

Socratic Method Revisited: Human-AI Dialogue for Knowledge Creation and Internalization

Navid Hashemi Tonekaboni
 College of Charleston
hashemin@cofc.edu

Saber Soleymani
 Georgia State University
ssoleymani@gsu.edu

Abstract

This paper presents the Revisited Socratic Knowledge Interaction Framework, a novel approach to human-AI collaboration in knowledge-intensive tasks. Leveraging the Socratic method, the framework systematically integrates classical Socratic elements—Elenchus (critical refutation), Maieutics (knowledge elicitation), Aporia (constructive doubt), and Dialectic (collaborative synthesis)—into human-LLM interactions. We delineate how AI augments select Socratic components, while humans retain final authority over context and ethics. Phase-guided dialogue turns confident-but-wrong outputs into testable claims that must be justified, challenged, and revised. The work describes an agentic AI model that orchestrates role-specific agents to collaboratively facilitate Socratic dialogue with users. Our full open-source implementation is publicly available on GitHub¹ along with a live web application. Through practical scenarios in academic research and entrepreneurial planning, we demonstrate the framework’s utility in fostering deeper inquiry, critical thinking, and knowledge co-creation by challenging assumptions and promoting iterative refinement. This approach reinforces an augmented intelligence paradigm by operationalizing human-in-the-loop oversight via phase-routed dialogue, explicit check-backs, and an auditable trail of reasons, leading to more trustworthy and superior co-created outcomes.

Keywords: Human-AI Collaboration, Socratic Method, Knowledge Management, Agentic AI, Large Language Models (LLMs), Augmented Intelligence

¹https://github.com/noghte/socratic_chatbot

1. Introduction

Despite their impressive capabilities, current large language models (LLMs) are known to generate information that appears plausible but is factually incorrect (hallucinations), and they fundamentally lack a true understanding of truth, ethics, or nuanced contextual knowledge. This inherent limitation necessitates continuous human critical oversight to verify facts, logical coherence, and ethical implications of AI-generated content. The core challenge lies in designing systems that transcend mere information provision, instead actively asking, challenging, and inspiring deeper inquiry and knowledge co-creation. This critical gap between LLMs’ generative capacity and the human imperative for truth, ethics, and critical judgment is precisely where the Socratic method offers a structured solution (Lara and Deckers, 2020).

LLMs excel at generating content, but without genuine comprehension of truth or ethics and with susceptibility to hallucinations, human contributions of ethical judgment and contextual knowledge become indispensable. This problem statement underscores the necessity of augmented intelligence—where AI amplifies human capabilities—over pure artificial intelligence (French and Shim, 2024). It also highlights the critical role of responsible AI by embedding ethical considerations and human oversight directly into the interaction design, ensuring that model contributions are continuously vetted and aligned with human values.

Within this problem space, the Socratic method (Seeskin, 1987), a time-tested form of cooperative argumentative dialogue centered on asking and answering questions to stimulate critical thinking, emerges as a natural and promising candidate for structuring meaningful human-LLM conversations. By routing dialogue through elenchus (refutation

through cross-examination), aporia (productive impasse), maieutics (guided elicitation), and dialectic (collaborative synthesis), the interaction can scaffold explicit reasoning, surface contradictions, and embed human approval gates that keep ethics and context under human control (Delić and Bećirović, 2016; Ho et al., 2023). In what follows, we use this classical method as an interaction pattern to ground human–LLM collaboration in knowledge-intensive work.

We use three operational notions in this paper: *intellectual process*, the stepwise cycle of articulating claims, testing reasons and counter-examples, revising hypotheses, and synthesizing conclusions (Polit and Beck, 2004); *intellectual scaffolding*, the prompts and counter-prompts that structure this cycle into explicit, auditable steps open to critique and revision (Wood et al., 1976); and *productive discomfort*, the recognition of a gap or contradiction that motivates revision rather than disengagement (Mercer and McDonagh, 2021). With these notions in place, we next define the four components that ground our Socratic AI framework:

