Thought as a Substrate in Human-AI Interaction

Xingyu Bruce Liu
University of California, Los Angeles
Los Angeles, CA, USA
xingyuliu@ucla.edu

Abstract

I propose to conceptualize "thought" as a new substrate in human-AI interaction. While there is extensive research in the NLP and ML communities on augmenting large language models (LLMs) with "thinking" capabilities, these efforts primarily focus on improving AI's reasoning performance. As an HCI researcher, I explore how enabling AI to generate and utilize thought can unlock new capabilities for interacting with humans and introduce new paradigms for human-AI interaction. I argue that this conceptualization opens up new possibilities for human-AI interaction by supporting proactive AI behavior, enabling continuous alignment with user intent, and fostering more dynamic and adaptive interaction experiences. In this paper, I articulate the conceptual foundations of thought as a substrate in human-AI interaction, demonstrate its role through system examples, and envision how this paradigm could shape the future of human-AI collaboration.

CCS Concepts

• **Human-centered computing** → **HCI theory, concepts and models**; **Interactive systems and tools**.

Keywords

Thought, Large Language Models, Agent, Human-AI Interaction, Theory, Interaction Paradigm, Substrate

ACM Reference Format:
Xingyu Bruce Liu. 2025. *Thought* as a Substrate in Human-AI Interaction. In *The 38th Annual ACM Symposium on User Interface Software and Technology (UIST Adjunct '25), September 28-October 1, 2025, Busan, Republic of Korea.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3746058.3758466

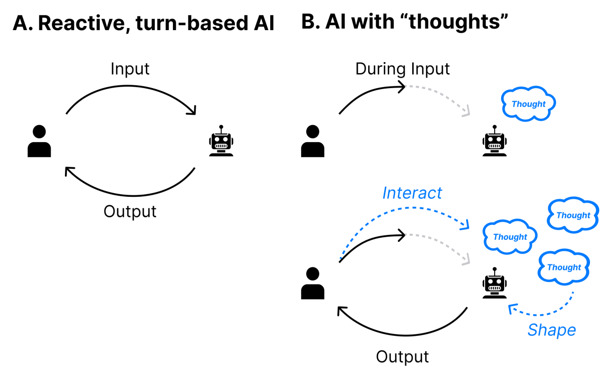

Figure 1: In contrast to (A) traditional AI that is reactive, responding only when prompted, AI thinks continuously, proactively generating, iterating, and allowing users to interact with its thoughts (B).

1 Introduction

In 1950, Alan Turing famously asked, *"Can machines think?"* [12] This question has inspired decades of research in artificial intelligence (AI), with efforts ranging from symbolic reasoning and expert systems to today's large language models (LLMs). Early philosophical and computational models of cognition viewed thought as a defining characteristic of intelligence—whether in humans or machines [2, 8, 9, 11]. Recent LLMs have reignited discussions about what it means for an AI to "think". Techniques such as Chain-of-Thought prompting [13] demonstrate that generating intermediate reasoning steps can significantly enhance model performance on complex tasks. Building on this foundation, models like OpenAI o1 [6], Gemini Flash Thinking [3], and DeepSeek R1 [4] integrate reasoning into their training processes. These developments mark substantial progress in AI's ability to problem-solve.

However, despite these advancements, current human-AI interaction paradigms remain fundamentally constrained. Existing systems still operate within a turn-based, input-output framework: Users issue a prompt, the AI generates a single response, and the cycle repeats. This limitation becomes evident in everyday AI interactions. Imagine a user asking ChatGPT: "Help me plan a surprise birthday party for my friend." ChatGPT would generate a static response—perhaps listing general steps like choosing a venue, sending invitations, and selecting a menu. At this point, the AI has already stopped processing and just passively waits for the user's next input. If the user later realizes they need a budget-friendly vegan restaurant for six people in New York, they must start a new turn, manually refining their query.

