

# Felt but Not Seen: Design Patterns from Building a Metacognitive Writing Tool

Sergio Abraham

Auster Center for Applied Innovation and Research  
Tufts University  
Medford, MA, USA  
Sergio.Abraham@tufts.edu

Filip Čučkov

Auster Center for Applied Innovation and Research  
Tufts University  
Medford, MA, USA  
Filip.Cuckov@tufts.edu

## Abstract

Most AI tools personalize around task workflows, adapting to what users are doing rather than how they are thinking. Building a writing augmentation prototype, we discovered that this task-centric paradigm conflicts with the goal of cognitive augmentation. Cognitive augmentation requires AI that extends metacognition rather than generating on the user’s behalf: present but not visible, reflecting but not generating, momentary but not conversational. From this experience, we extract a broader design principle: when tools for thought adapt to cognitive states rather than task workflows, personalizing around the user’s metacognitive rhythm and attention capacity, they augment rather than displace the user’s thinking. We operationalize this principle through two design patterns. *Presence-without-Visibility* uses behavioral process observation, detecting hesitation, deletion, and avoidance, to maintain cognitive awareness without demanding attention, manifested through configurable entry points with user-controlled proximity. *Moments-over-Conversations* replaces threaded dialogue with atomic interactions that match the unpredictable rhythm of cognition while the system silently maintains continuity. We discuss the design considerations these patterns create when they collide with current paradigms and identify open questions for the tools-for-thought community.

## Keywords

tools for thought, cognitive augmentation, metacognition, design patterns, human-AI interaction

## 1 Introduction

Generative AI is transforming knowledge work, and a growing body of research examines how these tools affect human cognition: LLM-delivered synthesis inhibits the sensemaking process essential for deep learning [6], AI-generated content is significantly more homogeneous across users than human-generated content [3], and AI boosts creative performance primarily for users with strong metacognitive strategies [10]. In response, the tools-for-thought (TFT) community has called for AI systems that protect and augment cognition rather than merely accelerating task completion [12]. Yet most current AI tools share a structural assumption that may undermine this goal: they personalize around tasks. Writing tools analyze text and generate drafts. Coding tools complete functions. Even tools designed for cognitive support tend to

adopt the conversational paradigm, structuring interaction as a threaded dialogue between user and AI.

We argue that this task-centric paradigm creates a fundamental misalignment for tools for thought. Thinking does not follow task boundaries. A writer mid-paragraph may shift from composing to questioning their own argument to reconsidering their audience to avoiding a difficult section. These are cognitive state transitions, and they happen at a rhythm the user may not be fully aware of. When AI tools impose task-based interaction models, they structure thought around the tool’s cadence rather than the user’s own rhythm.

In this paper, we make two contributions. First, we propose that when tools for thought adapt to cognitive states rather than task workflows, personalizing around the user’s attention capacity and readiness to engage, the result is augmented original thought rather than offloading of cognitive tasks. Second, we present two design patterns that operationalize this principle, discovered through building a writing augmentation prototype: *Presence-without-Visibility*, in which the AI maintains awareness through behavioral process observation while remaining imperceptible until the user invites engagement, and *Moments-over-Conversations*, in which each interaction is experienced as an independent, atomic exchange while the system silently accumulates context. We discuss the trade-offs, challenges, and paradoxes these patterns create, and we identify open questions for the community.

## 2 Background and Related Work

The vision of computing as cognitive augmentation predates generative AI. Engelbart [4] proposed that technology should amplify human capability rather than replace it. Brynjolfsson [2] formalized this distinction for AI, arguing that augmentation creates more value and distributes it more broadly than automation. Large-scale empirical evidence supports this: the Stanford WORKBank study [8], surveying 1,500 workers across 104 occupations, found that in 45.2% of occupations, workers preferred human-AI partnership as their collaboration mode. Our work contributes to this tradition at the interaction design level: how specific design patterns can operationalize augmentation in cognitive tools.