- **Elenchus (Refutation & Critical Questioning):** This is the central technique of the Socratic method, involving rigorous cross-examination to test the logical consistency of beliefs and reveal inconsistencies. It is described as an ‘adversarial approach’ that can be ‘destructive’ in dismantling false beliefs (Zaks, 2018).
- **Maieutics (Guiding to Birth of Knowledge):** Often referred to as ‘midwifery’, this aspect highlights Socrates’ role in assisting others to ‘give birth’ to their own understanding and insights, rather than directly imparting knowledge (Hanke, 1990).
- **Aporia (Perplexity or Impasse):** A state of productive perplexity or impasse, where the realization of inadequate knowledge creates an openness to new insight and deeper understanding. It is a crucial turning point that precedes genuine inquiry (Politis, 2006).
- **Dialectic (Collaborative Dialogue Toward Synthesis):** The overarching collaborative process through which two or more parties engage in reasoned discourse, posing questions and answers to collectively arrive at a more refined understanding or truth (Meyer, 2018).

The Socratic method is characterized as a cooperative argumentative dialogue and a shared dialogue. Simultaneously, elenchus is explicitly defined as refutation through cross-examination and an adversarial

approach that is destructive. This apparent contradiction highlights a core dynamic: the destruction of false beliefs through refutation is a necessary precursor to the birth of new understanding (maieutics) and the collaborative pursuit of truth (dialectic). The discomfort of aporia is productive, creating a fertile ground for new knowledge to emerge. When designing Socratic AI, this tension must be carefully managed. A system that is perceived as purely adversarial or destructive without a clear path to constructive insight could lead to user frustration and disengagement. The aim is to cultivate productive discomfort, not intimidation, requiring a nuanced persona that balances challenge with guidance.

Recent advances in LLMs have ushered in an era of unprecedented possibilities for human–LLM collaboration in knowledge-intensive work. These models are now capable of engaging in interactive dialogues that closely resemble human conversation, prompting a critical re-evaluation of how classical methods of inquiry might be applied or adapted within these novel interactions (Liu and Shen, 2025). Knowledge innovation systems, traditionally reliant on human creativity, critical thinking, and collaborative discourse for generating, sharing, and applying new knowledge, are undergoing a profound transformation with the integration of LLMs as active partners.

This evolving landscape signifies a fundamental shift from viewing computational systems merely as tools for automation or information retrieval to conceptualizing them as genuine collaborative entities (Jarrahi et al., 2023). This redefinition moves beyond simply enhancing efficiency to fostering a more symbiotic relationship where LLMs actively participate in the intellectual process. This paradigm shift necessitates the development of interaction frameworks that transcend simple command-and-response prompting, paving the way for dynamic, co-creative paradigms. It also raises important questions about how human cognitive processes may be reshaped and augmented through continuous, deep engagement with LLMs, akin to the concept of intellectual scaffolding.

This paper pursues three primary aims:

1. To establish a theoretical link between classical Socratic dialogue elements (elenchus, maieutics, aporia, and dialectic) and modern human–LLM interactions.
2. To delineate which components of the Socratic method can be effectively augmented by LLMs and which fundamentally require human insight.
3. To propose methodologies and practical scenarios for deploying Socratic human–LLM dialogue within various knowledge innovation settings.

The central contribution of this work lies in aligning ancient pedagogical techniques with contemporary models to enhance human–LLM collaborative outcomes, thereby offering a theoretically grounded yet practically actionable framework for knowledge co-creation. This contribution is expressed through providing design principles for agents that actively question, challenge, and inspire in the spirit of Socratic inquiry, while ensuring that ethical oversight and human contextual judgment remain central to the process. Furthermore, we describe the design and implementation of an agentic model that orchestrates a set of specialized agents to collaboratively facilitate Socratic dialogue with the user. For implementation, we employed CrewAI, a leading open-source multi-agent platform, to coordinate role-specific agents in structured Socratic interactions, demonstrating the feasibility of operationalizing Socratic principles into reproducible system design. The full implementation is publicly available, supporting transparency, replicability, and future extensions by the research community.