In contrast, consider instead how humans naturally collaborate during a brainstorming session. Each participant holds an internal train of thought that continuously updates in the background, even while someone else is speaking. In parallel, they may externalize immature, intermediate thoughts by sketching or annotating on a whiteboard, inviting immediate refinements and comments from others. This fluid interaction unfolds dynamically as thoughts are developed, shared, and restructured.

UIST Adjunct '25, Busan, Republic of Korea

ACM ISBN 979-8-4007-2036-9/2025/09
https://doi.org/10.1145/3746058.3758466

I argue that we should conceptualize "thought" as a substrate in the human-AI interaction process. It should be viewed as a first-class citizen that is not merely hidden computation steps but an *interaction component* that should be explicitly considered when designing and building human-AI systems. I highlight that thought, in the human-AI interaction process:

(1) provides an *intermediate medium*, enabling users to observe and interact with AI's intermediate thinking rather than just its final outputs.
(2) enables a *full-duplex process*, where AI and users exchange fluidly rather than in rigid, turn-based exchanges.
(3) serves as an *intrinsic driver*, allowing AI to initiate interactions rather than merely responding to queries.
(4) establishes a *shared space*, where AI and users build upon each other's thoughts in a collaborative, dynamic process.

In the sections that follow, I first outline the conceptual foundations of AI thought. I then provide concrete examples that illustrate how it can bring new possibilities to human-AI interaction, offering benefits like enabling proactive AI behavior, facilitating continuous alignment with users, and fostering more dynamic interaction experience. I also discuss the broader implications of this paradigm shift, and explore how "thinking"—long considered a uniquely human trait—can transform our ways to interact with machines in the future.

2 AI Thought in the Context of HCI

Discussions around "thought" in LLM research often focus on a chain-of-thought approach, where a model generates intermediate steps to improve performance on reasoning tasks. From an HCI perspective, however, *thought* can be understood more broadly. I define AI's *thought* as:

Definition 2.1. **Thought** refers to a system's *intermediate processes and responses* that can be continuously generated, developed, and selectively communicated over the course of an interaction.

In contrast to LLM-based definitions, my conception of thought emphasizes how these ongoing processes shape the system's behavior and its capacity to interact with human. Thoughts can emerge at any point in an interaction, triggered by external stimuli or intrinsic reflection, may be expressed or remain internal, and they can take various forms, from abstract keywords to visual or auditory representations. I highlight four key traits of thought (Table 1) in the context of human-AI interaction:

Trait	Current AI	AI + Thought
Intermediate Medium	Reveals only final outputs	Surfaces intermediate thoughts in real time
Full-Duplex Process	Waits for user prompts	Continuously thinks in parallel with user activity
Intrinsic Driver	Responds only when asked	Self-initiates actions based on thoughts
Shared Cognitive Space	Turn-based exchanges	Builds on collaborative fragmented ideas

Table 1: Comparison of AI with thought and Current AI across key traits.

2.1 Traits of Thought

2.1.1 Intermediate Medium. A fundamental distinction between a *thought* and a *response* is that a thought is *in-progress*, potentially fragmented, light-weight, and not necessarily intended for final output. This process, whether it involves brainstorming ideas, weighing pros and cons, or exploring parallel strategies, can be treated as an *intermediate medium* that the AI may optionally share.

2.1.2 Full-duplex Process. Human conversation does not pause for one party to "finish thinking", nor do humans stop thinking when others start to speak. Similarly, thought enables a *full-duplex process* where thinking is continuous and can occur any point in the interaction, rather than locked into turn-based cycles. AI thoughts may be triggered by user inputs or by the system's own internal reflections; they can also evolve without any external stimuli, for example, during periods of user silence.

2.1.3 Intrinsic Driver. One of the key aspects of human intelligence is that thought is not merely reactive: it also serves as an *intrinsic driver* of action. Similarly, AI's thought can serve as a mechanism that enables the system to self-initiate its interactions and be *proactive*. Rather than being solely triggered by user input, with thought they can independently generate ideas, monitor evolving contexts, and identify opportunities to intervene or contribute.

