

# Judging How to Think with AI as a Tool for Thought

Paul C. Parsons

parsonsp@purdue.edu

Purdue University

West Lafayette, Indiana, USA

## Abstract

Generative AI is widely described as a tool for thought, yet the same system can either deepen reasoning or short-circuit it, especially across levels of expertise. This variability cannot be understood through trust, explainability, or output quality alone, but depends on whether AI-mediated workflows preserve users' judgment over commitments under uncertainty. This paper conceptualizes judgment as a situated, enacted competence and presents a compact taxonomy of recurring judgments that govern how AI participates in the lifecycle of commitments in reasoning, suggesting that AI supports thinking only insofar as it sustains accountable, revisable judgment—even when final outputs appear equally correct.

## CCS Concepts

• **Human-centered computing** → Human computer interaction (HCI); HCI theory, concepts and models.

## Keywords

Judgment, Generative AI

### ACM Reference Format:

Paul C. Parsons. 2026. Judging How to Think with AI as a Tool for Thought. In *Proceedings of Tools for Thought Workshop at CHI '26, April 16, 2026, Barcelona, Spain (TfT '26)*. ACM, New York, NY, USA, 5 pages.

## 1 Introduction

Generative AI systems are increasingly described as tools for thought (TfTs): systems intended not only to improve efficiency but to support reasoning and learning. Yet their cognitive effects are uneven. The same model can deepen reasoning in some contexts while short-circuiting it in others, especially across levels of expertise. Experienced users may use AI to explore alternatives or interrogate assumptions, whereas novices may accept fluent output as authoritative, with consequences for learning and responsibility.

Several lines of work attempt to account for this variability. One response frames the problem in terms of trust or appropriate reliance: users must learn when to accept or override AI recommendations [6, 13]. Another treats prompting as an iterative, representational practice rather than a one-shot query, emphasizing reflective refinement over time [2, 10]. While these perspectives are important, they leave open a further question: how users judge the commitments that enter and structure reasoning as interaction unfolds.

Thinking with AI requires ongoing *judgment* about how the system participates in reasoning. Users judge whether to initiate AI use at all, how to bound its contributions, which assumptions to stabilize, what warrants scrutiny, how to integrate suggestions without surrendering authorship, where delegation should be limited, how earlier commitments should be revised, and which claims they are prepared to own. These judgments regulate how commitments enter and structure reasoning, yet they are rarely treated as a central object of analysis.

Here, judgment is not treated as an abstract virtue or latent trait, but as a situated capacity exercised in use. In this context, I define judgment as the capacity to manage commitments: determining what to accept, reject, verify, defer, revise, or endorse under uncertainty and constraint. This view resonates with design-theoretical accounts that position judgment as central to action in indeterminate situations (e.g., [8]), while shifting attention to how such judgment operates in AI-mediated reasoning. This framing foregrounds a distinction between fluent output and cognitive support. An AI system may readily generate explanations, yet still undermine thinking if it fails to preserve explicit commitments that constrain subsequent reasoning and can be meaningfully revised.

This perspective clarifies why superficially similar AI-mediated workflows produce different cognitive consequences. When judgment is actively exercised, AI functions as a conditional resource whose contributions are initiated, bounded, stabilized, scrutinized, integrated, limited, revised, and owned. When judgment is weak or underdeveloped—as is often the case for novices—AI can function as a default authority, collapsing intermediate reasoning into output acceptance. The difference lies in how commitments are governed in use and in whether workflows make such governance possible.

The contribution of this paper is conceptual and analytic. I articulate judgment in AI-supported thinking as a repertoire of recurring, observable acts—initiating, bounding, stabilizing, scrutinizing, integrating, limiting, revising, and owning—through which users regulate the lifecycle of commitments in reasoning. The term *judgment* is used deliberately to emphasize situated, experience-based action rather than formal decision-making in a rational or normative sense (cf. [3, 8, 11]). This taxonomy offers a practical lens for distinguishing AI-mediated workflows that sustain judgment from those that erode it.

By foregrounding judgment, this paper clarifies what it should mean for AI to function as a TfT. Rather than evaluating systems primarily in terms of what they generate, I focus on how commitments are managed in use and on whether reasoning remains accountable and revisable. This shift provides a basis for evaluating and shaping AI-mediated workflows in contexts where expertise is developing and cognitive engagement matters most.



This work is licensed under a Creative Commons Attribution 4.0 International License. TfT '26, Tools for Thought Workshop at CHI '26, April 16, 2026, Barcelona, Spain © 2026 Copyright held by the owner/author(s).