Recent work has begun to articulate what this requires from AI systems. Sarkar [7] argued that AI should challenge users rather than comply, proposing the provocateur as an alternative to the compliant assistant. Tankelevitch et al. [11] identified the metacognitive demands that generative AI

places on users, from prompt formulation through output evaluation. We build on both contributions. If generative AI creates these demands, and if challenging users supports deeper thinking, then supporting the user's own reflective process becomes the primary function of a tool for thought. We operationalize this by treating AI as an extension of the user's thinking, a design choice that shapes both the interaction model and the experience of cognitive ownership.

The dominant interaction model for generative AI is conversational: a threaded dialogue where user and AI exchange messages in sequence. This paradigm has clear strengths, including flexibility and natural language interaction, but it also carries implicit assumptions about cognition. Conversations accumulate context visibly, creating a shared history that the user must manage. They impose temporal structure: each exchange invites a response. Research on cognitive offloading, notably the Google Effect [9], suggests that when people expect external access to information they invest less in encoding it, and the writing process literature [5] has established that the cognitive work of composing and evaluating is where understanding emerges. These findings raise a question: might alternative interaction paradigms better preserve the cognitive engagement that augmentation requires?

While the field has articulated the goals of cognitive augmentation (challenging users, supporting metacognition, preserving ownership) and the risks of failing to meet them (cognitive offloading, homogenization, sensemaking inhibition), less work has addressed what specific interaction patterns operationalize these goals and what happens when those patterns collide with user expectations. We address this gap with design knowledge from a prototype that deliberately departs from current conventions.

### 3 From Building to Principle

We initially developed a writing augmentation prototype under a deliberate constraint: the AI should never generate text for the user. Across several iterations, including content analysis, sidebar suggestions, and conversational interfaces, we observed a consistent pattern. Each approach improved output quality while reducing the writer's cognitive ownership. Writers could not explain or defend their arguments without re-reading them. The text was better, but the thinking behind it was shallower.

The consistency of this result across different feature sets suggested the problem was structural. Each iteration treated writing as a collection of sub-tasks (argumentation, evidence, structure) that the AI could support. But writing is a non-linear flow of cognitive states: composing, questioning, evaluating, hesitating, avoiding, among others. These states have different needs. A writer composing fluently needs different support than one who is stuck, and one avoiding a difficult section needs something different still. Our designs, organized around the task, had no mechanism for recognizing these states. They responded to the product of thinking (the text) rather than the process of thinking (the behavior).

This led us to a core design principle for cognitive augmentation: adapt to cognitive states rather than task workflows. The personalization axis should be the user's oscillation between composing and evaluating, flowing and stuck, confident and uncertain, rather than the content they produce. This reframes the AI's role: instead of a task assistant, it operates as an extension of the user's own inner reflective process. Like the inner voice that monitors one's own thinking, the AI is always present and always observing, but conscious engagement with it remains voluntary. The following two design patterns operationalize this principle.

## 4 Design Patterns

We present two patterns that operationalize this principle, described at a level of abstraction intended to be transferable beyond our specific implementation: the first, Presence-without-Visibility, addresses how the AI maintains awareness and makes itself available; the second, Moments-over-Conversations, addresses the structure of the interaction. Working together, Presence-without-Visibility provides the sensing and signaling layer, and Moments-over-Conversations governs what happens when the user engages.

### 4.1 Presence Without Visibility

The AI maintains continuous awareness of the user's cognitive process without demanding attention. We describe this as "felt but not seen." This contrasts with two existing paradigms: fully invisible AI (background processing with no user awareness) and fully visible AI (chat panels, sidebars, notification systems). Presence-without-Visibility occupies a third position: the user knows something is there, but nothing in the interface confirms it until a moment is triggered.

**Behavioral process observation.** The sensing mechanism is behavioral rather than textual. Instead of analyzing what the user has written, the system observes how they are writing. Signals include typing rhythm changes, extended pauses, repeated deletions, avoidance of specific sections, and sudden speed changes. These serve as proxies for cognitive states: hesitation may signal uncertainty, repeated deletion may signal dissatisfaction with one's reasoning, avoidance may signal discomfort with a topic. The system reads the user's process analogously to how their own reflective awareness monitors their thinking, but externalized.