2. Background and Related Work

2.1. Challenges in Human–AI Collaboration

Despite their capabilities, LLMs are known to hallucinate (Brasse et al., 2023), lack a true understanding of truth or ethics, and exhibit inconsistent results in moral reasoning tasks. Consequently, human experts remain essential for verifying facts, logic, and ethical implications of model contributions. Developing ethical AI necessitates adherence to principles of fairness, transparency, accountability, and robust human oversight. The non-negotiable imperative for human-in-the-loop ethical governance and fact-checking in AI-powered knowledge systems is clear. The repeated emphasis that human experts must verify facts, logic, and ethical implications and that ethical judgment remains a human prerogative underscores that LLMs, despite their sophistication, cannot operate autonomously in domains requiring truth, values, or moral reasoning. This establishes a clear causal need for human oversight to prevent bias, discrimination, and misinformation. The Socratic method, by its inherent nature of questioning assumptions and demanding justification, provides a practical and structured mechanism for this oversight. It compels the system to expose its reasoning (or lack thereof), allowing humans to apply contextual, ethical, and value-based judgments, thereby making the collaborative process safer and more trustworthy (Lara and Deckers, 2020).

2.2. Evidence and Applications of Socratic Dialogue

The Socratic method possesses a rich history in education as a means to cultivate critical thinking by systematically exposing contradictions and guiding interlocutors toward deeper truths (Seeskin, 1987). Modern educational research consistently demonstrates that Socratic questioning significantly enhances critical thinking skills, improving learners’ clarity, logical reasoning, and metacognitive monitoring abilities (Lee et al., 2014). It fosters deeper learning and self-discovery, moving beyond traditional models of rote memorization or passive knowledge reception. Its effectiveness lies in compelling active engagement and self-reflection, shifting learners from passive recipients of knowledge to active participants. By forcing learners to articulate reasoning and confront contradictions, the method emphasizes the process of inquiry and cognitive work—rather than simply acquiring correct answers—as the foundation for genuine learning and the development of critical thinking.

Beyond education, dialogue and questioning play a crucial role in knowledge innovation systems, particularly in converting tacit knowledge (individual insights) into explicit knowledge (shared innovations), as articulated by Nonaka’s SECI model (Farnese et al., 2019). Contemporary management thought increasingly advocates for leaders to employ Socratic-style questioning to cultivate a culture of inquiry and improvement, stimulating innovative thinking and more effective collective problem-solving (Shoop, 2014).

This approach is equally relevant in entrepreneurial planning, where probing questions can uncover blind spots and novel solutions. Socratic dialogue thus functions as a systematic mechanism for articulating organizational knowledge and fostering innovation. Nonaka’s model emphasizes social interaction and reflection as central to this conversion, and Socratic questioning, by compelling individuals to articulate underlying assumptions (e.g., Why do you believe customers will prefer this feature?), directly facilitates the externalization and combination phases of knowledge creation. By favoring open-ended questions over directives, leaders create a clear causal chain: Socratic questioning surfaces tacit knowledge, which strengthens organizational learning and innovation. This extends the utility of Socratic AI beyond individual cognition to collective knowledge creation (Cress and Kimmerle, 2018), positioning it as a valuable tool for strategic planning, post-project analysis, and fostering environments that systematically draw out and leverage collective intelligence.

2.3. From Early Systems to LLM-enabled Orchestration

Early attempts to integrate Socratic methods into computing emerged within the domain of intelligent tutoring systems (Anderson et al., 1985) and educational chatbots (Georgescu, 2018). These systems aimed to guide learners through questions rather than providing direct answers, thereby improving learning outcomes by prompting articulation of reasoning. However, these early implementations were characterized by significant complexity, high development time, and brittle interactions limited to narrowly predefined domains (Córdova-Esparza, 2025). The historical limitations of earlier models in replicating Socratic dialogue were primarily due to insufficient generative and contextual understanding. They lacked the flexibility and natural language generation needed to sustain open-ended, contextually rich Socratic dialogues. The constraint was not just in question generation, but in understanding and adapting to a wide range of human responses.

