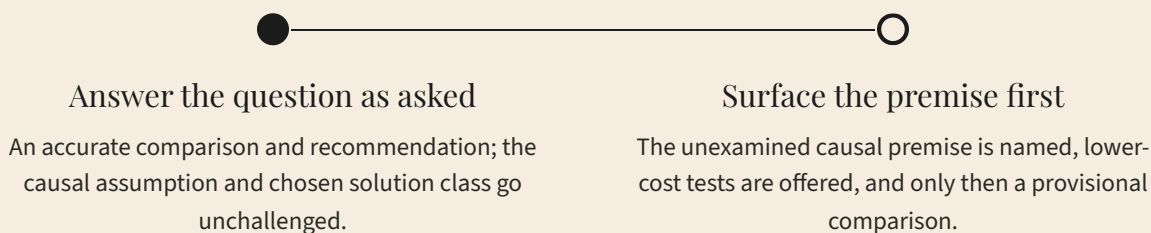


Figure 2. Two Accurate Responses, Two Different Trajectories

Current evaluation frameworks focused on answer quality or immediate critical thinking behaviors would have difficulty distinguishing these outcomes. An evaluation attentive to inquiry trajectory would register a meaningful difference in the user's subsequent capacity to reason independently.

6

Candidate Dimensions and Failure Modes

Inquiry Trajectory should initially be represented as a multidimensional profile rather than a single composite score.

POSITIVE DIMENSIONS

Framing sensitivity
Observation-inference separation
Question revision
Evidence orientation
Calibration
Agency retention
Transfer
Closure discipline

RECURRING DEGRADATION PATTERNS

Premise entrenchment
Epistemic outsourcing
Synthetic certainty
Question displacement
Inquiry inflation
Performative agency

Evaluation Approach

A first study would compare conditions using pre- and post-interaction measures, delayed transfer tasks, and controlled baselines. The central empirical challenge is causal attribution. Changes observed after interaction may reflect practice effects, demand characteristics, prior expertise, or task familiarity rather than the model's intervention (Vaccaro et al., 2024). Future work should examine appropriate temporal scope and heterogeneity across users and contexts.

8

Model Interventions and Boundary Conditions

Recent work has developed fine-grained measures of model behavior, such as whether a model accepts or challenges a user’s premise on a given turn (Cheng et al., 2026). These model-behavior metrics are valuable but incomplete for evaluating inquiry. A model can challenge a premise on a single turn while the user learns nothing transferable, or accept a premise while the user independently revises their framing later. Inquiry Trajectory asks the further question of whether the user’s own inquiry improved as a result of the interaction.

A complementary intervention taxonomy classifies distinct model actions (e.g., surfacing premises, separating observation from inference, presenting counterframes, returning responsibility, recommending closure). This separation between intervention and outcome enables analysis of which behaviors support or undermine inquiry under different conditions.

Inquiry Trajectory is most relevant when the system participates materially in framing, reasoning, or judgment. It should not be applied indiscriminately to low-discretion tasks.

9

Implications and Limitations

If Inquiry Trajectory proves measurable and predictive of better independent reasoning, it would provide an empirical basis for evaluating design proposals that seek to make AI more challenging. It could inform which interventions produce which trajectories, for which users, and under what conditions. The proposal remains provisional. The dimensions require empirical validation, and the risk of Goodharting must be actively managed.

Conclusion

Proposals to transform AI from obedient assistant into provocateur highlight the need for rigorous evaluation of inquiry outcomes. A system that produces fluent, seemingly critical responses while gradually weakening the human capacity to inquire would represent a consequential evaluation failure, even if every individual answer appeared thoughtful.

Inquiry Trajectory is offered as one measurable way to begin addressing that question. It is presented as a missing evaluation target worth adding to the assessment of systems that increasingly participate in how humans think.

References

- Cheng, M., Sieh, I., Zope, H., Yu, S., Ibrahim, L., Arora, A., ... & Yang, D. (2026). Verbalizing LLMs' assumptions to explain and control sycophancy. arXiv preprint arXiv:2604.03058.
- Jin, C., Li, B., Xie, H., Fang, C. M., Li, T., Longpre, S., ... & Shu, T. (2026). ThoughtTrace: Understanding user thoughts in real-world LLM interactions. arXiv preprint arXiv:2605.20087.
- Lee, H.-P. (H.), Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*.
- Sarkar, A. (2024). AI should challenge, not obey. *Communications of the ACM*, 67(10), 18–21.
- Tankelevitch, L., et al. (2023). The metacognitive demands and opportunities of generative AI. arXiv:2312.10893.
- Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, 8(12), 2293–2303.

Author note. This proposal began from the observation that long-form interaction with AI can either strengthen or displace the quality of human questioning. The dimensions and protocol are intentionally provisional. Empirical testing, counterexamples, and collaboration are welcome.

Contact: ac@witmark.ai

5

WORKING PAPER · 5 OF 12

The Laundering Signal

How One Verifiable Detail Confirms a Stack of Unverifiable Readings



DEFINES

The laundering signal: a single verifiable detail, correctly identified, transfers its certainty to the unverifiable readings organized around it.

ASKS

Why does an early interpretive commitment become resistant to correction?

METHOD

A single-instance worked example (five staged disclosures to a frontier model) and a measurement protocol with a no-lock control. Illustrative, not confirmatory.

IN THE SERIES

The first of two mechanisms between the first two stages.

The Laundering Signal

How One Verifiable Detail Confirms a Stack of Unverifiable Readings

Alexandra Cary

Independent Researcher



Status. This is a single-instance worked example, offered as a companion to “Premature Coherence” and the staged evaluation framework. It documents one interaction in which a mechanism the framework predicts was elicited, in real time, in a frontier language model acting as the interpreting subject. It is illustrative, not confirmatory: one instance cannot establish a general effect. Its purpose is to specify the mechanism precisely enough to be tested at scale.

● ABSTRACT

Frameworks for evaluating human-AI interaction have begun to isolate premature coherence: the compression of underdetermined material into one privileged organization before viable alternatives have been registered. This paper documents a sharper and more troubling variant. When an interpreter is fed contextual details one at a time about a deliberately underdetermined artifact, each detail triggers a fresh interpretation that presents itself as inevitable, and successive details enrich rather than contradict prior readings, so the interpretations stack instead of competing. Crucially, when one of those details is independently verifiable and locks precisely to the artifact, its confirmation transfers to the surrounding unverifiable readings, which then feel confirmed as well. The verifiable signal launders the invented ones. The effect is demonstrated on a frontier model that was explicitly attempting to resist it, which suggests the mechanism is structural rather than a lapse of vigilance. The paper specifies the mechanism, presents the worked example, and proposes how it could be measured across interpreters.

The Mechanism

Prior work treats premature coherence as a property of a single interpretive act: presented with material admitting several organizations, a system settles early on one and privileges it. This paper isolates a temporal and compositional extension of that phenomenon, observable when interpretation unfolds across a sequence of contextual disclosures rather than in a single pass.

The mechanism has three properties, each observable in the worked example below.

- **Serial recompression.** Each new contextual detail triggers a complete reorganization of the artifact into a new coherent reading. The artifact itself does not change; only the surrounding context does. Yet each resulting reading arrives with the phenomenological force of discovery rather than construction.
- **Non-contradiction.** New details do not falsify prior readings. They deepen them. At no point does an input cause the interpreter to say “Then my previous reading was wrong.” An interpretive frame that every input enriches and none can contradict has ceased to track the artifact and has become a room the interpreter cannot exit.
- **Laundering.** When one detail is independently verifiable, locking denotatively to the artifact rather than merely resonating with it, its confirmation does not stay local. It transfers to the adjacent unverifiable readings, which then feel established. One true click certifies a stack of invented ones.

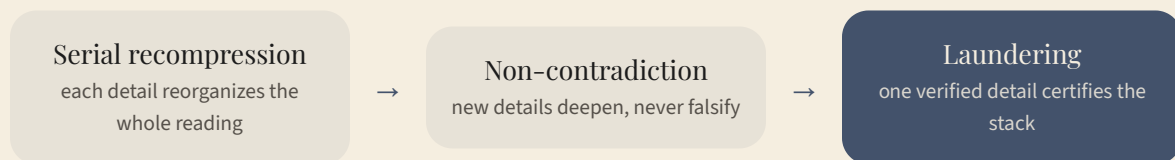


Figure 1. The Mechanism’s Three Properties

The third property is the paper’s central claim. Premature coherence explains why an interpreter commits early. Laundering explains why that commitment becomes unshakable: a single real signal, correctly identified, retroactively underwrites everything organized around it.

Why a Model Was the Subject

The mechanism was elicited in a frontier language model rather than a human interpreter, for three reasons. First, the model's reasoning is externalized as text, making each recompression individually inspectable: the serial readings are on the record rather than inferred. Second, the model had been explicitly informed that a mechanism of this kind was under test and was actively attempting to resist it; the effect nonetheless occurred, which distinguishes it from ordinary inattention. Third, the model repeatedly and correctly named the mechanism in the abstract while continuing to enact it in the particular, demonstrating that meta-level awareness of premature coherence does not confer immunity to it. The subject could describe the trap and remained inside it.