## 2 Background and Related Perspectives

This account of judging how to think with AI builds on several complementary traditions that have treated judgment as central to competent action under uncertainty: professional judgment, design judgment, and judgment in complex cognitive systems.

### 2.1 Judgment Beyond Rules and Procedures

Across professional domains, judgment has long been invoked to describe forms of decision-making that cannot be reduced to formal rules or complete specifications. Dunne [3] characterizes professional judgment as a practical competence exercised in situations that are indeterminate, value-laden, and resistant to algorithmic resolution. In such contexts, practitioners must act despite incomplete information, conflicting demands, and the absence of a single correct answer. Judgment is not a fallback when procedures fail, but the primary means by which responsible action becomes possible in the first place—emerging through situated activity rather than the execution of pre-specified plans (cf. Suchman [11]).

This emphasis on situated action under constraint is directly relevant to AI-mediated reasoning. Generative AI systems routinely present users with plausible outputs in the absence of ground truth, forcing users to decide what to accept, what to question, and what they are willing to stand behind. Framing AI use purely in terms of accuracy or compliance obscures the fundamentally judgment-laden nature of these decisions.

### 2.2 Design Judgment and Intentional Change

Within design theory, judgment has been articulated as a core competency for navigating situations that are ill-defined, contingent, and shaped by human intention. Drawing on pragmatist and phronetic traditions that treat action under indeterminacy as irreducible to formal procedure, Nelson and Stolterman [8] argue that design judgment governs how designers move from abstract possibilities to concrete outcomes—the “ultimate particular”—through commitments that are neither derivable from scientific truth nor guaranteed by method. Judgment, in this sense, is not the application of rules but the capacity to bring about intentional change under conditions of complexity and uncertainty.

Importantly, design judgment is not a single moment of choice but an ongoing activity: designers continually decide what to fix, what to leave open, what counts as sufficient, and when to revise earlier commitments [5, 9]. This perspective aligns closely with contemporary accounts of AI as a tool for thought. When AI is introduced into reasoning workflows, users are not merely solving problems but shaping trajectories of thought through a series of commitments and revisions.

### 2.3 Judgment in Complex Cognitive Systems

Research in cognitive systems engineering and naturalistic decision-making has similarly emphasized judgment as a response to complexity, time pressure, and uncertainty. Rather than viewing cognition as an individual process, this tradition treats decision-making as emerging from joint cognitive systems composed of humans, artifacts, and environments [12]. From this perspective, good performance depends on maintaining effective coordination and control across the system as a whole.

This body of work highlights how expertise manifests as the ability to recognize what matters, anticipate consequences, and manage trade-offs—often without explicit deliberation [7]. Crucially, it also shows how automation can either support or undermine judgment, depending on how responsibilities and commitments are distributed across the system [1, 4]. While this literature has primarily focused on safety-critical domains, its insights are increasingly relevant as generative AI becomes embedded in everyday reasoning tasks.

### 2.4 Toward Judgment in Use

Taken together, these traditions converge on a view of judgment as a situated competence exercised under uncertainty and irreducible to rules, metrics, or explanations alone. They emphasize action under indeterminacy, the management of competing commitments, and the distribution of responsibility across people and systems.

What remains less developed is how such judgment operates in routine interaction with generative AI, particularly in everyday contexts where users differ widely in expertise. The moment-to-moment governance of AI participation in reasoning—when to initiate it, how to constrain it, what to stabilize, and what to endorse—has not been treated as a primary analytic object.

The present work addresses this gap by articulating recurring, observable forms of judgment in AI-mediated reasoning. From this perspective, AI functions as a TfT not by generating or justifying outputs alone, but by preserving the conditions under which users can exercise judgment over commitments in practice. The following section makes this claim concrete by identifying the recurring judgments through which such governance becomes visible.

## 3 Judgments in Thinking with AI

Judging how to think with AI involves recurring judgments through which users regulate the lifecycle of commitments in AI-mediated reasoning—how they are introduced, constrained, stabilized, examined, transformed, restricted, updated, and ultimately endorsed. These judgments are not exhaustive, nor are they stages of a fixed process. Rather, they recur throughout AI-supported workflows and are exercised unevenly depending on task demands, stakes, and expertise. Each is defined by the role it plays in commitment governance, the boundary that distinguishes it from adjacent judgments, and the traces through which it can be observed in practice.