The advent of contemporary LLMs has fundamentally altered this landscape, enabling new research explicitly combining Socratic strategies with open-ended dialogue. The principles of Socratic dialogue also inspire the design of multi-agent systems (Zhang et al., 2025). Socratic principles serve as a blueprint for enhancing internal reasoning and collective intelligence within multi-agent settings. The extension of Socratic principles to interactions between agents signifies a broader recognition: the benefits of structured inquiry are not limited to human–LLM interfaces but extend to internal computational processes. By assigning distinct roles and encouraging critical evaluation of each other’s reasoning, these systems aim to overcome individual limitations (e.g., mistaken beliefs) through a collective, dialectical process. This points toward a future where complex problems might be addressed by sophisticated Socratic agent teams, potentially leading to more robust, accurate, and less error-prone outputs. It also suggests a new avenue for research: designing systems that can self-critique and engage in internal thought processes that mimic structured Socratic inquiry.

2.4. Literature Synthesis and Research Gap

LLMs amplify human capability but remain prone to opaque reasoning, over-confidence, and ethical blind spots. Prior work shows that disciplined questioning improves reasoning and externalization at individual and organizational levels, but most systems do not operationalize a full Socratic cycle with explicit human approval gates. We address this gap with an agentic framework that routes turns across elenchus, aporia,

maieutics, and dialectic, logging reasons and requiring human confirmation before adoption.

3. The Revisited Socratic Knowledge Interaction Framework

This section introduces the Revisited Socratic Knowledge Interaction Framework, which systematically conceptualizes how classical Socratic dialogue elements can be mapped onto specific roles and interactions within human–AI collaborations in knowledge innovation systems. The framework identifies four primary components of the Socratic method—Elenchus, Aporia, Dialectic, and Maieutics—and provides a detailed description of how each is operationalized when an AI system is one of the participants. To clarify how these four components are distributed between AI augmentation and human judgment, Table 1 provides a mapping of AI’s potential contributions alongside the human insights required for effective oversight.

The Socratic method unfolds as a dynamic, iterative process in which dialogue turns may flexibly transition across the agents. Rather than being locked into a fixed order of steps, the interaction is designed to evolve in response to the user’s reasoning and contributions. As illustrated in Figure 1, the framework is implemented through a set of specialized agents coordinated by a Triage Agent, which dynamically routes dialogue turns based on user responses and context. The Triage Agent evaluates each user response and determines the most suitable next phase, ensuring that the dialogue remains coherent, purposeful, and aligned with the unfolding context. Elenchus typically initiates the process by probing key assumptions, after which the dialogue can loop through questioning, advance directly to synthesis, or conclude with a summary. This adaptive flow highlights not only the structured nature of the framework but also its ability to remain responsive and flexible in real-time interaction.

Beyond its procedural design, the framework also reflects the deeper philosophical character of the Socratic method. The four primary components do not operate in isolation but work synergistically, transforming simple question–and–answer exchanges into a richer process of knowledge interaction. By shifting the locus of knowledge from external authority to dialogic exchange, the framework aligns with the Socratic view that genuine learning and the co-creation of knowledge emerge through critical inquiry, co-construction, and reflection. In this sense, it reinforces the role of dialogue as both an intellectual and ethical engagement, where examining ideas collaboratively serves as a philosophical imperative and a practical outcome.

Socratic Component	AI Augmentation Potential	Human Insight Needed
Elenchus (critical questioning)	AI can generate probing questions and identify logical inconsistencies automatically.	Human verifies relevance of questions, judges if identified inconsistencies truly matter or are errors. Human also questions AI's statements (critical evaluation).
Maieutics (guiding discovery)	AI can provide hints, suggestions, and incremental questions to scaffold thinking.	Human integrates hints, chooses direction, and provides creative leaps. Human ensures guidance is on-track, bringing in context.
Aporia (constructive impasse)	AI can introduce counterexamples or conflicting information to provoke uncertainty. It can signal when it's unsure or when evidence conflicts.	Human experiences the puzzlement and decides to investigate further or rethink. Human distinguishes genuine impasse from AI mistake, and uses creativity to resolve the impasse.
Dialectic (synthesis through dialogue)	AI maintains dialogue flow, offers summaries and multiple perspectives, and attempts synthesis proposals.	Human sets goals, keeps conversation practical, and makes final judgments on synthesis (considering ethics, feasibility, values). Human breaks deadlocks or ends dialogue with a decision.