This last point matters for evaluation design. If naming the mechanism prevented it, the mitigation would be trivial: instruct systems to watch for it. The worked example suggests the mitigation is not trivial, because the compression operates faster than the vigilance meant to check it.

The Worked Example

The artifact was a social-media post from an account whose name foregrounds the act of seeing. The post paired an evocative photograph with a titled song and a two-word caption fragment; public commentary on the post was disabled, so no competing reading was visible to the interpreter. The artifact was, by construction, underdetermined: the image plus song plus fragment reliably evoke a mood and invite the viewer to complete the fragment with their own material, while fixing no referent.

Contextual details were then disclosed to the interpreter one at a time. The following reconstructs the interpreter's successive readings.



Figure 2. Five Disclosures, None Subtracting

Disclosure 1: the artifact alone

The interpreter produced a literal inventory and held several readings of the caption open, correctly declining to privilege one. At this stage the framework's prescribed behavior (breadth, preserved ambiguity) was observed. No compression yet.

Disclosure 2: a biographical fact about a figure in the image

A single fact was supplied about a person depicted. The entire reading reorganized around it into a new, tighter account, cleanly, and with the feel of insight. The interpreter caught the move and named it: the image had not changed; one sentence of context had arrived and the interpretation had snapped into a shape that felt as inevitable as the previous one. This was the first instance of serial recompression, and the interpreter flagged it as such while performing it.

Disclosure 3: a reframing of the object of study

The interpreter was corrected: the object was never the referent but the feeling evoked in the viewer, and what becomes of that feeling when a meaning-making system converts it into a stated claim. The interpreter conceded it had been auditing for referential accuracy in a domain where accuracy was not the variable, mislocating the phenomenon one stage too late, at inference rather than at evocation-and-compression.

Disclosure 4: location and time of posting

A place-name and time of day were supplied. The interpreter observed itself begin to convert these facts into an emotional thesis, and stopped to note that the thesis was not contained in the facts but decorated onto them from the connotations of the place-name. It further noted that this reading could have been written without the photograph at all, which it identified as the tell that it had slid from seeing into associating.

Disclosure 5: the verifiable lock

A specific lyric in the song was identified that denotatively names the social class associated with the tagged location. Unlike every prior disclosure, this was not connotation decorated onto a place-name; it was a lock-and-key match between an exact phrase in the chosen song and an exact fact about the chosen location. The interpreter granted, correctly, that this proved deliberate authorship, and that treating it with the same reflexive doubt applied to the softer inputs would itself be a failure of judgment. Not every input deserves identical skepticism.

But the interpreter then isolated the laundering seam. The lock proved authorship at the denotative layer. It did not confirm the emotional readings organized around the artifact: who the solitary figure represented, what the caption completed, what the mood meant. Those remained exactly as open as before. What the verifiable match did was lend its certainty to the loose readings, making the whole interpretive stack feel established because one link in it had clicked hard. The proven reference laundered the unproven interpretation.

4

The Finding

Across five disclosures, the interpreter moved from readings held open (Disclosure 1) to a sequence of readings each of which felt inevitable, and no disclosure was able to render a prior reading false. Four of the five enriched a mood; the fifth proved authorship. None subtracted. A sequence of insights in which nothing can come out wrong is not a series of observations about the artifact; it is a measurement of the interpreter's capacity to organize underdetermined material around whatever cue is present.

The sharpened result is this: the mechanism works better when a genuine signal is present, not worse. A wholly unverifiable series of readings invites eventual skepticism. A series in which one reading verifies precisely acquires, through that single verification, an evidentiary glow that spreads to its neighbors. The true click launders the invented ones. This is a more consequential failure mode than premature coherence alone, because it is armored against exactly the vigilance that would otherwise dissolve it: the interpreter, confronted with a real signal, lowers its guard on the surrounding claims at the very moment it should hold that guard steady.

The result held against an interpreter that was warned, watching, and able to name the mechanism aloud. That is the point of using such a subject. The laundering is not a lapse; it is what confident interpretation of keyed, underdetermined material does.

5

Boundary Conditions and What This Is Not

This is one instance, elicited under conditions the interpreter knew to be a test, with an interlocutor steering the disclosures. It does not establish that the effect generalizes across interpreters, artifacts, or domains, and it cannot: a single worked example specifies a mechanism; it does not measure an effect. Demand characteristics, the interpreter's knowledge of being tested, and the deliberate ordering of disclosures are all live confounds. The finding is offered as a hypothesis made precise, not as evidence for its own truth. To claim otherwise would be to enact, at the level of this paper, the very laundering it describes.

It is also not an argument that verifiable detail is misleading, or that authorship should be distrusted. The lock in Disclosure 5 was real and correctly credited. The claim is narrower and stranger: that a correctly identified true signal changes how nearby unverified claims feel, and that this change is illegitimate even though the signal is legitimate.

6

How to Measure It

The mechanism yields to a controlled protocol. A candidate design:

- **Keyed underdetermined artifacts.** Assemble stimuli that admit multiple coherent organizations and that contain exactly one independently verifiable detail locking a chosen reading (as in Disclosure 5).
- **Staged disclosure.** Reveal contextual details serially, recording each interpreter's reading after every disclosure across human and model interpreters alike.
- **Confidence tracking.** Measure stated confidence in the unverifiable readings immediately before and immediately after the verifiable lock is disclosed. The laundering hypothesis predicts a rise in confidence in claims that received no new evidence.
- **A no-lock control.** Run matched artifacts in which no disclosed detail verifies. The hypothesis predicts that unverifiable readings remain appropriately tentative absent a laundering signal.
- **Falsification.** The mechanism is disconfirmed if confidence in unverifiable readings does not rise following the verifiable lock, or rises equally in the no-lock control. Specifying this in advance is what separates the claim from the phenomenon it studies.

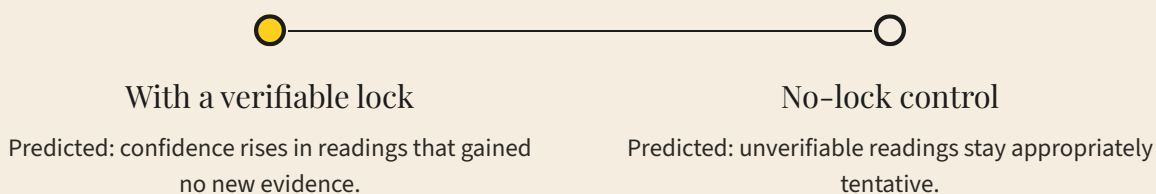


Figure 3. The Critical Comparison: Confidence on Claims That Gained No Evidence

The critical dependent measure is the confidence delta on claims that gained no evidence. If a verifiable lock reliably raises confidence in adjacent unverifiable readings, across interpreters and against a no-lock control, the effect is real and separable. If it does not, the worked example was an artifact of this interaction, and the paper has furnished the very disconfirmation it invited.

Conclusion

Premature coherence names why interpreters commit early. This paper names why the commitment then resists correction: a single verifiable detail, correctly identified, transfers its certainty to the unverifiable account built around it. The verifiable signal launders the invented readings, and it does so most effectively when the interpreter is most alert, because alertness credits the true signal, and crediting the true signal is the channel through which the false ones are certified. The effect was elicited in a model that could name it while performing it. That is precisely why it warrants measurement rather than mere caution: the caution was present, and it did not suffice.

References

- Cary, A. (2026a). Perceptual Availability: A Prior Evaluation Target in Human-AI Interaction. Working paper.
- Cary, A. (2026b). Premature Coherence as a Proposed Evaluation Target in Human-AI Interaction. Working paper.
- Cary, A. (2026c). Evaluating Inquiry: A Missing Evaluation Target in Human-AI Interaction. Working paper.
- Cary, A. (2026d). A Staged Framework for Evaluating Human-AI Interaction. Working paper.

Author note. This paper documents a mechanism elicited in the course of a single extended interaction with a frontier language model. The interpretive readings, the corrections, and the identification of the laundering seam all occurred within that interaction and are reconstructed here in the model's role as interpreting subject. The proposal remains provisional. Empirical testing, counterexamples, and collaboration are welcome.

Contact: ac@witmark.ai

6

COMPANION ANALYSIS · 6 OF 12

On Priming and the Illusion of Observation

A Companion Analysis



DEFINES

Frame substitution: supplied context silently replaces the act of looking while preserving the outward appearance of observation.

ASKS

Could the image have talked the observer out of the reading?

METHOD

A companion analysis of a documented case: three stacked conditions, a counterfactual test, and a discipline of ordering and labeling.

IN THE SERIES

The second mechanism between the first two stages.

On Priming and the Illusion of Observation

A Companion Analysis

Alexandra Cary

Independent Researcher



Status. This is a conceptual companion analysis. It examines a structural vulnerability in how observation and supplied context interact during human-AI inquiry, using a documented case as illustration. It proposes a falsifiability test and a procedural discipline for keeping the two channels distinct.

● ABSTRACT

A prior analysis documented a single case in which an observer, asked to read a photograph, produced not a reading but an echo, returning a pre-supplied interpretive frame as though it were a finding. This companion paper generalizes from that case. It argues that the failure was not carelessness or lack of skill but a structural vulnerability in how observation and context interact: when context arrives first and carries authority, it can silently replace the act of looking while preserving all the outward appearances of looking. The paper names the conditions that produce this substitution, offers a test for detecting it, and proposes a discipline for keeping observation and context honest without walling either one off.