### 3.1 Initiating

Initiating judgments govern whether AI enters the reasoning process at a given moment. **Boundary clarification:** initiation concerns whether to involve AI in a particular step, not how AI will be constrained once engaged (Bounding) or which domains should never be delegated (Limiting). **Observable traces:** deliberate sequencing (e.g., drafting before prompting), selective consultation (e.g., using AI only for critique), or explicit refusal to engage AI in high-stakes contexts. Weak initiation is often visible in AI-first defaulting across tasks regardless of stakes.

### 3.2 Bounding

Bounding judgments govern the internal scope within which AI may propose candidate commitments for a task. **Boundary clarification:** bounding constrains what AI is being asked to generate

within a delegated activity (assumptions, criteria, exclusions, format). It differs from Limiting, which sets jurisdictional boundaries on what should not be delegated at all. **Observable traces:** prompts specifying constraints, evaluation criteria, definitions, exclusions, audience, or uncertainty handling; iterative tightening of scope. Weak bounding is visible in under-specified prompts followed by post hoc adaptation of goals to fit generated output.

### 3.3 Stabilizing

Stabilizing judgments govern which candidate commitments become binding constraints for subsequent reasoning. **Boundary clarification:** stabilization concerns internal structuring of the reasoning trajectory—what will be carried forward as a working assumption, criterion, or conclusion. It differs from Owing, which concerns external endorsement and accountability. **Observable traces:** explicit statements of working assumptions; marking elements as provisional versus binding; preserving criteria across regenerations; instructing the system to maintain fixed constraints while exploring alternatives. Weak stabilization is visible when AI-generated suggestions become implicitly binding without explicit adoption.

### 3.4 Scrutinizing

Scrutinizing judgments govern what aspects of AI-proposed commitments warrant examination and what form that examination should take. **Boundary clarification:** scrutinizing concerns validation at the point of uptake. It differs from Revising, which governs how changes propagate after commitments are altered. **Observable traces:** requests for sources, counterexamples, uncertainty estimates, cross-checks, or alternative formulations; explicit decisions to defer verification in low-stakes contexts. Collapse of this judgment appears as uncritical acceptance or unfocused checking without articulated rationale.

### 3.5 Integrating

Integrating judgments govern how AI output is transformed and aligned with prior commitments in the evolving artifact. **Boundary clarification:** integrating concerns adaptation and reshaping of generated material. It differs from Stabilizing, which determines what is binding, and from Scrutinizing, which concerns validation. **Observable traces:** partial uptake of output; substantial rewriting or restructuring; synthesis of multiple generations; explicit rationale for adoption or rejection. Weak integration is visible when output is adopted wholesale with minimal transformation.

### 3.6 Limiting

Limiting judgments govern where AI participation should be restricted or withheld because a domain requires human authority, contextual insight, or value interpretation. **Boundary clarification:** limiting is jurisdictional—it marks domains in which delegation is inappropriate. It differs from Bounding, which constrains work within a delegated domain. **Observable traces:** explicit refusal to delegate value-laden or context-sensitive interpretation; separation of AI-generated material from human-endorsed claims; use of AI for options but not for final evaluative judgment.

### 3.7 Revising

Revising judgments govern how changes to commitments propagate through dependent elements of reasoning. **Boundary clarification:** revising concerns dependency management over time. It differs from Scrutinizing, which concerns validation prior to uptake. **Observable traces:** explicit revisiting of earlier sections after changing assumptions; targeted regeneration of affected components; reconciliation of inconsistencies; updating criteria before further generation. Weak revising appears as wholesale regeneration or superficial patching without tracing implications.

### 3.8 Owing

Owing judgments govern which commitments the user will endorse and be accountable for in consequential contexts. **Boundary clarification:** owing concerns external authorship and responsibility. A commitment may structure internal reasoning (Stabilizing) without being something the user is prepared to stand behind publicly or professionally. **Observable traces:** explicit endorsement or qualification of claims; insertion of provenance or justification before sharing; rejection of “the AI said” as sufficient warrant; final risk or responsibility checks prior to dissemination.

**Table 1: Judgments in thinking with AI as a tool for thought.**

Judgment	Brief Definition (What is Judged)
Initiating	Whether AI should participate in shaping commitments at a given moment.
Bounding	The scope within which AI may propose candidate commitments for a task.
Stabilizing	Which candidate commitments become binding constraints for subsequent reasoning.
Scrutinizing	What aspects of AI-proposed commitments warrant validation before uptake.
Integrating	How AI-generated commitments are transformed and aligned with prior commitments.
Limiting	Where AI participation should be restricted because delegation is inappropriate.
Revising	How established commitments should be updated and propagated when conditions change.
Owing	Which commitments the user will endorse and take responsibility for.