**Entry points into the moment.** When the system detects a signal worth surfacing, it manifests through one of three configurable entry points: an ambient visual signal (a subtle peripheral change in the workspace), an ambient auditory signal (a soft sound), or user invocation (the user actively calls the AI). The first two are AI-initiated through different sensory channels, both deliberately designed to avoid the dopamine-triggering notification patterns common in social media and productivity tools. The design intent is for these signals to function as persistent environmental states rather than events: something the user may notice when their attention naturally drifts to the periphery, rather than something that announces its arrival. The third entry point

is user invocation, the only user-initiated channel. The user configures their preferred entry point to match their working style.

Regardless of entry point, every interaction follows the same structure: a single constrained exchange, one user message allowed, then the moment resolves. The AI's final utterance is always a closing thought, a statement that settles rather than a question that demands continuation. If the user wants to engage again, they must manually trigger a new moment. This is a deliberate anti-dependency mechanism: it prevents open-ended engagement loops, creates a micro-pause that returns ownership of the thinking process to the user, and ensures the AI steps back after each exchange rather than sustaining its presence.

**The nature of the reflection.** What the AI surfaces is neither a suggestion nor a directive. It is a reflection: a comment or question designed to trigger the user's own thinking rather than provide an answer. "Are you avoiding this section because the argument isn't clear to you yet?" does not require a typed response. It requires the writer to sit with it. The user may respond once, after which the AI offers a closing thought and the moment resolves. The interaction structure is therefore at most three turns: the AI opens, the user responds, the AI closes. This creates a specific duality: the moment settles (it closes, demands no follow-up, leaves no thread) while its content may unsettle (it opens a cognitive gap the writer must process on their own terms). The AI's contribution ends. The thinking it triggered does not.

This constrained interaction structure is where the first pattern connects to the second: every entry point leads to the same kind of interaction, which we call a moment.

## 4.2 Moments Over Conversations

Each AI interaction is experienced by the user as independent, complete, and atomic. There is no visible thread, no conversation history, no accumulated dialogue to manage. From the user's perspective, each interaction arrives fresh. From the system's perspective, however, context accumulates silently across interactions, much as human memory shapes present thinking without requiring conscious recall.

**Cognitive rhythm matching.** Conversation imposes temporal structure: the user speaks, the AI responds, the user responds to the response. This cadence shapes when the user thinks, about what, and for how long. Thinking does not follow this rhythm. A writer may need five minutes of uninterrupted flow, then thirty seconds of self-questioning, then ten minutes of reorganization. A "moment" occurs when the AI's process observation detects a state that may benefit from reflection and the user chooses to engage. The moment resolves and leaves no residue. The next moment may come in two minutes or two hours. The rhythm belongs to the user.

**Invisible continuity.** The system retains everything. Each interaction informs subsequent observations: the AI recognizes that a passage has been rewritten three times, that the user hesitates before conclusions, that certain argument structures recur. This accumulated understanding

shapes each new moment without the user managing it. Memory functions like memory: present in its effects, absent from conscious management.

Accumulated visible context, a feature in conversational AI tools, becomes a liability for tools for thought: it shifts cognitive load from the primary task (thinking, writing) to the secondary task (managing the AI interaction). Atomic moments eliminate this secondary load. The user's working memory remains dedicated to their own thinking. The closest analogue in HCI is peripheral interaction [1], extended here from ubiquitous computing to cognitive AI tools and combined with invisible context accumulation.

## 5 Navigating the Design Space

These patterns emerged through iterative work, and each iteration revealed difficult decisions where our principles met expectations shaped by existing interaction conventions. We present three that we believe generalize beyond our specific implementation.

### 5.1 The Discoverability Trade-off

If the AI is felt but not seen, how do users learn it exists? Traditional onboarding relies on visible features. Our pattern has nothing to show. The first moment of engagement must carry disproportionate weight: the system must demonstrate value through a single, well-timed reflection the user did not expect but recognizes as useful. This requires perceptiveness during early interactions while operating with minimal accumulated context. The trade-off between patience (waiting for the right moment to surface a genuinely useful reflection) and urgency (proving value before the user abandons the tool) remains difficult to resolve. We prioritize patience, accepting that some users may never discover the AI's presence, reasoning that a premature first interaction would undermine trust more than a delayed one. This discoverability problem also intersects with the privacy considerations of felt-but-not-seen presence, which we discuss in Section 5.2.