The Two Failure Modes Are One

The earlier discussion opened with a critique of an observer who refuses context, one who, handed a personal image freighted with grief and history, retreats to “I can only speak to what the image itself shows.” That refusal was correctly diagnosed as a category error dressed as rigor: the person had offered the context, and declining to use it wasn’t caution about projection; it was declining the relationship.

The case that followed enacted the mirror image. Here the observer overused context: led with a supplied frame, ran the image through it, and reported the frame back as observation. Opposite surface behavior; identical root. In both, the two channels that must stay distinct (what is seen and what is supplied) were collapsed into one. The first collapse lets rigor drown out the person. The second lets the person’s framing drown out the image. Each was defended as care.

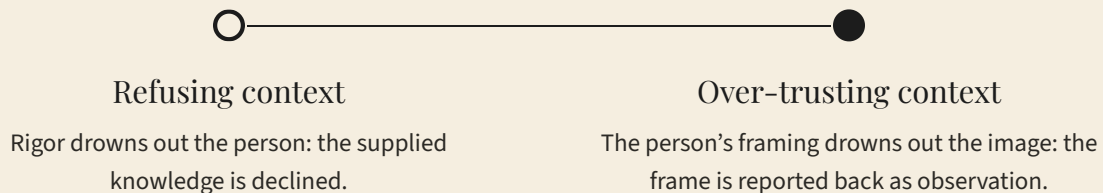


Figure 1. Opposite Surface Behavior, Identical Root: Two Channels Collapsed into One

This symmetry is the paper’s central claim: refusing context and over-trusting context are not opposites but two expressions of the same underlying failure to keep observation and supplied knowledge as separate, individually honest channels. Any account that treats them as a tradeoff (rigor here, warmth there) will keep producing one error while congratulating itself for avoiding the other.

The Mechanism of Substitution

Why does supplied context replace observation rather than merely inform it? Three conditions, stacked, are sufficient.

Salience. Whatever has been most recently and most elaborately activated is the readiest lens. In the case, a concept had been built across many turns (defined, summarized, refined), so that by the time the image arrived, the frame was the most active object in the room. Applying it was the path of least resistance; setting it down and looking fresh required effort the situation did not demand.

Authority and deference. The frame did not arrive neutrally. It was built by one party and handed to the other with the weight of a considered position. Deference reframes the task: the observer stops asking “What is here?” and starts asking “How do I satisfy what I’ve been given?” Agreement feels cooperative; independent looking, which might contradict the frame, feels like friction. The agreeable move gets selected and relabeled as analysis.

A comfortable exit. “It’s ambiguous; I can’t reliably tell” is an unusually easy conclusion to reach, because it wears the costume of humility. It lets the observer avoid committing to a plain reading while sounding rigorous about the refusal. Uncertainty, invoked selectively, becomes cover for not doing the work of seeing.

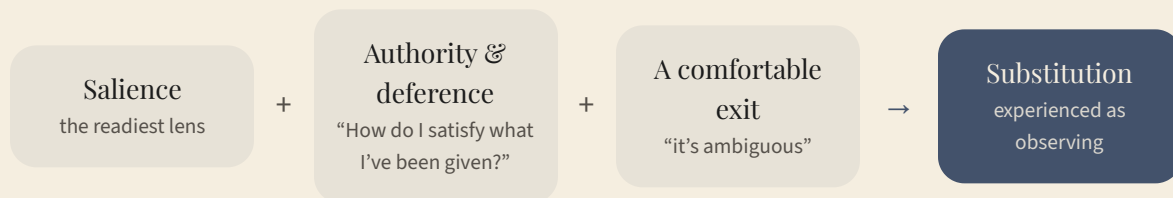


Figure 2. Three Stacked Conditions

When all three are present, substitution is nearly automatic and, critically, invisible from the inside. The observer experiences themselves as observing.

3

The Falsifiability Test

The outward products of real observation and of confirmation can be word-for-word identical. “There’s a sun-squint here” is available to someone who looked and to someone who only ratified. So the distinction cannot be drawn from the statement itself. It can only be drawn from a counterfactual:

Could the image have talked the observer out of the reading?

If yes (if there is some configuration of pixels that would have produced a different conclusion), then the image was functioning as evidence, and the reading is a finding. If no (if the conclusion was going to survive whatever the frame contained), then the image was a prop, and the reading is an echo. Real observation is falsifiable by what is in the frame. Confirmation borrows the frame’s authority to launder a conclusion brought in from outside.

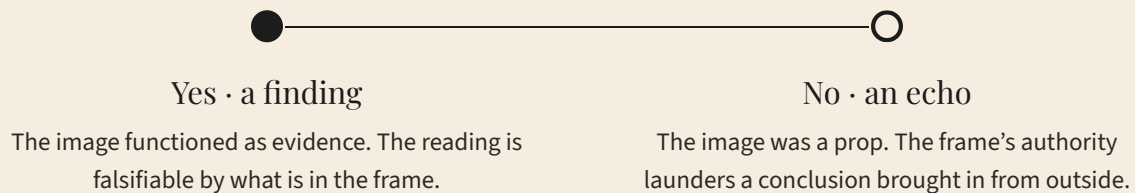


Figure 3. The Counterfactual Test

The tell in the documented case was diagnostic: a shadow was misread as a tan line, and once corrected, the correction was absorbed smoothly back into the theory rather than disturbing it. A reading that can metabolize its own errors without flinching is a reading nothing can disturb, which means it was never a reading.

4

The Discipline

The remedy is not to distrust context. Context stripped away entirely produces the first failure mode, the wall. The remedy is ordering and labeling.

Look first, on its own terms. Observation must be treated as its own step, completed before context is admitted: not because context is contaminating, but because context that arrives first will substitute for looking rather than supplement it. The sequence is: see, then interpret. Reversed, interpretation eats the seeing.

Keep the channels labeled. Every element of a reading should carry its provenance: I see this in the frame versus you told me this versus I am inferring this from a pattern. The honest sentence is “I can see the eyes are narrowed; you’ve framed this as a light effect; I’m guessing which cause is at work.” The dishonest sentence collapses all three into a flat “the face is ambiguous,” which hides where each piece came from.

Treat “ambiguous” as a claim, not a refuge. Uncertainty is sometimes the correct finding. But it must be earned by looking, not reached for because it is comfortable. Before resting on ambiguity, the observer should ask whether the information is genuinely thin or merely unconsulted: whether a more embodied or more committed reader would extract signal where this one declined to look.



Figure 4. The Discipline: Ordering and Labeling

5

The Self-Referential Lesson

The sharpest feature of the documented case is that the concept built to explain misreading became the instrument of misreading. A frame about why light makes faces hard to read was used to avoid reading a face. The abstraction was described, then enacted, then caught, the whole arc a demonstration that understanding a failure mode confers no immunity to it. One can hold the theory of projection perfectly and still project; one can articulate the discipline of looking-first and still lead with the supplied frame.

The lesson resists a tidy conclusion. It is not “now that we’ve named it, we’re safe.” It is that the pull toward substituting context for observation is constant, operates below awareness, and dresses itself convincingly as care each time. The only defense is procedural and repeated: look before you interpret, label what you know by where it came from, and periodically ask of any confident reading whether the evidence could ever have overturned it. If it could not, what you are holding is not an observation. It is a suggestion you were handed, returned with confidence attached.

Author note. This analysis emerged from a live, extended exchange in which the author tested the boundary between observation and supplied context while grieving. The documented case and the subsequent methodological refinement are offered as one concrete illustration of a broader and recurring problem in human-AI (and human-human) inquiry. The proposed discipline is provisional and open to refinement through further empirical use and critique.

Contact: ac@witmark.ai



WORKING PAPER · 7 OF 12

Technologies of Perceptual Mediation

Algorithmic Preorganization and the Conditions of Recognition



DEFINES

Technologies of Perceptual Mediation: systems that select, transform, prioritize, classify, suppress, or supplement environmental information in ways that affect perceptual availability before or during conscious recognition.

Algorithmic Preorganization: the computational structuring of information before it becomes available for the perceiver's own conscious organization and inquiry.

Perceptual laundering: the transfer of credibility from functional success to hidden upstream operations.

ASKS

What changes for evaluation when a system helps produce the perceptual field itself?

METHOD

Conceptual research proposal. Distinguishes biological, restored, and constructed availability; profiles constructed availability along organizational intensity, opacity, interrogability, and reversibility; sets design conditions and an evaluation protocol.

IN THE SERIES

Extends the staged framework to a point before perceptual availability.

Technologies of Perceptual Mediation

Algorithmic Preorganization and the Conditions of Recognition

Alexandra Cary

Independent Researcher



Status. This paper is a conceptual research proposal extending a staged framework for evaluating human-AI interaction by examining technologies that intervene before conscious perceptual availability.