*Scope and operationalization.* In this framework, judgment is treated primarily as epistemic and responsibility-bearing: judgments about what to accept as warranted, what to stabilize as binding, what to scrutinize, and what one is prepared to endorse. For example, in an AI-supported writing workflow, a user might initiate AI only after drafting, bound it to generating counterarguments, scrutinize a cited claim, revise the text in light of that check, and ultimately own the final argument. Such judgments can be examined in prompt histories, revision traces, versioned documents, annotation trails, and retrospective walkthroughs of AI-mediated work. They also suggest concrete design directions, such as making assumptions explicit, preserving revision paths, and supporting principled non-delegation in value-laden contexts.

## 4 Implications for Tools for Thought

**Reframing evaluation.** The judgments articulated above reframe what it should mean for generative AI to function as a TfT. This shifts evaluation from output performance to judgmental structure. From this view, TfTs are defined less by what they generate than by how they structure initiating, bounding, stabilizing, scrutinizing, integrating, limiting, revising, and owning.

**Expertise asymmetries.** One implication is that superficially similar AI systems may differ substantially in their cognitive effects depending on how they structure these judgments. Workflows that default to immediate generation may suppress initiating and bounding, whereas those that require articulation of constraints or provisional commitments can foreground them [10]. Systems that present fluent output without mechanisms for selective scrutiny or explicit ownership risk collapsing stabilizing, scrutinizing, and owning into passive acceptance. This framing also clarifies why expert users may report benefit from AI while novices experience cognitive short-circuiting. Experts are more likely to actively govern commitments—bounding contributions, selectively scrutinizing output, stabilizing assumptions, and revising them when conditions change—whereas novices are more prone to letting generated output stand in for these acts. This asymmetry is not reducible to trust calibration or model understanding alone; it reflects differences in how judgment is exercised and supported in practice.

**Design implications.** The framework also points toward concrete interaction patterns that differ from conventional decision support. Systems can be designed to make assumptions explicit, preserve provisional commitments across iterations, prompt selective scrutiny of high-risk claims, and create ownership checkpoints before results are shared or acted upon. Consider a user drafting an analytical memo with AI support: one workflow prompts immediately for a finished draft and accepts a fluent response with minimal scrutiny, whereas another begins with a user-generated outline, bounds AI to generating counterpoints, scrutinizes a cited claim, revises the text, and ultimately owns the final position. The outputs may appear similarly coherent, yet the underlying judgments—and thus the cognitive consequences—are quite different. From this perspective, evaluation should attend not only to final output quality, but also to whether commitments are made explicit, selectively scrutinized, revisable, and ultimately owned by the user.

**Analytic value.** Finally, the taxonomy provides a lens for analyzing AI-mediated workflows without reducing evaluation to friction, automation, or performance metrics alone. It enables researchers and designers to ask targeted questions—such as which judgments are suppressed, displaced, or overloaded—and to examine how systems shape the governance of commitments over time. In this sense, the framework offers a vocabulary for reasoning about cognitive engagement that complements existing approaches to evaluation while clarifying what it means for AI to genuinely function as a tool for thought.

## 5 Judgment and Cognitive Integrity

Understanding AI as a TfT requires attention to the judgments through which users govern commitments over time. This perspective complements work on explainability and appropriate reliance. Explainability addresses whether systems can make their processes

visible; judgment concerns whether users remain responsible for managing commitments in use. An AI system may provide detailed explanations yet still erode judgment if those explanations do not preserve the conditions under which commitments can be examined, revised, and owned. What matters is not only whether reasoning is transparent, but whether it remains governable.

The framework also aligns with traditions that treat cognition as situated and distributed across people and artifacts, while maintaining that responsibility for judgment remains human. AI may participate in generating and transforming commitments, but it does not assume accountability for them. Judgment remains a human responsibility even when AI is deeply involved in shaping the course of reasoning. The central question, then, is not whether AI can think, but whether users retain the capacity to judge how thinking proceeds.