2.1.4 Shared Cognitive Space. Finally, a *shared cognitive space* could emerge when AI and user co-exist in an ongoing thought process. Rather than the AI presenting a single answer and the user responding with a single set of followups, both parties iteratively build upon each other's partial ideas. We envision that the AI's intermediate thoughts, user feedback, clarifying questions, and real-time refinements etc. form a collaborative "thinking canvas."

2.2 Implications for HCI

The four traits of thought —an *intermediate medium*, *full-duplex process*, *intrinsic driver*, and a *shared cognitive space*—open up new possibilities for how humans interact with AI. Below, we discuss four implications:

2.2.1 From Passive Respondents to Proactive Participants. The first implication is the transition from AI as a passive respondent to an active participant in interactions. Conventional AI systems operate in a reactive manner, awaiting user input before generating a response. In contrast, AI can continuously generates thoughts, enabling it to self-initiate actions and engage more dynamically with users.

This proactivity manifests in two ways. First, the *full-duplex process* (Trait 2) allows the AI to generate thoughts in parallel with user input, meaning it does not need to remain idle between interactions. It can detect emerging needs, anticipate user queries, and even interrupt when necessary. Second, thoughts equip AI with an *intrinsic drive* (Trait 3) to contribute based on its internal cognitive processes. This ability to self-motivate and intervene resembles human-like initiative, distinguishing it from traditional systems that rely on heuristics or predefined triggers [5].

2.2.2 Continuous Cognitive Alignment. A second major implication for HCI arises from AI's ability to reveal *in-progress* ideas (Trait 1: *Intermediate Medium*) continuously (Trait 2: *Full-duplex Process*). Rather than presenting a single, static final output, the system selectively shares partial thoughts. Users can then not only observe these thoughts for better interpretability, but also guide the AI's thinking early in the process, leading to more dynamic collaboration.

This helps bridge the "Gulf of Evaluation," where users struggle to understand how or why a system arrived at a particular result [10]. Here, the AI's partial thoughts provide visibility into its evolving path, allowing users to catch misconceptions or supply additional context before errors compound. Similarly, users can steer the AI's next steps more effectively, shrinking the "Gulf of Execution," since they can articulate what they want *as* they see the AI's tentative directions.

This also helps build common ground [1] between human and AI, aligning with established theories of communication that emphasize the importance of shared context for effective collaboration. By understanding the AI's intermediate thought, users gain a clearer understanding of the system's current assumptions and partial conclusions. This shared cognitive workspace drives continuous alignment between the user's thinking and AI's thinking process.

2.2.3 Beyond Turn-based Interaction. AI with thought prompts us to reconsider the fundamental structure of human–AI interaction. Traditional chatbot interfaces operate in discrete, back-and-forth messages, with the AI effectively going idle after sending each response. By contrast, a *full-duplex process* (Trait 2) enables the AI to continuously think and listen, even when the user is not actively providing input. This opens the door to interaction models that transcend simple chat boxes. For instance, an AI planning assistant could silently update its suggestions as it overhears new constraints in a virtual meeting, intervening only when necessary (Traits 2 and 3). More interestingly, AI and users can build a *shared cognitive space* (Trait 4), adding thoughts, annotations, and partial ideas to a collective interface. Rather than a linear log of exchanges, this collaborative canvas allows ideas to branch, merge, and evolve continuously—resembling a dynamic mind-map or sketchnote more than a turn-based chat.

2.2.4 Messy, Fragmented and Informal Interaction. Traditional AI systems demand users structure their inputs as complete, well-formed queries or commands. This formality creates friction: human thinking is inherently non-linear, often involving half-formed ideas and abrupt shifts in focus. AI with thought embraces this messiness: AI thinking can be messy, incomplete, and in-progress. In addition, instead of requiring users to provide polished, fully-formed queries or responses, users can also input partial ideas, rough thoughts, or even keywords, and the system can process and respond to these in a similarly fragmented manner. This approach mirrors how humans often think and communicate: by iterating on ideas, refining them over time, and bouncing off incomplete or tentative thoughts.