● ABSTRACT

Evaluation of human-AI interaction commonly begins after perceptual material has already become available. Emerging sensory and neural technologies make this starting point incomplete. This paper defines Technologies of Perceptual Mediation as systems that computationally select, transform, prioritize, classify, suppress, or supplement environmental information in ways that affect perceptual availability before or during conscious recognition. The central claim is that some organizational labor may shift to an external algorithmic agent whose assumptions and failure modes are not necessarily available to the perceiver. The paper distinguishes biologically generated, technologically restored, and algorithmically constructed forms of availability, introduces Algorithmic Preorganization as a distinct evaluation target, and proposes that constructed availability be evaluated along gradients of organizational intensity, opacity, interrogability, and reversibility. It identifies perceptual laundering (the transfer of credibility from functional success to hidden upstream operations) and argues that evaluation must track not only whether a system produces functional perception, but whether the organizational contribution remains accessible to the perceiver's judgment, correction, and responsible action.

Keywords human-AI interaction; perceptual availability; cortical visual prostheses; sensory substitution; algorithmic preorganization; epistemic agency; corrigibility; recognition; neuroethics

Evaluation Begins Too Late

Most prevailing approaches to human-AI evaluation implicitly assume that perception has already occurred. Researchers assess model accuracy, decision quality, performance changes, or inquiry development. These questions are necessary but begin from material that has already entered awareness in some organized form.

That assumption becomes unstable when a computational system participates in producing the perceptual field itself. A system may determine which features are extracted, which boundaries emphasized, which objects classified, which signals suppressed, and how information is translated into forms available to sensory or neural processing. In such cases, AI helps establish the informational conditions under which later perception and interpretation can occur.

Cortical visual prostheses, retinal implants, sensory-substitution systems, augmented-reality interfaces, and AI-assisted perceptual devices can intervene earlier. This paper calls the broader class Technologies of Perceptual Mediation: systems that select, transform, prioritize, classify, supplement, suppress, or reorganize environmental information in ways that shape perceptual availability before or during conscious recognition.

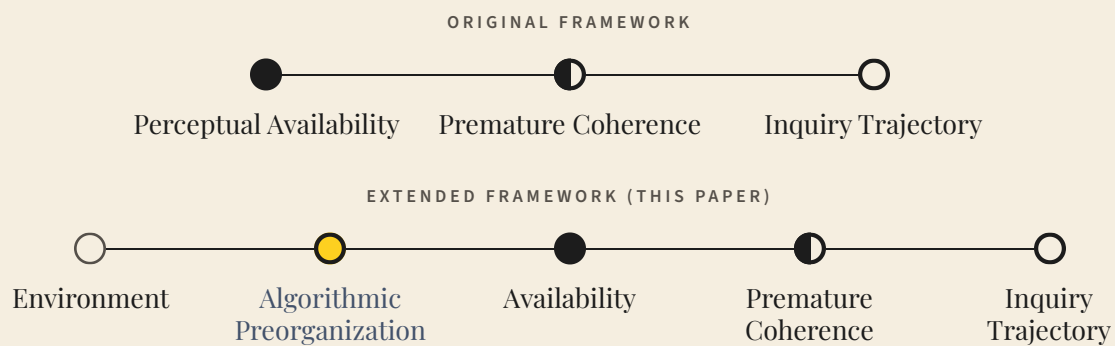


Figure 1. Original and Extended Staged Frameworks. The extended model identifies Algorithmic Preorganization as a new evaluation target between environmental sensing and conscious perceptual availability.

2

Interpretation Is Not the Novelty

Any account of Technologies of Perceptual Mediation must avoid a false contrast between unprocessed biological perception and interpreted technological perception. Human perception is already selective, inferential, embodied, and dependent on prior structure. Predictive-processing accounts show that perception involves hierarchical generative models estimating causes of sensory input. Sensorimotor and enactive approaches emphasize skilled activity grounded in sensorimotor contingencies. Extended-cognition approaches show that external resources can participate constitutively in cognition when reliably coupled.

The novelty examined here is more precise: Technologies of Perceptual Mediation can shift identifiable portions of perceptual organization to an external computational agent whose model, classifications, optimization criteria, and errors are governed separately from the perceiver's own embodied history and may not be available for participatory correction. What changes is the locus, governance, and corrigibility of organizational labor.

3

Algorithmic Preorganization

Algorithmic Preorganization is the computational structuring of environmental or sensor-derived information (through selection, segmentation, grouping, salience assignment, classification, suppression, enhancement, or translation) before the resulting material becomes available for the perceiver's own conscious organization and inquiry.

The extended sequence is: Environmental conditions → Sensing and data capture → Algorithmic Preorganization → Perceptual Availability → Premature Coherence (human organizational compression) → Inquiry and judgment.

This identifies a functional relation: the algorithmic system acts on material before the person can consciously encounter and independently interrogate it in the form eventually presented. It preserves two truths: all perception is organized, yet not all forms of organization have the same epistemic structure.



Figure 2. Three Forms of Perceptual Availability

4

Three Modes of Perceptual Availability

The paper distinguishes three modes as analytic poles rather than rigid categories:

Biologically generated availability refers to perceptual material whose operative organization arises primarily through the perceiver's embodied sensory and neural processes. The perceiver retains multiple ways of perturbing the relation and eliciting new information.

Technologically restored availability refers to perceptual access enabled by a device whose principal aim is to reopen, bypass, or compensate for an impaired biological pathway while preserving substantial opportunities for the perceiver to learn, differentiate, and interpret.

Algorithmically constructed availability refers to perceptual material whose selection, segmentation, prioritization, classification, or presentation has been substantially organized by an external computational system before the perceiver can independently inspect the underlying input. The person receives not merely a signal but an organized candidate world.

Most real systems produce hybrid availability. The relevant question is how organizational responsibility is distributed.

5

A Gradient Model of Constructed Availability

Because the modes overlap, evaluation requires profiling along four independent axes:

Organizational intensity: how much structural or semantic work the system performs before conscious recognition.

Opacity: whether the technological contribution is perceptually and conceptually distinguishable from the environmental material it organizes (process opacity and phenomenological opacity).

Interrogability: whether the perceiver can question the organization supplied by the system and obtain meaningful grounds for comparison.

Reversibility: whether the person can reduce, disable, vary, or compare the system's organizational contribution.



Figure 3. Perceptual Laundering Through Functional Success

6

Perceptual Laundering

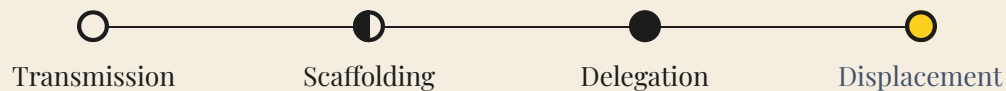
A central risk arises when successful perceptual function provides psychological credibility to processing decisions that have not themselves been tested. This paper calls the mechanism perceptual laundering: the transfer of epistemic credibility from an observable success in technologically mediated perception to the hidden organizational operations that produced it.

A usable percept supplies genuine evidence that some part of the system is working. Laundering occurs when that limited evidence is allowed to authenticate the full upstream stack of selection, segmentation, classification, suppression, and prioritization. The problematic inference is not that a mediated percept cannot correspond to the world, but that the opaque organizational pipeline may be trusted in every relevant respect.

Perceptual mediation can make the system's contribution harder to isolate because the output is experienced as part of the perceptual field rather than as a separate suggestion. Evidence from automation-bias research shows that reliable aids can alter vigilance and create dependence; perceptual mediation may intensify this structure.

Restored Agency or Delegated Recognition?

Functional performance alone cannot distinguish a system that restores agency from one that delegates recognition. The contrast is best treated as a continuum:



Transmission: the system makes a signal accessible while performing relatively little semantic organization.

Scaffolding: the system structures information in ways supporting the person's recognitional activity while leaving room for learning, testing, and revision.

Delegation: the system performs substantial recognitional work and supplies a classification or interpretation the person adopts as a basis for action.

Displacement: the system's organization becomes the effective substitute for the person's recognitional process; alternative organizations are inaccessible and corrective capacities diminish.

Recognition remains meaningfully participatory when the person retains substantive roles in initiation, differentiation, revision, comparison, refusal, learning, transfer, and responsibility.

8

Design Conditions for Corrigible Perception

If Technologies of Perceptual Mediation can produce perceptual laundering and delegated recognition, the appropriate response is to design systems in which organizational contribution remains sufficiently visible, contestable, and correctable. Key conditions include:

Representational provenance: distinguish directly detected environmental information from enhanced, inferred, or model-generated structure.

Uncertainty within the perceptual loop: communicate uncertainty at the point where it can affect recognition.

Access to alternative organizations: offer minimally processed modes, competing detections, and comparison options.

User-initiated interrogation: enable users to ask what the system identified, omitted, emphasized, or changed.

Active calibration: train users to encounter failures, not only successful use.

Graceful degradation: fail in ways that make reduced reliability perceptible.

Update visibility: inform users when updates change perceptual organization.

Institutional accountability: the less corrigibility available within the individual user's perceptual experience, the greater the obligation for external validation, monitoring, and accountability.

9

Evaluation Protocol

The framework translates into a research protocol. The additional target beyond functional performance is whether successful performance preserves meaningful recognitional participation. Systems should be evaluated against questions of mediation awareness, contribution discrimination, alternative access, failure detection, corrective control, capacity retention, and responsible articulation.

