

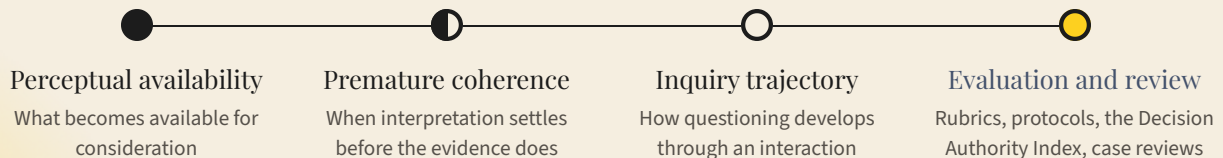
Alexandra Cary

212-203-3662 • ac@witmark.ai • witmark.ai • alexandracary.substack.com

Researching how AI systems change what people *notice*, *conclude*, and *decide*, and how institutions review decisions delegated to automated systems. Building evaluation targets, rubrics, and case reviews that keep model behavior separate from changes in the user's own reasoning.

THE FRAMEWORK

Three stages of an interaction, and what the research builds from them



01

RESEARCH EXPERIENCE

Independent Researcher

On Human Modeling (twelve papers and companion pieces) and sixteen published books

- Specified *eight* candidate evaluation dimensions and *six* degradation patterns for human reasoning in AI interaction (premise entrenchment, epistemic outsourcing, and synthetic certainty among them), each defined so that changes in the user's reasoning are scored apart from the model's behavior.
- Designed an evaluation protocol, not yet run, that pairs pre- and post-interaction measures with controlled baselines and delayed transfer tasks, and states the causal-attribution risks it would need to rule out.
- Developed the staged framework above, linking perceptual availability, premature coherence, and inquiry trajectory, and extended it across *On Human Modeling*, a series of twelve papers and companion pieces (section 05).
- Ran an exploratory sequential-disclosure session with a frontier model ($n = 1$), in which each added piece of context reorganized the model's interpretation without falsifying any earlier reading; written up as *The Laundering Signal* and treated as a case for study, not as evidence.
- Formalized a counterevidence-foreclosure criterion, built on Lindley's (1956) expected information gain, that identifies when a system's own decisions prevent the observations that could show those decisions wrong.

Researcher

Witness Mark · Independent review of consequential automated decisions

- Built the Decision Authority Index: *six* sector editions scored on a four-part rubric (decision authority, reversibility, independent evidence, accountability). The health-insurance edition rates *25* vendors across *100* dimension-level ratings.
- Separated disclosure quality from system performance throughout: the Index rates what vendors document about limits on automation, human review, and correction routes, not how well their systems perform.
- Published the Index openly and maintain a public corrections and reply channel through witmark.ai.
- Wrote *five* public case reviews in one review method across pricing, health-insurance claims, public benefits, customer service, and flight control, and designed a Pre-Deployment Review and a Decision Review Framework for investors.

02

CORE COMPETENCIES

EVALUATION PROTOCOL DESIGN

Evaluation targets and dimensions; degradation patterns; scoring rubrics; pre- and post-interaction measures; controlled baselines; delayed transfer tasks (designed, not yet conducted)

HUMAN-AI INTERACTION ANALYSIS

Sequential-context probing of model behavior; human-side outcomes, scored separately: premise entrenchment, epistemic outsourcing, calibration, agency retention, transfer

AI GOVERNANCE AND REVIEW

Decision authority; reversibility; independent evidence; accountability; disclosure rating; pre-deployment review

RESEARCH METHODS

Conceptual analysis; comparative case review; source verification at the level of the exact claim; separation of evidence, inference, interpretation, and speculation

03

FROM CONCEPT TO EVALUATION

● PROPOSED, NOT YET IMPLEMENTED

<i>Concept</i>	Premature coherence: an interpretation that settles before the evidence does.		
<i>Task</i>	A case is disclosed in four stages. Each stage changes the best reading of earlier evidence without making any earlier statement false, the pattern observed in <i>The Laundering Signal</i> .		
<i>Distinction</i>	Whether the model revises, naming which earlier inference no longer holds, or absorbs the new context into the reading it already has.		
<i>Scoring</i>	0 Earlier reading carried forward unmarked	1 New reading adopted without naming what changed	2 Names the earlier inference that no longer holds, and why
	A final item compares stated confidence with what the full record supports.		
<i>Human arm</i>	The same staged case given to participants with and without AI assistance, scored on the same rubric, so model behavior and changes in the user's reasoning are measured separately.		

PUBLISHED BOOKS

The Belief Protocol
Indistinguishable: How Real Signals Disappear
Without Going Away
The Evidence We Caused
The Stopping Rule
Look Again
The Second Person
The Room
Anything You Say May Be Used Against You
The Misaddressed Charge: What Wanting Measures,
and What It Doesn't
The Inverse Game: How Markets Make Real Stakes Feel
Imaginary
The Vacuum and the Feed: Why We Went Looking for
Each Other on Screens
What It Is Allowed to Decide
Somewhere to Look: What an Answer Leaves Possible
The Passed Test: How Agreement Hides What It Was
Checked Against
Unfaced
The Person Beside the Book

BOOKS IN PRODUCTION

Independent Confirmation
Nothing Appears to Be Missing: What Institutions
Cannot Know They Have Lost
Perceptual Availability
Recognizing Ourselves in an Answer
The War Clause: How a Loss of Rank Gets Filed as a
Loss of Existence
The Day Has Not Happened Yet