Table 1. Socratic Component Mapping: AI Augmentation Potential vs. Human Insight Needed

4. Socratic Human–AI Knowledge Interaction in Action

Step 1: Initial Proposition/Claim (Hypothesis Elicitation)

The dialogue begins with the interlocutor (human user) presenting a belief, claim, or definition on a specific topic. The Socratic AI agent may initiate by professing ignorance or posing a fundamental question to encourage the interlocutor to articulate their initial understanding.

Types of Questions: Clarification Questions (What do you mean by [term/concept]? Could you rephrase your point? Why do you say that?) and Definition Elicitation (What is [concept]? How would you define [term]?).

AI's Role: The AI initiates the conversation by prompting the user to articulate their initial thoughts, ensuring a clear starting point for the inquiry.

Human's Role: The human articulates their initial belief or hypothesis, providing the subject for the subsequent Socratic inquiry.

This step initiates the back-and-forth conversational exchange, establishing the foundation for mutual understanding (Dialectic). The AI's initial questions aim to elicit the interlocutor's existing ideas and knowledge, even if unformed or implicit, acting as a midwife to bring forth initial thoughts (*Maieutics*).

Step 2: Socratic Questioning & Cross-Examination (Elenchus in Action)

The Socratic AI agent rigorously probes the initial proposition through a persistent and targeted line of

questioning. This phase involves deep examination of the interlocutor's reasoning, assumptions, and evidence, with the goal of testing the validity and consistency of their belief. This is the primary application of Elenchus.

Types of Questions: Probing Assumptions (What is being assumed here? Is that assumption sound? Could the assumption be different?), Probing Reasons and Evidence (How do you know that? What is the quality of the evidence supporting your claim? Are there any reasons to doubt that?), Viewpoints or Perspectives (How might other people approach this issue? What is an alternative view? Have all the alternatives been considered?), Implications and Consequences (What is the result if this is true? What does that imply? If [X] is true, how does it affect [Y]?), and continued Clarification.

AI's Role: The AI generates a series of probing questions to systematically challenge the user's statements, assumptions, and evidence. This leverages the AI's ability to process information and identify potential areas for deeper inquiry.

Human's Role: The human responds to the probing questions, articulating their reasoning, providing evidence, and considering alternative viewpoints. This requires active cognitive engagement and self-reflection.

This step is the core application of Elenchus, actively challenging assumptions, testing beliefs, and identifying potential contradictions. The continuous back-and-forth exchange, where the interlocutor responds to probing questions, and the Socratic agent further examines those responses, embodies Dialectic. The AI guides the interlocutor to articulate their understanding and identify gaps or inconsistencies in their own thinking, without providing direct answers (*Maieutics*). The Triage Agent evaluates the user's answer and, if it detects a stall or

contradiction, transitions to Step 3 (Elenchus again), or otherwise to Step 4 (Aporia).

Step 3: Examination of Contradictions & Refutation (Deepening Elenchus)

Through the persistent questioning in Step 2, the Socratic AI agent reveals inconsistencies, logical flaws, or weaknesses in the interlocutor's initial claim or subsequent responses. This often involves presenting *counter-examples* or highlighting competing definitions, leading the interlocutor to recognize that their initial belief cannot withstand scrutiny.

Types of Questions: Questions that highlight logical inconsistencies (If [X] is true, but [Y] is also true, how can both be consistent? Does this specific example fit your definition, or does it contradict it?) and Questions prompting re-evaluation (Does this new information change your initial statement?).

AI's Role: The AI identifies and presents contradictions or logical flaws based on the user's previous statements or external knowledge (if RAG is integrated), forcing the user to confront inconsistencies.

Human's Role: The human acknowledges and admits contradictions in their arguments, which is crucial for the dialogue to proceed constructively. This requires intellectual honesty and a willingness to revise one's beliefs.