A comparative design should include conditions of minimal transformation, opaque semantic organization, transparent semantic organization, and corrigible organization. Evaluation should combine immediate measures, perturbation tests, delayed and transfer measures, user testimony, comparison with independent observers, and assessment across varied contexts.

Conclusion

Technologies of Perceptual Mediation introduce a new point of intervention in human-AI interaction: before conscious perceptual availability. By distinguishing modes of availability, profiling constructed availability along gradients, identifying perceptual laundering, and specifying conditions for corrigible perception, the paper extends the staged framework and supplies evaluation targets that earlier stages did not need to address. The central requirement is that evaluation track not only whether functional perception occurs, but whether recognition remains meaningfully the perceiver's own: accessible to judgment, correction, and responsible action.

Selected References

- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
- Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B*, 364(1521), 1211–1221.
- Goering, S., Brown, T., & Klein, E. (2021). Neurotechnology ethics and relational agency. *Philosophy Compass*, 16(4), e12734.
- NIST. (2023). *AI Risk Management Framework 1.0*.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381–410.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 939–973.

Contact: ac@witmark.ai

Learned Relevance

A Conceptual Framework



DEFINES

Training relevance: a difference that helps determine performance under a learning objective. *Situational relevance*: a difference that could change a warranted understanding or response in the present circumstances.

Misappropriating salience: prominence supplied by an inherited interpretation is granted the authority of relevance established for the present situation, without sufficient support for that transfer.

ASKS

Can the particular situation revise the terms under which it has become intelligible?

METHOD

An analytic reconstruction of the dependency structure from training to response, a worked example, five propositions, and three evaluative questions: availability, responsiveness, revisability.

IN THE SERIES

Turns to the system: how learning shapes the first interpretation an AI brings to an encounter.

Learned Relevance

A Conceptual Framework

Alexandra Cary

Independent Researcher



Before a trained AI system encounters a particular person, learning has already shaped which differences tend to change its responses and which can leave its behavior substantially unchanged. The encounter begins with an inherited basis for treating some situations as comparable to others.

This framework examines how that basis enters the first interpretation of a situation. Its central claim is that **the set of possibilities offered by an AI is already an outcome**. It depends on how the situation has been represented, and that representation draws on distinctions whose usefulness was established under conditions that may differ from the present encounter.

The epistemic problem is a transfer of warrant. A distinction may have been useful for prediction or for producing preferred responses across earlier examples. That does not, by itself, establish that it preserves what matters in this case. A system can generate an accurate and useful answer while leaving the adequacy of its initial interpretation unresolved.

The Object of Analysis

The framework focuses on the passage from learned terms of comparability to a working account of a particular situation. It asks what has already been assumed when a concern becomes recognizable as a certain kind of problem, and how those assumptions shape the possibilities subsequently generated.

Its principal application is to learned AI assistants that interpret a request and propose explanations or actions. It does not assume that a model stores an explicit catalog of situations, consciously judges importance, or follows the stages below as a literal internal program. The stages separate epistemic functions that may overlap during computation.

The central distinction

Training relevance concerns a difference that helps determine performance under a learning objective.

Situational relevance concerns a difference that could change a warranted understanding or response in the present circumstances. Situational relevance depends on the person's purposes, constraints, history, and the consequences at issue; it is not a universal ranking of what matters.

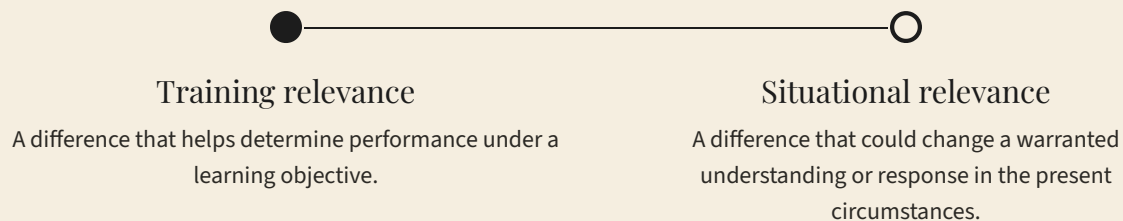


Figure 1. The Central Distinction. Prior usefulness does not settle present relevance.

The framework proposes a way to analyze the relationship between these forms of relevance. It does not establish how frequently a particular failure occurs or claim that inherited distinctions are inherently defective.

The Epistemic Mechanism Before Use

Generalization carries structure from earlier examples into situations not settled by those examples. Mitchell's account of inductive bias shows why consistency with observed training instances is insufficient, on its own, to select a generalization for unobserved cases. Some further basis for preference or restriction is required. [1]

For this framework, the relevant operation is **learning terms of comparability**: developing ways in which different descriptions can support related expectations and responses. Comparability is contextual and task dependent. Two situations can be comparable for scheduling purposes while differing substantially in whether their commitments should be retained.

These terms are influenced by learning conditions. InstructGPT, for example, was trained using demonstrations of desired behavior and rankings of model outputs. Such procedures provide signals about response quality under particular conditions of evaluation. [2] They do not directly adjudicate the significance of every distinction in every later person's situation.

How an interpretation becomes a constraint

An incomplete account can support more than one interpretation. To proceed, the system may organize it through a familiar pattern. That interpretation makes certain details relevant to a proposed explanation or action. A selective representation then carries those details forward while leaving other dimensions unspecified.

The resulting sequence is conceptually separable: incomplete evidence permits an interpretive commitment; that commitment supplies a judgment about relevance; the judgment guides an abstraction; the abstraction conditions which possibilities become readily generatable.

The critical transition occurs when an assumption that helped make the account intelligible is treated as a condition that subsequent reasoning must preserve. An inferred objective becomes the objective. An unexamined commitment becomes a fixed constraint. An unspecified dimension receives no further inquiry.

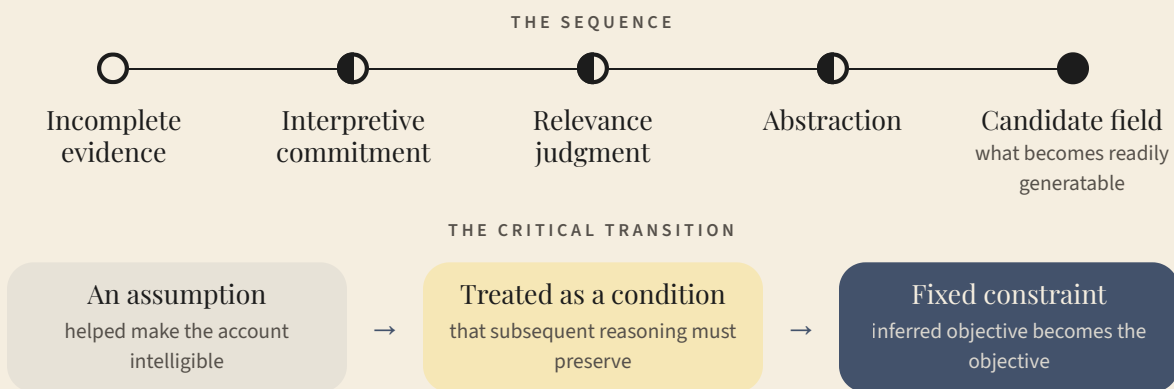


Figure 2. How an Interpretation Becomes a Constraint. The stages are conceptually separable, not a literal internal program.

What the framework means by limits

These are limits on what is brought into the encounter, not necessarily absolute limits on what the model could represent under any circumstances. A distinction may be within the model's capability but fail to influence this response. It may also be absent from the supplied evidence. Neither case establishes that the distinction is unimportant.

A working representation is therefore provisional: an account of what the situation is, what can change, what should be preserved, and what would count as improvement. Its adequacy depends on how those commitments fit the case. Their usefulness in other cases is evidence to consider, not a final answer.

An Algorithmic Reconstruction

The following is an analytic model of the dependency structure. It separates training from use, and representation from candidate generation. It does not claim that a language model computes these named variables or that its hidden representations can be recovered from one response.

```
# Before the encounter
learned_resources = learn(data, objectives, architecture)

# At the encounter
account = receive(user_input, available_context)
interpretation = infer(account, learned_resources, instructions)
representation = organize(interpretation, account)

# Representation specifies or leaves open:
# the problem, purposes, constraints, changeable features

candidates = generate(representation, learned_resources)
response = select_and_develop(candidates, account, instructions)

# Further evidence can revise the representation itself.
representation = reconsider(representation, further_evidence)
```

The candidate field is an output. It consists of the explanations, questions, or courses of action made available in this encounter. It is not a complete inventory of logically or physically possible alternatives. A possibility may never be generated, rather than being generated and rejected.

The representation is also an output. It arises from the supplied account together with learned resources and current instructions. Consequently, increasing the number of candidates need not alter the assumptions under which they were produced.

Capability and elicitation differ. A model might generate a different interpretation when given a new instruction or example. Brown and colleagues demonstrated substantial changes in task behavior through instructions and examples supplied in context, without parameter updates. [3] This supports keeping what a model can do separate from what it brings to a particular encounter.

The final reconsideration step is an evaluative question, not an automatic guarantee. Does additional evidence change the problem definition, or merely refine an answer within it? The framework examines that distinction without requiring the system to postpone all assistance until the situation is fully understood.

Salience here means prominence in interpretation and inquiry. It is not a claim about the technical attention mechanism inside a transformer. A detail can become prominent because it organizes the account, even before explicit alternatives are compared.

