Guided Sensemaking: Agents in Collaborative Deliberation

Aaditya Bhatia, United States Military Academy
Navdeep Kaur Bhatia, Independent Researcher
Jack Park, TopicQuests

Main themes: design strategy + interaction paradigm

Target domains: education, collective intelligence

Cognitive 'target(s)': critical thinking, metacognition, collective sensemaking

Type of contribution & main idea

Conceptual design and interaction paradigm for AI-augmented collective sensemaking; we propose a multi-agent system that fosters critical thinking and collaborative discourse in a classroom setting.

The AI tool's key characteristics

- Socratic critique as cognitive partnership
- Structural sensemaking via discourse graphs
- Human-centered, non-authoritative AI agents
- Reflective design - encourage the user to slow down and engage in deliberate analytical thinking

The following AI agents augment user thinking-

1. Socratic Guide acts as a cognitive partner by privately asking probing questions; it challenges assumptions, asks for concrete evidence to back claims, and demands clarification by surfacing counter arguments.
2. Reflector extracts argumentative structure and develops a visualization; it enables reasoning about complex discussions without relying heavily on working memory.
3. Curator acts as a sensemaking facilitator for the group; it organizes user contributions as nodes in a discourse graph, which enables users to conceptualize the problem space around a given prompt.

The tool is designed to support thinking by helping users articulate, connect, and reflect on ideas. AI agents are not judges of correctness; they enable reflective thinking instead of providing answers.

		Interaction design/usage strategy	Intermediary task outcomes	Final task outcomes
Interaction & outcomes	How people use and interact with the AI tool for a specific activity (first column) and what outcomes this provides, enables, or leads to (second and third column).	User answers an overarching question through an ideation panel for text composition and a dynamic visualization of their argument structure. Instead of a standard chat, the system uses AI agents to generate exploratory questions and visual discourse graphs. User interacts indirectly by revising their prose to address these prompts rather than replying to the agents; this design choice introduces a deliberate pause to encourage deeper reflection.	At the individual level, the agents generate a personal discourse graph that supports reflection-in-action by helping users track how their reasoning evolves. This visualization fosters clearer arguments as users constructively respond to critique.	Final outcome would be a collaborative discourse graph with reduced redundancy, clear argumentative structure, and well-justified contributions, allowing the users and facilitators to conceptualize the entire problem space. Users gain a shared understanding of how ideas relate and evolve. Each user contributes to this graph in a meaningful, measurable way.
Objective value	The 'objective' quality of the process and the outcomes – i.e. what makes them 'good' or desirable (e.g. critical thinking, understanding, new insights, learning).	Users conduct critical thinking using an AI agent-supported scaffolding. Their thoughts and ideas undergo refinement before being shared with the group, improving the signal-to-noise ratio. Watching a well-curated collaborative discourse graph grow and evolve in real-time helps the group converge on ideas and insights that meaningfully address the overarching discussion prompt.	Personal discourse graph help users visualize reasoning, assumptions, and structural gaps. High-quality intermediary outcomes include clearer claims, identifiable assumptions, and better evidence contributed to the evolving discourse. Assessment criteria includes user engagement, evolution of thought measured through revision frequency and self-reported metacognitive gains.	The final collaborative discourse graph transforms a set of user posts into an evolving cognitive artifact that externalizes collective reasoning. The final discourse graph is highly structured but provides links to original user posts, preserving user agency. The final outcome can be measured by completeness of the problem space and objective measures including the number of positions per issue and arguments per position.
Perceived value & UX	How the process of using the AI tool and its outcomes is experienced by users (e.g. how they experience the cognitive work they are doing).	The tool encourages reflections through guided prompts, which users should find collaborative due to a low-pressure process. The feedback from Socratic Guide leads to a productive pause that helps users refine ideas privately before sharing them publicly. This can be assessed by behavioral metrics such as revision frequency, prompt engagement, contribution rates, and qualitative reflections on how the AI shaped users' willingness to think and contribute.	Users may feel increased confidence on their thought process through a visual representation of the evolution of their ideas. Users may not feel corrected or judged due to the partnership aspect of the AI agent. However, some users may find the saved evolution of their personal discourse graph intrusive and uncomfortable.	People might experience increased clarity about their own thinking, an improved ability to trace how their positions changed, and greater confidence when responding to critique. The AI tool is likely perceived as a structured mirror or facilitator. Reviewing the final artifact is likely to lead the users to recognize its depth and balance; this is when they might recognize how their own thinking evolved during the interaction.

How do people's goals 'interact with' the AI tool's goals? Is there a tension?

Users are likely to approach the system with short-term goals of making contributions quickly while minimizing cognitive effort. Users are accustomed to chat interfaces where AI would generate text, lookup information, and summarize ideas. Their long-term goals could be to improve critical thinking, prioritize quality over speed, and learn slowly rather than delivering results quickly.

The AI tool's goals are to promote reflection, increase structural coherence, and enable collective sensemaking, which largely align with users' long term aims, but is likely to conflict with user's desire for speed or delegation.

These tensions are intentional as the system introduces productive friction that encourages users to slow down and engage analytically and deeply. By leveraging this friction, the design fosters a more learning-oriented discourse.

How do you expect people to continue using this AI tool?

People would continue to use the tool based on the value they find in reflective sensemaking over speed. In time-pressured environments, user's short term goal of accomplishing the task quickly could cause friction, especially for those with limited experience with the platform. To support sustained use, the system will allow for adaptive levels of guidance, enabling users to soften introspective dialogue when appropriate while preserving the reflection process.

What you would like to discuss

How much productive support or unwanted friction is caused by AI-mediated reflective pauses? How to ensure that AI agents provide a thinking scaffold without becoming content generators? How might users feel about using an interface where AI agents are not instantly helpful.

What would you like to take away from the workshop?

Protecting human cognition is a growing concern. We hope to deeply understand where the challenge lies and what approaches are effective in fostering critical thinking. We want to gain perspective from related projects, identify potential collaborators, and get concrete feedback on our approach to prepare for an empirical evaluation of the system.

Key references (e.g. of main theories, empirical evidence, measurement methods etc.)

[1] Lucas Anastasiou and Anna De Liddo. 2026. From Shallow Discourse to Structured Deliberation: The BCause Approach. In Human-Computer Interaction – INTERACT 2025, Carmelo Ardito, Simone Diniz Junqueira Barbosa, and Raquel Prates (Eds.). Springer Nature Switzerland, Cham, 94–99. doi:10.1007/978-3-032-05008-3_19.

[2] Aaditya Bhatia and Gita Sukthankar. 2025. Using LLMs to Structure and Visualize Policy Discourse. In Artificial Intelligence for Research and Democracy, Wenpeng Yin, Jihyun Janice Ahn, Rui Zhang, Lifu Huang, Rafik Hadfi, Takayuki Ito, Susumu Ohnuma, and Shun Shiramatsu (Eds.). Vol. 14917. Springer Nature Singapore, Singapore, 69–76. doi:10.1007/978-981-97-9536-9_5.

[3] Anjali Singh, Karan Taneja, Zhitong Guan, and Avijit Ghosh. 2025. Protecting Human Cognition in the Age of AI. doi:10.48550/arXiv.2502.12447.

How to proceed with this work/idea?

We intend to develop a lightweight prototype that explores different levels of AI guidance and reflective prompting. Insights from the workshop will inform how the interaction is operationalized and how the system might be evaluated in classroom or organizational settings, guiding future empirical studies.