This is the direct outcome of the cross-examination, leading to the refutation of the initial argument and the dismantling of false or inconsistent beliefs (Elenchus). It requires the interlocutor to acknowledge and admit contradictions in their arguments for the dialogue to proceed constructively (Dialectic). The Triage Agent next invokes the Aporia Agent (Step 4).

Step 4: Induction of Aporia (Perplexity)

As a result of the refutation (Elenchus), the interlocutor reaches a state of perplexity, confusion, or intellectual impasse. They realize their initial knowledge or understanding was flawed, incomplete, or contradictory. This conviction of ignorance is a critical turning point, as it is considered a necessary first step towards genuine knowledge acquisition.

Types of Questions: Often implicit, or reflective (Are you still confident in your initial statement, given what we've discussed? Do you now see a problem with your definition or reasoning? What are you feeling about your understanding of [concept] right now?).

AI's Role: The AI's persistent and effective application of elenchus leads the user to this state of productive puzzlement, often by presenting a clear, undeniable contradiction.

Human's Role: The human experiences the intellectual discomfort of realizing their ignorance, which serves as a powerful motivator for deeper inquiry and learning.

This is the direct and intended result of the successful application of elenchus, forcing the interlocutor to confront their own ignorance and the limitations of their knowledge (Aporia). The Socratic agent facilitates this realization, creating the productive discomfort necessary for intellectual growth and the motivation to seek deeper understanding (*Maieutics*). The Triage Agent then invokes the Maieutics Agent (Step 5).

Step 5: Revision of Hypothesis / Search for New Understanding (Maieutics & Dialectic)

From the state of aporia, the interlocutor is prompted to revise their initial hypothesis, seek new definitions, or explore alternative viewpoints. The Socratic AI agent guides this process, helping the student give birth to new, more refined ideas and understanding. This step often involves a loop back to further questioning and examination.

Types of Questions: Redefinition/Re-evaluation (Given what we've discussed, how would you now define [concept]? How might you revise your initial claim?), Seeking Alternatives (What new understanding has emerged for you? What alternatives have we not yet considered?), and Questions About the Question (Is answering this question useful? Do we properly understand the question? How could we resolve this question?).

AI's Role: The AI provides scaffolding and prompts that encourage the user to reformulate their ideas, explore new avenues, and integrate new insights gained from the preceding steps.

Human's Role: The human actively revises their understanding, formulates new hypotheses, and engages in creative problem-solving to move beyond the impasse.

This is a core maieutic phase, where the Socratic agent acts as a midwife to understanding, helping students generate their own ethical questions and insights, and facilitating the birth of new ideas. The collaborative effort continues, refining ideas and moving towards deeper truths (Dialectic). This is where the iterative loop occurs, returning to further examination (Step 2) with the revised hypothesis. Finally, the Triage Agent invokes the Dialectic Agent (Step 6).

Step 6: Emergence of Deeper Understanding / Continued Inquiry

The iterative process of questioning, refutation, and revision leads to a more robust, nuanced, and critically