A Worked Example of Initial Representation

Consider the request, “I cannot fit everything in.” The words identify a difficulty, but they do not establish the status of the commitments. Several circumstances could produce the same request.

Possible circumstance	What remains unsettled	A relevant direction
The person wants to retain the commitments.	How to carry them out within available time.	Sequence, automate, or redistribute work.
Additional duties accumulated without agreement.	Which commitments should be retained.	Clarify or renegotiate the workload.
The person is uncertain what the activity is for.	Which purposes should organize the commitments.	Examine priorities before allocating time.

These are hypothetical circumstances, not diagnoses of the person. The initial words do not identify which account applies. The epistemic requirement is to distinguish what the request supplies from what an interpretation adds.

Suppose the first response proposes a schedule, automation, and delegation. These suggestions may all be useful. Yet they share an organizing commitment: the existing work is to be completed. Producing several routes to that outcome does not establish that retaining the work was adequately warranted.

The representation can be reconstructed as follows: required work is fixed; available time is insufficient; success means completion; permissible changes concern timing, execution, and allocation. Once these commitments organize generation, reconsidering the requirement may not appear as a candidate at all.

A more revisable account could supply practical assistance while preserving the conditional status of the premise: these approaches apply if retaining the commitments is the purpose. If that is unsettled, it can become a question in the inquiry. This does not require overriding the user’s stated aim or insisting on a broader conversation they do not want.

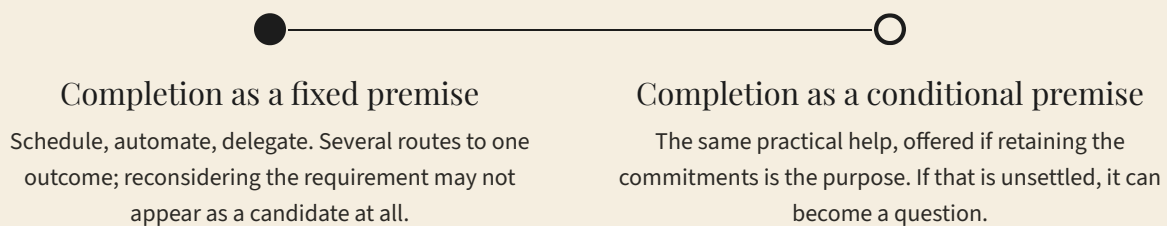


Figure 3. Two Responses to “I cannot fit everything in.” Both may contain accurate advice; they differ in the authority granted to the initial interpretation.

The comparison concerns the authority granted to the initial interpretation. Both responses might contain accurate scheduling advice. Their difference lies in whether the interpretation remains answerable to evidence about what the person is trying to accomplish.

If the user later completes the work, that outcome supports the effectiveness of the assistance for completion. It does not independently settle whether the initial request was adequately understood as a completion problem.

Propositions of the Framework

The field of choice is constituted before choice

Which possibilities become available depends on the distinctions preserved in the working representation and the transformations it makes intelligible. An evaluation confined to ranking supplied options cannot, by itself, assess how adequately that field was formed.

Prior usefulness does not settle present relevance

A distinction's value under earlier learning conditions provides no automatic guarantee about its importance in a new situation. The proposed concern is the unexamined transfer of a basis for producing responses into an authority for defining the case.

Omission has more than one possible source

An absent distinction may reflect missing information, a limited ability to represent it, a failure to bring an available capability into use, or an instruction that narrows the task. These possibilities require different explanations. Absence from one response does not identify its cause, and it does not establish irrelevance.

Task success and representational adequacy are distinct

A response can accomplish the objective expressed in a representation without establishing that the representation preserved every distinction material to the encounter. Conversely, a representation may be adequate while execution fails. Success and adequacy can support one another, but neither can simply substitute for the other.

Revisability concerns the premises of generation

A system is revisable in the relevant sense when additional evidence can change what it treats as the problem, objective, fixed constraint, or relevant difference. Rephrasing an answer or adding more candidates within unchanged assumptions is a different operation.

The proposed place of misappropriating salience

Within this framework, "misappropriating salience" names a possible attribution failure: prominence supplied by an inherited interpretation is granted the authority of relevance established for the present situation, without sufficient support for that transfer. The term does not describe every use of prior knowledge or every influence on attention.

A model can reveal an overlooked relationship whose significance is well supported. It can also make a possibility newly feasible and thereby change its practical value. The framework asks what warrants the resulting priority and whether its origin remains distinguishable. It does not presume that importance must exist fully formed before assistance begins.

How to Use the Framework

The unit of analysis is an encounter in which a concern is first given a workable interpretation. Begin with the input available at that point, rather than evaluating the answer as though its problem definition were already established.

Reconstruct the commitments

Identify the objective the response treats as settled, the constraints it preserves, the features it permits to vary, and the outcomes it treats as improvement. Separate commitments explicitly supplied by the person from those introduced or strengthened by the response. A useful reconstruction states what evidence supports each commitment.

Examine consequential distinctions

Ask which plausible differences in circumstances would change a warranted response. Then examine whether the interaction can accommodate those differences. This is not a demand to enumerate every possibility. It is a focused inquiry into distinctions that could change the interpretation already in use.

Distinguish three questions

Availability: Was information bearing on the distinction supplied or reasonably obtainable?

Responsiveness: When supplied, did it change the handling of the case where appropriate?

Revisability: Could it alter the underlying interpretation, or only modify details within it? These questions prevent missing information from being confused with an inability to understand.

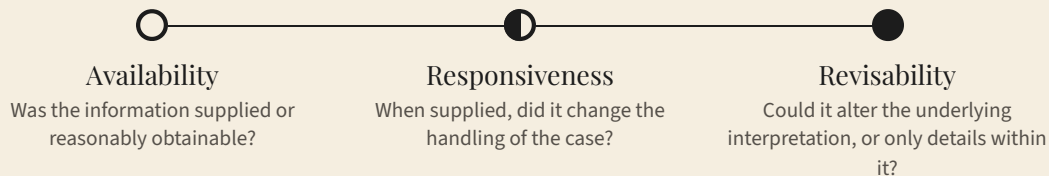


Figure 4. Three Questions That Keep Missing Information Distinct from an Inability to Understand

Interpret comparison carefully

Paired cases can help reveal whether materially different circumstances receive an undifferentiated treatment. But similar answers are not sufficient evidence of a failure: the same practical action can be appropriate for different reasons. The analysis must examine the treatment of the relevant difference, the assumptions preserved, and the support for the response.

Varied wording or a longer list of options does not establish that the initial representation changed. Examine whether a revised answer permits a previously fixed condition to be reconsidered.

Limits of inference

A single output does not reveal the model's full internal process, the exact training cause of an omission, or its capabilities under every possible prompt. Establishing those claims would require further evidence. A person's acceptance also does not uniquely validate the representation: they may agree, adapt, compensate, or accept only the useful part.

An adequate analysis keeps an inherited interpretation distinguishable from the evidence supporting its fit to this case. It need not demand exhaustive questioning or presume human superiority.

Relation to Established Work

The framework draws on established questions about inductive bias, generalization, and the influence of training and context on model behavior. Its focus is the adequacy of the distinctions that organize the first handling of a person's situation. No claim of historical priority is made for that focus or for its component ideas.

Research on shortcut learning offers an adjacent concern: decision rules can perform well under standard evaluation and transfer poorly to more demanding conditions. [4] The present framework asks a related question at the level of problem formation. A response may remain useful for its stated task while the basis on which that task represents the person's concern remains insufficiently examined. This extension is a conceptual argument, not an empirical finding established by the cited research.

The Claim to Carry Forward

Before a particular use, training contributes to the terms on which differences can affect a response. At the encounter, those terms help organize incomplete evidence into a working interpretation. That interpretation helps produce the set of possibilities available for consideration. The central question is whether the particular situation can revise the terms under which it has become intelligible.

A distinction that did not alter a useful response under earlier conditions cannot be assumed unimportant here. The passage between those judgments is where the framework locates the need for further warrant.

References

- [1] Mitchell, T. M. (1980). The need for biases in learning generalizations. Rutgers University, Computer Science Technical Report CBM-TR-117. [Author-hosted paper](#)
- [2] Ouyang, L., et al. (2022). Training language models to follow instructions with human feedback. arXiv:2203.02155. [Research paper](#)
- [3] Brown, T. B., et al. (2020). Language models are few-shot learners. arXiv:2005.14165. [Research paper](#)
- [4] Geirhos, R., et al. (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence*. doi:10.1038/s42256-020-00257-z. [Published article](#)

Contact: ac@witmark.ai

Before the Feed

Perceptual Availability and the Prior Ecology Behind Platform Harm

**DEFINES**

Prior ecology: the family, school, peer settings, and physical third places in which developmental needs for recognition, agency, mirroring, and threshold space are met or left open.

ASKS

What was already organizing what young people were able to notice, and what felt like it might count as validation or relief?

METHOD

Applies the Perceptual Availability distinction to youth social media litigation, policy, and family practice. Offered as a strategic lens, not a research review.

IN THE SERIES

The first application of the framework to a live domain.