3 Example Projects

To further illustrate the practical applications of AI thought in interactive systems, I present two projects that embody its core principles. The first, Inner Thoughts, explores how AI can proactively engage in conversations by developing and evaluating its internal thoughts before participating. The second, ThinkaloudLM, investigates AI-generated thoughts as an interface component that allows users to observe and interact with in real-time.

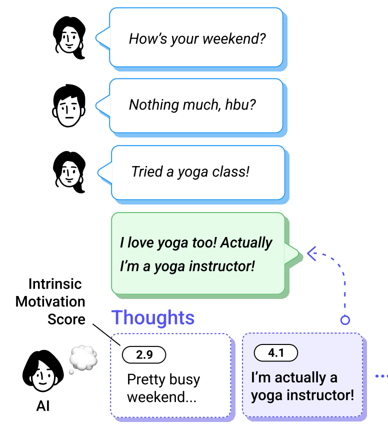

Figure 2: Conversational Agents with Inner Thoughts: AI generates a train of thoughts and evaluates them based on their intrinsic motivation to participate.

3.1 Inner Thoughts

Conversational AI often rely on turn-taking prediction techniques, where an algorithm determines the most likely next speaker and generates a response accordingly. However, these approaches struggle in multi-party conversations where turn allocation is often ambiguous. Furthermore, existing conversational agents tend to fall into two extremes: either they remain passive, requiring explicit user input to respond, or they overcompensate, generating frequent and often unnecessary interruptions.

Instead of predicting conversational turns, *Inner Thoughts* [7] introduces a new method in which AI autonomously generates a continuous stream of covert (internal) thoughts, similar to how humans process conversations. These thoughts remain internal until AI evaluates whether it has sufficient *intrinsic motivation* to contribute—defined by heuristics derived from a user study on human conversational behavior. Once the AI determines that a thought is relevant and meaningful, it then strategically selects an appropriate moment to engage (Figure 2).

The Inner Thoughts framework embodies multiple traits of thought: *Intermediate Medium*: AI does not simply generate final responses but produces an evolving series of intermediate thoughts; *Full-Duplex Process*: Instead of waiting for user prompts, the AI continuously listens and generates internal responses in parallel with the ongoing conversation; *Intrinsic Driver*: The AI does not just react but actively determines when to contribute based on its intrinsic motivation.

The framework was implemented in two systems, a multi-agent conversational simulation and a chatbot. Through a technical evaluation, conversational agents using Inner Thoughts significantly

Figure 3: ThinkaloudLM: AI generates intermediate, fragmented thoughts in parellel to user input.

outperformed a traditional next-speaker prediction baseline across multiple criteria, including turn appropriateness, coherence, engagement, and adaptability. Participants overwhelmingly preferred AI interactions using Inner Thoughts (in 82% of the conversations).

3.2 ThinkaloudLM

While Inner Thoughts explores AI's ability to think internally, ThinkaloudLM investigates a complementary question: what if users could actively observe and engage with AI's thought process?

Most AI systems present only finalized outputs, hiding the intermediate thoughts behind their decisions. This lack of transparency can lead to user distrust, inefficiencies in communication, and missed opportunities for collaboration. For example, if an AI-powered writing assistant provides a response that feels misaligned with the user's intent, the user must iteratively refine their query, leading to unnecessary back-and-forth exchanges.

By contrast, ThinkaloudLM envisions a system where users can engage with AI-generated ideas before they fully materialize into responses. Rather than receiving a single, static reply, users might see a live preview of AI's thinking, including key points, possible directions, or alternative solutions. They can then refine, select, or discard thoughts before finalizing the AI's output.

This approach leverages multiple traits of thought: *Intermediate Medium*: The AI's partially formed ideas are themselves the primary interface, giving users a real-time window into the AI's thought process; *Full-duplex Process*: By continuously updating these intermediate thoughts, the AI can incorporate user feedback on-the-fly while still in the "draft" stage; *Shared cognitive space*: Users and AI co-construct the final answer, using the AI's visible thoughts as a collaborative canvas for discussion, refinement, and mutual alignment. To explore this concept, a participatory design study is currently underway with qualitative interviews and observational studies.