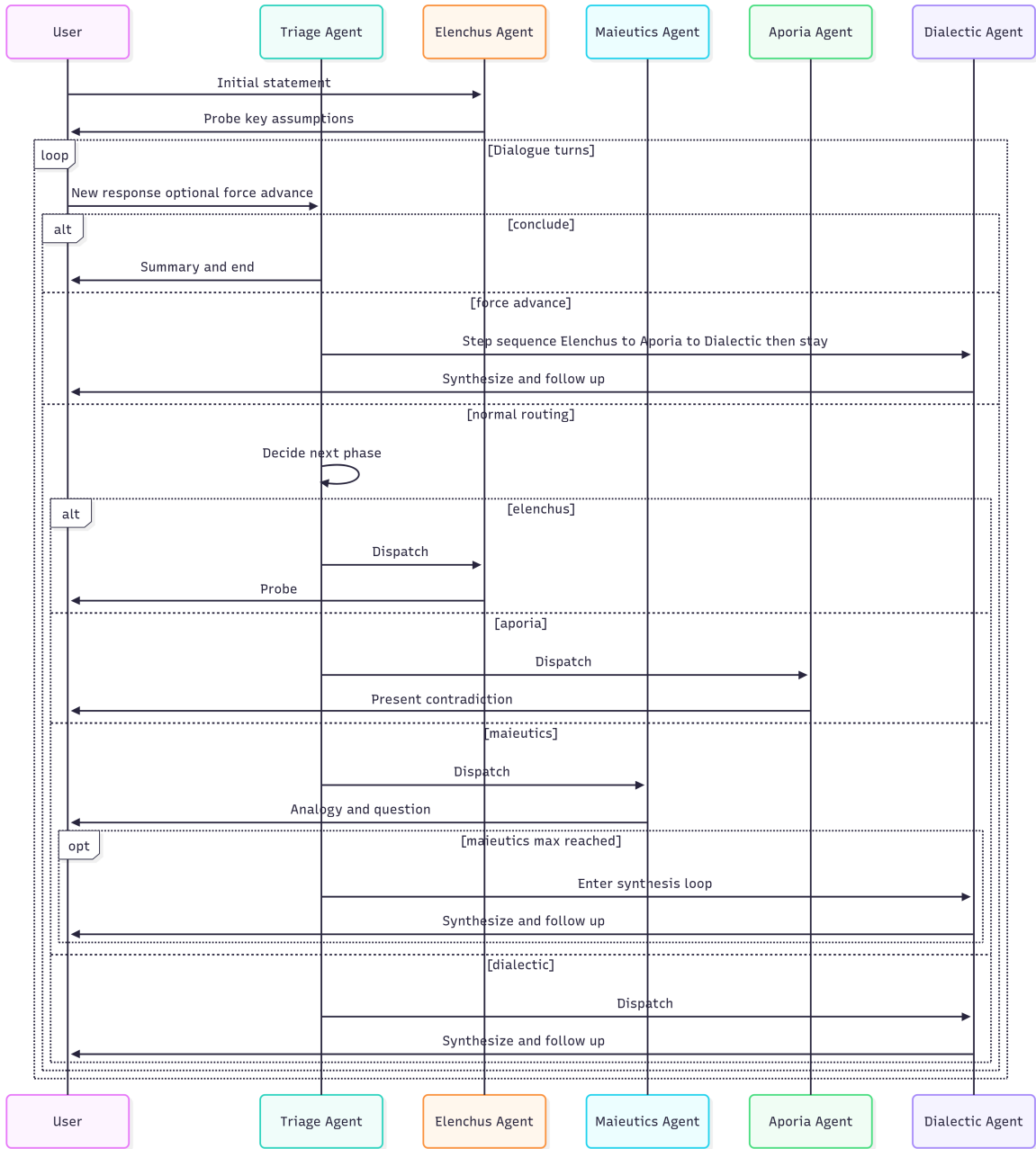


Figure 1. Sequence diagram of the Revisited Socratic Knowledge Interaction Framework. Dialogue always begins with Elenchus, and subsequent turns are dynamically routed by the Triage Agent. Depending on user responses and context, the dialogue may loop through Elenchus, Aporia, and Maieutics, advance directly to Dialectic for synthesis, or conclude with a summary.

examined understanding. While not always yielding a definitive answer, the goal is intellectual and ethical growth, equipping the learner for future critical thinking and problem-solving. The dialogue may conclude with a refined understanding or continue with new propositions for further inquiry.

Types of Questions: Reflective Questions (What have you learned from this discussion? How does this new understanding change your perspective or approach?) and Application Questions (How might you apply this understanding in a different situation?).

AI's Role: The AI facilitates reflection on the learning process and encourages the application of newly acquired knowledge, helping to solidify understanding and promote transferability.

Human's Role: The human synthesizes the insights gained, reflects on their intellectual growth, and considers how to apply the new understanding in practical contexts.

The ultimate goal of giving birth to the truth and fostering active, independent learning is realized through this refined understanding (*Maieutics*). The collaborative process successfully leads to mutual understanding and refinement of ideas, even if not a final, absolute truth (Dialectic). The Triage Agent then concludes the cycle and can either end the dialogue or loop back to Step 2 if the user wishes further inquiry.