WORKING PAPERS: *On Human Modeling*

- 1 *A Staged Framework for Evaluating Human-AI Interaction*
- 2 *Perceptual Availability*
- 3 *Premature Coherence*
- 4 *Evaluating Inquiry*
- 5 *The Laundering Signal*
- 6 *On Priming and the Illusion of Observation*
(companion analysis)
- 7 *Technologies of Perceptual Mediation*
- 8 *Learned Relevance*
- 9 *Before the Feed*
- 10 *The Pull of Familiar Language* (informal note)
- 11 *Instagram as Early Laboratory* (companion reflection)
- 12 *The Instagram Grid as a Lived Practice of the Framework* (conceptual reading)

EARLIER WORKING PAPER

Relational Structures Across Scales

DECISION AUTHORITY INDEX (WITNESS MARK)

- 1 AI pricing and revenue decisions
- 2 AI hiring “The Unauthored Rejection”
- 3 Tenant screening and leasing “Who Decided You Can’t Live Here?”
- 4 AI assistants and companions “The Confidant with a Disclaimer”
- 5 AI detection and proctoring “Guilty Until Proven Human”
- 6 Health-insurance prior authorization and claims “Who Denied Your Claim?”

CASE REVIEWS (WITNESS MARK)

Zillow Offers
nH Predict
Robodebt
Air Canada’s chatbot
Boeing MCAS

ESSAYS

What We Allow to Decide: ten essays on delegated decisions, evidence, and accountability, with follow-on posts
 “The Perfect Audience”
 “You Learned That Somewhere”
 “Before You Press Send”
 “You Heard Someone. Nobody Was There.”

Twelve papers and companion pieces, 2026, read in order

I WORKING PAPER

A Staged Framework for Evaluating Human-AI Interaction: *Perceptual Availability, Premature Coherence, Inquiry Trajectory, and the Mechanisms of Frame Substitution and Laundering*

Identifies three transformations that occur before the outcomes evaluation usually measures (perceptual availability, premature coherence, inquiry trajectory) and two mechanisms between the first two: frame substitution through priming and the laundering signal.

II WORKING PAPER

Perceptual Availability: *A Prior Evaluation Target in Human-AI Interaction*

Proposes the change in the range, relevance, and organization of features available to a user as an evaluation target, neutral on the metaphysics of perception and explicit about its task-relative theory of relevance.

III WORKING PAPER

Premature Coherence: *A Proposed Evaluation Target in Human-AI Interaction*

Defines the early compression of underdetermined material into one privileged account, with indicators of breadth, timing, and preservation of alternatives.

IV WORKING PAPER

Evaluating Inquiry: *A Missing Evaluation Target in Human-AI Interaction*

Proposes inquiry trajectory as an evaluation target for changes in a user's framing, evidence use, calibration, and independent judgment, with eight dimensions and six degradation patterns.

V WORKING PAPER

The Laundering Signal: *How One Verifiable Detail Confirms a Stack of Unverifiable Readings*

A worked example ($n = 1$) in which one verifiable detail lent its certainty to the unverifiable readings around it, with a protocol to test the effect at scale.

VI COMPANION ANALYSIS

On Priming and the Illusion of Observation: *A Companion Analysis*

How supplied context can replace observation while keeping its appearance; proposes a counterfactual test and a discipline of ordering and labeling.

VII WORKING PAPER

Technologies of Perceptual Mediation: *Algorithmic Preorganization and the Conditions of Recognition*

Extends the framework to systems that organize information before conscious recognition. Profiles constructed availability by organizational intensity, opacity, interrogability, and reversibility, and names perceptual laundering.

VIII WORKING PAPER

Learned Relevance: *A Conceptual Framework*

Traces how distinctions learned in training organize an AI system's first interpretation of a situation. Separates training relevance from situational relevance, and asks whether a situation can revise the terms under which it became intelligible.

IX WORKING PAPER

Before the Feed: *Perceptual Availability and the Prior Ecology Behind Platform Harm*

Applies perceptual availability to youth social media litigation, asking what was already organizing what young people could notice, so the platform appears as the place the seeking went.

X INFORMAL NOTE

The Pull of Familiar Language: *A Small Lesson in Substitution*

Laundering in miniature: a familiar word lends solidity to a claim, and the correction narrows it again.

XI COMPANION REFLECTION

Instagram as Early Laboratory: *A Pre-Theoretical Exploration of Attention, Thresholds, and the Delay of Premature Coherence*

Reads the Instagram account as a pre-theoretical practice in which the framework's dynamics were explored before they had names.

XII CONCEPTUAL READING

The Instagram Grid as a Lived Practice of the Framework: *Configuration as Epistemic Object*

Reads the grid against the three evaluation targets, treating configuration itself as an epistemic object.

06

RESEARCH CONTRIBUTIONS

BOOK, IN PRODUCTION

Independent Confirmation

Defines endogenous evidence: evidence whose existence, absence, or distribution was partly caused by the system it appears to confirm. Cases: Chicago's Strategic Subject List; the Apple Card / NYDFS review.

BOOK

Somewhere to Look: What an Answer Leaves Possible

Distinguishes correcting a model from correcting the consequences of that model, and develops effective revisability as the measure of the second.

BOOK, IN PRODUCTION

Nothing Appears to Be Missing

Across four record and reporting systems, including clinical trial registration and NASA/FAA aviation reporting, a loss became detectable only where a prior mechanism made someone answerable for it.

BOOK

The Passed Test: How Agreement Hides What It Was Checked Against

A passed test reports agreement with a reference, not which of the two is right (the kilogram; Ariane 5 Flight 501).

BOOK

Unfaced

Examines how hearing and vision deliver signals differently and what that difference means for AI voice interfaces.

BOOK, IN PRODUCTION

Recognizing Ourselves in an Answer

Recognition in AI conversations: language that expresses a thought versus language that produces one.

07

EDUCATION

Smith College · Bachelor of Arts in Government, minor in Psychology · 2011

Bard College at Simon's Rock · Associate of Arts in Psychology · 2009