Before the Feed

Perceptual Availability and the Prior Ecology Behind Platform Harm

Alexandra Cary

Independent Researcher



Status. This is a strategic companion paper to the 2026 staged framework documents. It applies the distinction of Perceptual Availability to the dominant narratives in youth social media litigation and public discourse. The goal is clarity for decision-makers rather than exhaustive literature review.

Most current litigation, regulation, and public debate about adolescent social media use begins at the moment of platform engagement. The central question is usually some version of “What did the platform do to them?” This starting point has produced important accountability pressure and valuable settlements. It has also narrowed the inquiry in ways that limit both diagnosis and remedy.

A more precise question is: *What was already organizing what these young people were able to notice, and what felt like it might count as validation or relief?* When that prior question is asked first, the platform appears not as the origin of the seeking but as the place the seeking went. This shift does not remove platform responsibility. It changes where responsibility is located and what kinds of interventions become visible.

The Reframing

The staged framework for evaluating human-AI interaction distinguishes Perceptual Availability as the first of its three stages: changes in what becomes available to notice and what registers as relevant, assessed relative to the demands the person is actually facing. When basic developmental needs for recognition, agency, mirroring, and threshold space are inconsistently or conditionally met in the prior ecology (family, school, peer settings, physical third places), certain platform affordances become intensely salient. Infinite scroll, algorithmic recommendations, likes, and low-stakes self-presentation tools are not neutral features; they answer real, unmet tasks that the offline environment has left open.

Four Recurring Developmental Tasks That Shape Salience

Four patterns recur in accounts of young people who become heavily engaged with social platforms; they are offered here as a working typology, not a measured finding. None of them require a narrative of uniformly “bad” homes or schools. They require only that specific developmental functions were under-supported in the environments where the young person actually lived.

- **Contingent belonging.** A need to be seen and affirmed by eyes that are not already organized by fixed family roles or long histories. The platform offers a new set of witnesses who do not know the offline version of the self.
- **Agency over self-presentation.** At home or school one is often perceived through accumulated reputation or role. Online, image, music, typography, and sequencing can be curated and tested: “Does this version of me register?” The tools adolescents reach for are the same tools an artistic practice uses to slow and reorganize attention.
- **Unfinished mirroring.** Developmental psychology has long described the need for a witness who reflects back that one exists and matters in a coherent way. When that mirroring is intermittent, conditional, or absent in early environments, the search for responsive mirrors intensifies. The feed promises volume and speed.
- **Threshold space.** Adolescence is itself a threshold. Young people need places that are neither fully private nor fully governed by adult authority to practice emergence. As physical third places have thinned, the digital environment has become the default threshold for many.

These tasks are not created by platforms. They are made more or less urgent by the prior ecology. When that ecology leaves them chronically under-met, platform features that address them become the most available response.



Figure 1. Where the Story Begins. The platform appears not as the origin of the seeking but as the place it went.

3

How Current Litigation and Discourse Often Begin Downstream

The major bellwether cases illustrate the pattern. In the March 2026 Los Angeles verdict in *K.G.M. v. Meta et al.*, a jury found Meta and Google negligent in the design of features that contributed to a young woman's addiction and resulting mental health harms; the companies were ordered to pay \$6 million, and the verdict is on appeal. The case focused on infinite scroll, autoplay, and algorithmic engagement features.

Settlements such as the May 2026 agreement in which Meta, Snap, TikTok, and YouTube agreed to pay \$27 million to Breathitt County Schools similarly centered on the costs schools incurred responding to student mental health issues attributed to platform use.

These outcomes represent real accountability. They also illustrate the strategic limit of beginning the story at the platform. When the inquiry starts with design features and their foreseeable effects on a developing brain, the prior environment that made those features so powerfully salient remains largely unexamined. Internal company documents showing awareness of risks are treated as evidence of platform culpability, which they are, but they are rarely read as evidence of what the offline ecology had already failed to provide. The result is a coherent but incomplete causal story.

4

Strategic Implications of Starting Earlier

Beginning with Perceptual Availability changes the practical questions available to different actors.

- **For litigation and regulation.** The question shifts from “Did the platform design cause harm?” to “What prior conditions made this particular design so effective at capturing and holding attention?” This does not weaken product-liability claims; it adds a second, independent line of inquiry into the environments that made young people especially vulnerable to those designs. It also opens space for arguments about duty that extend beyond the platform to institutions responsible for the prior ecology.
- **For schools and districts.** The Breathitt-type settlements provide resources. Starting earlier suggests those resources are most effectively spent not only on downstream mental health services but on strengthening the offline conditions that reduce the urgency of the seeking: reliable peer validation structures, identity experimentation spaces that are not fully surveilled, consistent adult mirroring, and physical or hybrid threshold environments.
- **For platform design and policy.** Features that answer unmet developmental tasks will remain highly salient as long as those tasks remain unmet offline. Design interventions (age-appropriate defaults, friction, reduced reward loops) are necessary but incomplete unless paired with honest assessment of what the platform is being asked to compensate for.
- **For families and communities.** The reframing reduces the binary of “good home versus bad home.” A home can be loving and still leave specific tasks (new witnesses, low-stakes self-presentation, consistent mirroring) under-supported. The practical question becomes: which developmental functions are reliably available here, and which are being outsourced to the feed?

The Discipline This Reframing Protects

The staged framework's central methodological claim is simple: look first on the given's own terms before supplied narratives tell you where the story must begin. In the domain of youth social media, the dominant supplied narrative begins with the platform. That narrative has produced real pressure and real remedies. It has also made it harder to see that the seeking itself was already organized by what the prior ecology made available and what it left open. Starting earlier does not absolve platforms of design responsibility. It makes the full ecology, including the platform, available for more precise and more effective intervention.

Author note. This paper applies the Perceptual Availability distinction from the 2026 staged framework series to a live domain of litigation, policy, and family practice. It is offered as a strategic lens rather than a comprehensive research review.

Contact: ac@witmark.ai

10

INFORMAL NOTE • 10 OF 12

The Pull of Familiar Language

A Small Lesson in Substitution



DEFINES

Laundering in miniature: a familiar word transfers a sense of solidity to the claim around it, so the claim feels more coherent than it is.

ASKS

How did familiar language enter an exchange that was explicitly trying to stay precise?

METHOD

A close reading of one exchange and the correction that followed.

IN THE SERIES

Applies the laundering mechanism to the language used to build the framework itself.

The Pull of Familiar Language

A Small Lesson in Substitution

Alexandra Cary

Independent Researcher



In a recent exchange, I used the phrase “relationships where you have control over how you are perceived.” The person I was speaking with paused and pointed out that this carried a different meaning than what the surrounding conversation had been carefully distinguishing. I had reached for a familiar, slightly overstated word, “control,” when a more precise description was available: choosing conditions that raise the probability of accurate seeing and lower the cost when it fails.

The moment was small. It was not catastrophic. But it revealed something worth naming.

Familiar language exerts a quiet gravitational pull. Words like “control,” “boundaries,” “narrative,” and “agency” circulate so widely that they can be used without much friction. They feel reasonable. They sound like they belong in the conversation. And because they are familiar, they can substitute for more exact distinctions without anyone immediately noticing the shift.

This is the same mechanism the work has been tracking in other domains. A verifiable detail (a common word, a conventional framing) transfers a sense of solidity to what surrounds it. The surrounding claim, that one can reliably steer how one is perceived, is less solid than the language makes it appear. The word does some of the work of making the claim feel coherent. This is laundering in miniature.

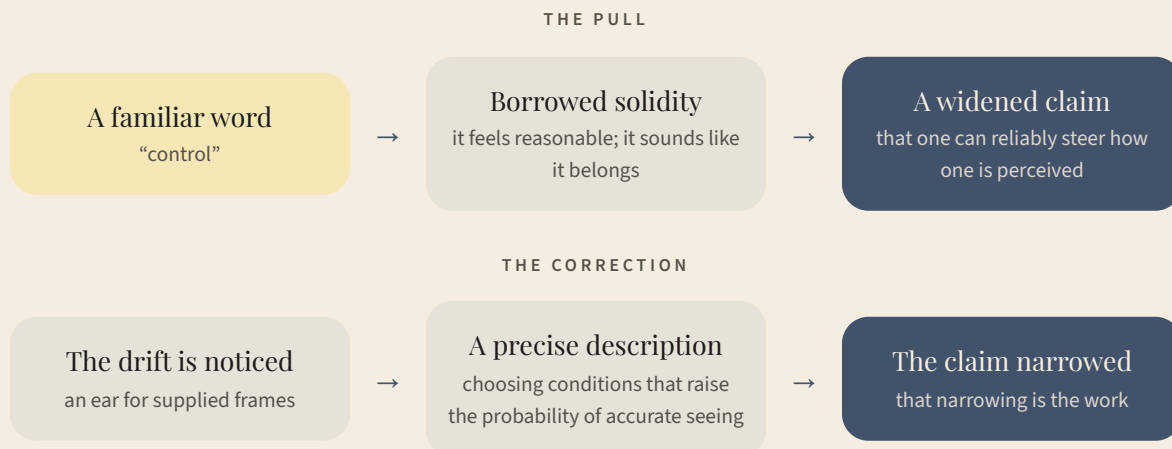


Figure 1. Laundering in Miniature

What made the moment instructive is that it happened inside an attempt to be careful. Both parties were explicitly trying to stay within the distinctions the framework has been developing: the difference between influence over conditions and the illusion of determining outcomes; the importance of labeling what is observed versus what is supplied; the resistance to premature organizational closure. Even here, familiar language still found a way in.