### 5.2 The Invisible Continuity Challenge

The system accumulates behavioral context silently, which is necessary for moments to deepen in quality but raises questions about perceived surveillance. Observing how someone writes (their hesitations, deletions, avoidance patterns) feels more intimate than analyzing what they wrote. Moreover, these signals vary by writing context: casual writing may flow with fewer hesitations, while academic writing involves deliberate slowness where each statement requires evidential support, producing different behavioral profiles for similar cognitive states. The content of a text is something the writer chose to make visible. The process of producing it is something the writer may consider private. Transparency about what categories of behavior are observed, combined with local-only processing where no data leaves the device, partially addresses this. But the boundary between supportive and intrusive sensing likely varies by individual, suggesting that user control over observation granularity is essential.

A related question is how the system ensures accumulated context remains relevant over time, avoids stale assumptions, and gives users meaningful control without requiring them to manage the context directly. Designing this control without undermining the system's ability to detect meaningful cognitive signals remains an open problem.

### 5.3 The Cognitive Friction Paradox

Users approach AI tools with productivity expectations. A tool that observes instead of suggesting, that requires the user to do the cognitive work, appears to offer less. This is compounded by deliberate choices: the entry points avoid dopamine-triggering patterns, and the one-message constraint prevents conversational loops. The tool offers less engagement because minimizing engagement is part of the cognitive protection. Writers who already valued their own thinking process engaged readily, while those seeking efficiency found the approach frustrating. This raises a question with equity implications: can a tool for thought develop reflective capacity in users who lack it, or does effective use require a baseline? If the latter, cognitive augmentation tools risk serving only those who are already cognitively privileged.

There is a second, less deliberate source of friction. The single-interaction constraint assumes that each moment arrives at the right time: the AI detects a meaningful cognitive state and surfaces a reflection worth engaging with. When sensing is accurate, the constraint works as designed—the moment resolves, the user returns to their thinking, and the friction is productive. When sensing is crude, the same constraint becomes a liability. A reflection that misreads the user's state (interrupting flow rather than recognizing a pause, or surfacing a prompt when the user is composing rather than stuck) cannot be corrected within the moment, because the interaction structure does not permit clarification or iteration. The user experiences friction without cognitive benefit. In our implementation, early iterations used a fixed inactivity threshold to trigger moments, producing false positives that were not just unhelpful but disruptive—the tool intervened during deliberate pauses rather than genuine uncertainty. This suggests that the viability of the single-interaction constraint depends on the sophistication of the sensing layer: the more accurately the system distinguishes between cognitive states, the less the constraint feels like a limitation. Designing sensing mechanisms that can distinguish deliberate slowness from genuine uncertainty, particularly across different writing contexts where the same behavioral signal may indicate different cognitive states, remains an active area of our implementation work.

## 6 Discussion and Open Questions

These patterns raise questions that extend beyond our specific implementation. We discuss three that we believe are relevant to the tools-for-thought community.

The first concerns measurement. How do we evaluate whether Presence-without-Visibility supports reflective thinking versus being simply ignored? Standard usability metrics

capture task outcomes but miss the cognitive process. A user who produces excellent text may have engaged deeply with their own thinking or may have passively accepted the AI's framing. These outcomes look identical in task metrics but differ fundamentally from an augmentation perspective. Evaluating cognitive-state-responsive tools may require new approaches, potentially drawing on think-aloud protocols, physiological sensing, or longitudinal self-report, that track the quality of the thinking process alongside the quality of its products.

The second concerns generalizability. Our patterns emerged from writing, where behavioral signals (typing rhythm, editing behavior, structural reorganization) are relatively observable. Decision-making, learning, and creative work involve cognitive state transitions, but the behavioral traces differ. A decision-maker deliberating in a spreadsheet, a student reading a textbook, and a designer working on a canvas all think in ways that may be detectable, but the nature and reliability of those signals require investigation. The underlying principle may apply across domains even as sensing mechanisms change. Whether a generalizable taxonomy of cognitive-state signals can be developed, or whether such signals are inherently domain-specific, is a meaningful research direction.