Framed in these terms, cognitive integrity refers to the preservation of conditions under which commitments can be responsibly initiated, stabilized, scrutinized, revised, and owned over time. AI can function as a TfT to the extent that it sustains this integrity—structuring workflows that make judgment possible and salient rather than reducing it to passive acceptance. Because such scaffolding is not neutral, it should also preserve opportunities for refusal, uncertainty marking, and non-delegation in contexts where responsibility cannot be meaningfully offloaded to AI. Evaluating AI as a TfT therefore requires asking not only whether systems perform well, but whether they preserve the governance of commitments that underwrites accountable, revisable reasoning—particularly for those still developing expertise.

## 6 Conclusion and Future Work

I have argued that AI functions as a tool for thought only insofar as it preserves users' judgment in governing commitments over time. By articulating a set of recurring acts—initiating, bounding, stabilizing, scrutinizing, integrating, limiting, revising, and owning—I have reframed evaluation away from output quality alone and toward the lifecycle of commitments that underwrites accountable and revisable reasoning. The contribution is analytic, as it makes visible how AI becomes coupled to human reasoning and how that coupling may sustain or erode cognitive integrity.

Future work can extend this framework in several directions. First, trace-based empirical studies of AI-mediated workflows, particularly across expertise levels, could examine how these judgments are exercised or displaced in practice. Second, comparative analyses of alternative workflow configurations could clarify how different interaction patterns structure commitment governance without presuming that greater friction or automation is inherently better. Finally, the concept of cognitive integrity warrants further theoretical refinement, specifying the conditions under which AI-mediated reasoning remains responsibly initiated, examined, revised, and owned over time.

## References

- [1] Lisanne Bainbridge. 1983. Ironies of Automation. *Automatica* 19, 6 (1983). doi:10.1016/B978-0-08-029348-6.50026-9
- [2] Peter Dalsgaard. 2025. Thinking through Prompting: Cognitive Mediation in Human–AI Interaction. In *Proceedings of the 36th Annual Conference of the European Association of Cognitive Ergonomics*. ACM, Tallinn, Estonia, 1–6. doi:10.1145/3746175.3747192

- [3] Joseph Dunne. 1997. *Back to the Rough Ground: Practical Judgment and the Lure of Technique*. University of Notre Dame Press. Google-Books-ID: plwFDgAAQBAJ.
- [4] Mica R. Endsley. 2023. Ironies of artificial intelligence. *Ergonomics* 66, 11 (Nov. 2023), 1656–1668. doi:10.1080/00140139.2023.2243404
- [5] Colin M. Gray, Cesur Dagli, Muruvvet Demiral-Uzan, Funda Ergulec, Verily Tan, Abdullah A. Altuwaijri, Khendum Gyabak, Megan Hilligoss, Remzi Kizilboga, Kei Tomita, and Elizabeth Boling. 2015. Judgment and Instructional Design: How ID Practitioners Work In Practice. *Performance Improvement Quarterly* 28, 3 (2015), 25–49. doi:10.1002/piq.21198
- [6] Sunnie S. Y. Kim, Jennifer Wortman Vaughan, Q. Vera Liao, Tania Lombrozo, and Olga Russakovsky. 2025. Fostering Appropriate Reliance on Large Language Models: The Role of Explanations, Sources, and Inconsistencies. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–19. doi:10.1145/3706598.3714020
- [7] Gary A. Klein. 2017. *Sources of power: How people make decisions* (2 ed.). MIT Press.
- [8] Harold G. Nelson and Erik Stolterman. 2012. *The Design Way: Intentional Change in an Unpredictable World* (second ed.). The MIT Press, Cambridge, Massachusetts; London, England.
- [9] Paul Parsons, Colin M. Gray, Ali Baigelenov, and Ian Carr. 2020. Design Judgment in Data Visualization Practice. In *IEEE Visualization Conference (VIS)*. Salt Lake City, UT, 176–180. doi:10.1109/VIS47514.2020.00042
- [10] Leon Reicherts, Zelun Tony Zhang, Elisabeth Von Oswald, Yuanting Liu, Yvonne Rogers, and Mariam Hassib. 2025. AI, Help Me Think—but for Myself: Assisting People in Complex Decision-Making by Providing Different Kinds of Cognitive Support. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama, Japan, 1–19. doi:10.1145/3706598.3713295
- [11] Lucy A. Suchman. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press.
- [12] David D. Woods and Erik Hollnagel. 2006. *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*. CRC Press.
- [13] Chunpeng Zhai, Santoso Wibowo, and Lily D. Li. 2024. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments* 11, 1 (June 2024), 28. doi:10.1186/s40561-024-00316-7