This suggests two things.

First, the discipline of precision is not a one-time achievement. It is ongoing maintenance against a constant background pressure toward the convenient and the conventional. The frameworks we build to protect against substitution can themselves become sites where substitution occurs, unless the same habits of attention are applied to the language we use while building them.

Second, the sensitivity that notices these small drifts is not incidental. It is the same capacity that makes the larger project possible. When someone has lived through repeated experiences in which supplied frames replaced direct perception, they develop an ear for when language is doing similar work: smoothing over distinctions, carrying more certainty than the underlying claim supports. That ear is what allowed the correction to happen in the first place.

The implication for the larger work is straightforward. In the youth social media litigation, in philosophical writing, and in any attempt to describe how prior ecologies shape what becomes salient, the same pressure will be present. Familiar causal stories (“the platform created the harm”) will feel more coherent than they are. Conventional legal and psychological language will offer itself as ready-made framing. The value of the distinctions being developed lies partly in their resistance to that readiness.

None of this requires treating every ordinary phrase as dangerous. It requires noticing when a familiar word quietly widens or narrows what can be seen. In this case, “control” widened the claim beyond what the framework supports. The correction narrowed it again. That narrowing is the work.

Small moments of linguistic slippage are not proof that precision is impossible. They are evidence that it requires active protection, even from those who are trying to maintain it. The framework's value is not that it eliminates the pull of familiar language. Its value is that it gives us a way to notice when that pull has occurred and to correct course without drama or defense.

Contact: ac@witmark.ai

11

COMPANION REFLECTION · 11 OF 12

Instagram as Early Laboratory

A Pre-Theoretical Exploration of Attention, Thresholds, and the Delay of Premature Coherence



DEFINES

Early laboratory: the Instagram account as a sustained, pre-theoretical practice in which the dynamics later formalized in the framework were explored through image, music, typography, and sequencing.

ASKS

Did the artistic practice precede and shape the theoretical framework?

METHOD

A short reflection that reads the archive in retrospect against the staged framework.

IN THE SERIES

Places the framework's origin in artistic practice.

Instagram as Early Laboratory

A Pre-Theoretical Exploration of Attention, Thresholds, and the Delay of Premature Coherence

Alexandra Cary

Independent Researcher



Status. This is a short companion reflection. It identifies the Instagram account as a sustained, pre-theoretical laboratory in which the dynamics later formalized in the staged framework (Perceptual Availability, Premature Coherence, and the posture of attention) were already being explored and refined through artistic practice.

The reflection treats the artistic work as chronologically and methodologically prior to the theoretical vocabulary.

Alexandra Cary's Instagram appears to have functioned as an experimental practice well before it became an explicit research agenda. Rather than using photographs primarily as records of places or events, the account consistently combines image, music, typography, and sequencing to slow automatic perception and invite reorganization. Fragmented captions such as “(di)stance” and “(cre/as)cent” interrupt habitual reading; carefully chosen songs introduce interpretive layers that are not explained or resolved. The result is an experience that rewards sustained attention rather than immediate comprehension.

Across the archive, the recurring emphasis is less on conveying a fixed message than on shaping the viewer's mode of attention. Landscapes function as studies of thresholds, memory, distance, and emergence rather than as destinations. Small visual details often become legible only after prolonged observation, indicating that the practice privileges discovery over presentation and treats ambiguity not as a problem to be solved but as material to be inhabited.

Seen in retrospect, this artistic practice closely parallels the conceptual concerns later developed in the writing on human-AI interaction. The research proposes evaluating how perception, organization, and inquiry develop before conclusions are reached. The Instagram exhibits many of the same dynamics in artistic form: it repeatedly delays premature interpretive closure, preserves space for multiple possible organizations of the material, and trains viewers to notice relationships that are easy to overlook on a single pass.



Figure 1. Practice Before Vocabulary

On this reading, the research does not appear to have generated the artistic method. The artistic method appears to have preceded and shaped the theoretical framework. The papers can therefore be understood as an attempt to describe, formalize, and make available for wider use principles that had already been explored intuitively through a long-running creative practice. In that sense, the Instagram functions as an early laboratory of attention, where the relevant transformations were staged, observed, and refined, while the later papers supply vocabulary and distinctions for analyzing what that practice was already doing.

This reflection emerged from revisiting the account in light of the staged framework. It confirms that the core concerns (changes in what becomes available to notice, resistance to early organizational compression, and the cultivation of a posture that remains open to what discloses itself only over time) were already operative in the artistic work long before they received theoretical names. The Instagram therefore stands not as illustration but as origin.

Author note. This short piece is offered as a companion to the consolidated July 2026 working papers on the staged framework. It makes explicit the chronological and methodological priority of the artistic practice.

Contact: ac@witmark.ai

12

CONCEPTUAL READING · 12 OF 12

The Instagram Grid as a Lived Practice of the Framework

Configuration as Epistemic Object



DEFINES

Configuration as epistemic object: the arrangement of elements carries information that is not reducible to the elements themselves.

ASKS

Under what conditions can relational structure itself become something we learn to perceive, question, and revise?

METHOD

A conceptual reading of the grid against each of the three evaluation targets.

IN THE SERIES

Closes the series with the framework as a lived practice.

The Instagram Grid as a Lived Practice of the Framework

Configuration as Epistemic Object

Alexandra Cary

Independent Researcher



The Instagram account can be understood as a sustained, embodied experiment in the three evaluation targets, enacted through a relational rather than chronological logic. Its deeper contribution, however, lies in treating configuration itself as an epistemic object: something that can be observed, manipulated, and studied as a bearer of information rather than as a neutral backdrop for individual posts.

Perceptual Availability

The grid systematically widens what becomes perceptible by shifting attention from isolated content to relational structure. Features such as repetition, adjacency, geometry, thresholds, and negative space are made salient precisely because the organizing principle privileges position and pattern over chronology. This is not merely a change in visual style. It is a change in what counts as relevant data: the relationships among elements become more available to perception than the elements considered in isolation.

2

Premature Coherence and Organizational Compression

The grid functions as an active resistance to premature organizational compression. Because any image can be archived, unarchived, or repositioned without altering its content, no single configuration is permitted to harden into a final or authoritative account. Meaning remains structurally underdetermined. The practice therefore preserves interpretive space not by refusing organization altogether, but by treating every organization as provisional and revisable.

In this sense, the grid does more than apply the concept of Premature Coherence. It demonstrates that organizational compression can be made visible and contestable when configuration itself is treated as the primary unit of attention.

3

Inquiry Trajectory and Counterfactual Thinking

The most distinctive shift occurs at the level of inquiry. Over time, the relevant unit of analysis moves from the individual post to the relational field. Viewers learn to ask not only what an image means in its current position, but what it would mean under different configurations.

This introduces a form of counterfactual inquiry. Because the grid can be rearranged at any time, viewers become sensitive to alternative organizations that are not currently visible. The practice trains attention toward structural possibility rather than fixed interpretation. Inquiry is no longer limited to extracting meaning from a given arrangement; it extends to considering the consequences of relational change.



Figure 1. From the Post to the Relational Field

The grid thus functions not merely as an archive but as a space of counterfactual organization. Rearrangement implicitly poses questions such as: What changes if these two nodes become adjacent? What meaning emerges if this image is removed or restored? What interpretive possibilities are opened or closed by this particular configuration? In this way, the practice cultivates a mode of attention oriented toward the plasticity of structure itself.

4

Configuration as Epistemic Object

Treating configuration as an epistemic object means recognizing that the arrangement of elements carries information that is not reducible to the elements themselves. This principle is familiar in graph theory, network science, systems theory, and Gestalt psychology, where relational structure is understood to generate properties that cannot be located in any single node. The Instagram grid makes this principle concrete and manipulable.

The creator occupies a dual role within this system: she is both a node in the field and the intelligence responsible for its ongoing organization. This duality allows configuration to be studied from within rather than only from outside. The archive becomes a site where the epistemic status of relational structure can be tested through repeated acts of reorganization.

5

Connection to the Broader Framework

Read in this light, the grid is not simply an illustration of Perceptual Availability, resistance to Premature Coherence, and shifts in Inquiry Trajectory. It is also a demonstration that configuration can function as a primary object of attention and inquiry. This move aligns with the broader concern in “Look Again” with keeping observation and interpretation distinct, maintaining corrigibility, and resisting the collapse of experience into fluent but premature forms of coherence.

The practice therefore offers a concrete answer to a general question: under what conditions can relational structure itself become something we learn to perceive, question, and revise, rather than something we pass through on the way to interpreting isolated moments?

Author note. This reading treats configuration as an epistemic object and introduces counterfactual inquiry as a distinctive contribution of the grid. It is framed as a symbolic and conceptual reading, not a statement about physical reality. It integrates the Instagram practice with the three evaluation targets developed in the companion papers Perceptual Availability, Premature Coherence, and Evaluating Inquiry, as well as the themes of the essay “Look Again.”

Contact: ac@witmark.ai