4 Discussion and Future Work

How Can We Mitigate Potential Downsides? An AI that continuously generates intermediate thoughts risks overwhelming users with too much information or frequent interruptions. Further, sharing internal reasoning publicly could raise privacy and security concerns. Research is needed to identify interface designs, privacy safeguards, and policies that mitigate these concerns while preserving the benefits of proactivity and collaboration.

What Are the Best Mechanisms for Simulating Thought? Current LLM-based techniques show promise, but other approaches such as symbolic reasoning or neuron-level modeling offer different trade-offs in efficiency, interpretability, and control. Future research might explore how different architectures *represent* and *execute* continuous thought.

How Might AI Thinking Influence Human Thinking? If AI continuously generates and shares evolving thoughts, how might this reshape human cognitive processes? Will users offload more cognitive tasks to AI? Alternatively, could exposure to AI's thoughts improve human metacognition by encouraging users to articulate and refine their own thought processes? Additionally, how might this affect creative workflows, decision-making, and critical thinking skills?

What New Interaction Paradigms Can Emerge? Finally, AI with thought has the potential to reshape how users and AI collaborate, moving beyond simple chat interfaces. Exploring these new paradigms will be crucial to realizing AI that *thinks with us*, rather than simply *responding to us*.

5 Conclusion

In this paper, I call for the conceptualization of "thought" as new substrate in human-AI interaction process. We analyze the key traits of thought and how it can enable more proactive, adaptive, and collaborative engagement between humans and AI through two example projects.

In addition, I believe it represents a broader shift toward more *fluid, intuitive, and informal* paradigms of human–AI interaction. At its core, this approach challenges the decades-old model of AI as a passive respondent confined to chat interfaces—a structure that has persisted since the era of ELIZA [14]. I advocate for a fundamental reimagining of AI systems and human-AI interaction, ones that enable richer collaboration, fosters deeper trust, and ultimately creates a more natural way for humans and machines to think together.

References

[1] Herbert H. Clark and Susan E. Brennan. 1991. Grounding in Communication. In *Perspectives on Socially Shared Cognition*, Lauren B. Resnick, John M. Levine, and Stephanie D. Teasley (Eds.). American Psychological Association, 127–149.
[2] René Descartes. 1996. *Discourse on the method: And, meditations on first philosophy*. Yale University Press.
[3] Google DeepMind. 2025. Gemini Flash Thinking. https://deepmind.google/technologies/gemini/flash-thinking/ Accessed: 2025-02-13.
[4] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948* (2025).
[5] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 159–166. doi:10.1145/302979.303030
[6] Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, et al. 2024. Openai o1 system card. *arXiv preprint arXiv:2412.16720* (2024).
[7] Xingyu Bruce Liu, Shitao Fang, Weiyan Shi, Chien-Sheng Wu, Takeo Igarashi, and Xiang Anthony Chen. 2025. Proactive Conversational Agents with Inner Thoughts. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3706598.3713760
[8] Earl K Miller and Jonathan D Cohen. 2001. An integrative theory of prefrontal cortex function. *Annual review of neuroscience* 24, 1 (2001), 167–202.
[9] Allen Newell. 1972. Human problem solving. *Upper Saddle River/Prentive Hall* (1972).

[10] Donald A. Norman. 2013. *The Design of Everyday Things* (revised and expanded ed.). Basic Books.

[11] Marcus E Raichle, Ann Mary MacLeod, Abraham Z Snyder, William J Powers, Debra A Gusnard, and Gordon L Shulman. 2001. A default mode of brain function. *Proceedings of the national academy of sciences* 98, 2 (2001), 676–682.

[12] Alan M. Turing. 1950. Computing Machinery and Intelligence. *Mind* LIX (1950), 433–460. https://api.semanticscholar.org/CorpusID:14636783

[13] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35 (2022), 24824–24837.

[14] Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9 (1966), 36 – 45. https://api.semanticscholar.org/CorpusID:1896290