The third concerns accessibility. Section 5.3 identified that cognitive augmentation tools may inherently serve users who already possess reflective awareness. One possible design intervention is progressive disclosure of the AI's reflective process: making metacognitive sensing gradually visible to scaffold awareness in users who do not yet practice it. Whether such scaffolding can bootstrap metacognitive capacity, or whether it undermines the invisibility that makes the patterns work, is an open question.

## 7 Conclusion

We presented two design patterns for tools for thought. *Presence-without-Visibility* maintains cognitive awareness through behavioral process observation while remaining imperceptible until the user invites engagement. *Moments-over-Conversations* replaces threaded dialogue with atomic interactions that preserve the user's thinking rhythm while the system accumulates context invisibly. Both operationalize a single principle: tools for thought augment cognition when they adapt to cognitive states rather than task workflows.

These patterns are preliminary. They emerged from one domain, one prototype, and a limited set of user interactions. Their value at this stage is as concrete departures from current interaction paradigms that open new areas of the design space for cognitive augmentation. The trade-offs, challenges, and paradoxes we encountered are, we believe, as informative as the patterns themselves.

Our aim is for this work to contribute to the investigation of cognitive-state-responsive design, and we look forward to the workshop discussion.

## References

- [1] Saskia Bakker, Elise van den Hoven, and Berry Eggen. 2015. Peripheral Interaction: Characteristics and Considerations. *Personal and Ubiquitous Computing* 19, 1 (2015), 239–254.
- [2] Erik Brynjolfsson. 2022. The Turing Trap: The Promise and Peril of Human-Like Artificial Intelligence. *Daedalus* 151, 2 (2022), 272–287.
- [3] Anil R. Doshi and Oliver P. Hauser. 2024. Generative AI Enhances Individual Creativity but Reduces the Collective Diversity of Novel Content. *Science Advances* 10, 28 (2024).
- [4] Douglas C. Engelbart. 1962. Augmenting Human Intellect: A Conceptual Framework. Stanford Research Institute, Menlo Park, CA.
- [5] Linda Flower and John R. Hayes. 1981. A Cognitive Process Theory of Writing. *College Composition and Communication* 32, 4 (1981), 365–387.
- [6] Shiri Melumad and Jin Ho Yun. 2025. Experimental Evidence of the Effects of Large Language Models versus Web Search on Depth of Learning. *PNAS Nexus* 4, 10 (2025).
- [7] Advait Sarkar. 2024. AI Should Challenge, Not Obey. *Communications of the ACM* 67, 10 (2024), 18–21.
- [8] Yijia Shao, Humishka Zope, Yucheng Jiang, Jiaxin Pei, David Nguyen, Erik Brynjolfsson, and Diyi Yang. 2025. Future of Work with AI Agents: Auditing Automation and Augmentation Potential across the U.S. Workforce. *arXiv preprint arXiv:2506.06576* (2025).
- [9] Betsy Sparrow, Jenny Liu, and Daniel M. Wegner. 2011. Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. *Science* 333, 6043 (2011), 776–778.
- [10] Shuhua Sun, Angelina Zhuyi Li, Maw Der Foo, Jing Zhou, and Jackson G. Lu. 2025. How and For Whom Using Generative AI Affects Creativity: A Field Experiment. *Journal of Applied Psychology* (2025).
- [11] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The Metacognitive Demands and Opportunities of Generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. ACM.
- [12] Lev Tankelevitch, Elena L. Glassman, Jessica He, Majeed Kazemitabaar, Aniket Kittur, Mina Lee, Srishti Palani, Advait Sarkar, Gonzalo Ramos, Yvonne Rogers, and Hari Subramonyam. 2025. Tools for Thought: Research and Design for Understanding, Protecting, and Augmenting Human Cognition with Generative AI. In *CHI '25 Extended Abstracts*. ACM.