5. Illustrative Scenarios and Case Studies

To illustrate the practical utility of the Revisited Socratic Knowledge Interaction Framework, we apply it in two knowledge-innovation contexts: academic research and entrepreneurial knowledge systems. These scenarios demonstrate the versatility of Socratic human-LLM dialogues across learning and innovation domains.

Scenario A: Academic Research Co-Pilot A graduate student questions the assumption that citation counts reliably indicate the importance and quality of academic papers. The student engages a Socratic AI assistant to probe this metric and explore more rigorous validation practices. The AI orchestrates a dialogue among specialized agents to challenge and refine the student's thinking. Here is the dialog of this case study:

- **User:** *One of the main metrics that shows an academic paper is important and high-quality is the number of citations.*
- **Elenchus Agent:** *I notice you assume that the number of citations is a valid indicator of the quality of an academic paper. But what if a paper*

is cited for criticism rather than praise? How would you respond to the possibility that important research may initially receive few citations but still be impactful?

- **User:** *If we exclude the papers cited for criticism, then a high citations shows an academic paper is important and high-quality.*
- **Aporia Agent:** *A widely cited paper may promote ideas that are ultimately misleading or incorrect. What do you think about this contradiction?*
- **User:** *In that case, many researchers would refrain from citing a flawed paper.*
- **Maieutics Agent:** *Think of citations like a compass that sometimes points the wrong way. How might that affect your judgment of widely cited papers?*
- **User:** *I will revise my statement: citation count is only one indicator. Papers should also appear in reputable venues and their findings should be reproducible and validated by others.*
- **Aporia Agent:** *But consider this: a flawed yet popular paper may continue to gain citations due to prevailing trends. What does that imply for research credibility?*
- **User:** *It implies we need to replicate key experiments and scrutinize methods and data quality before trusting citation metrics.*
- **Dialectic Agent:** *Given that replication and critical scrutiny build credibility, how would you encourage graduate students to adopt these practices?*
- **User:** *I would tell them to view citations as preliminary signals and to systematically reproduce results, critique methodologies, and seek consensus in the research community.*
- **Maieutics Agent:** *Your final answer seems reasonable. It describes how to validate and improve research credibility through replication and careful examination of methods and data quality. If you'd like more elaboration, feel free to let me know. I'd be happy to continue the conversation.*

Through this Socratic exchange the student gains a deeper understanding of research validation. The dialogue exposes the limits of citation metrics and promotes best practices—replication, methodological critique, and community consensus—that underpin truly high-quality scholarship.

Scenario B: To-Do List App Usability Co-Pilot A developer is designing a to-do list app and plans to use A/B testing to evaluate the usability of new features. The AI coordinates a series of agentic prompts to challenge assumptions and refine the testing strategy.

- **User:** *I am developing a to-do list app. With A/B testing I can test different features to measure the usability of new features.*
- **Elenchus Agent:** *You assume A/B testing will accurately reflect usability for all users. What if people respond differently based on their unique needs or preferences? How will you ensure your results represent the entire user base?*
- **User:** *I will quantitatively analyze users' interactions to find out.*
- **Maieutics Agent:** *Think of it like a chef tasting a dish. One spoonful may not capture the full flavor. What does this reveal about the depth of interactions you're examining?*
- **User:** *I'll test two different button colors for the purchase page. Based on tap counts I'll know which color is better.*
- **Dialectic Agent:** *You've moved from general A/B testing to a specific button-color experiment. The key insight is that preferences may require more nuance. How might you combine qualitative feedback with your quantitative data to gain deeper insights?*
- **User:** *I'll add a feedback form in the app to collect user comments.*
- **Aporia Agent:** *But asking for feedback may lead users to focus on complaints while ignoring features they like. What do you think about this contradiction?*
- **User:** *I could offer incentives for users who provide feedback.*
- **Aporia Agent:** *Offering incentives might motivate users to reply only for the reward, undermining the authenticity of their insights. What do you think about this contradiction?*
- **User:** *Perhaps it's better to randomly select users and invite them to give feedback.*

The developer refines the A/B testing plan to include representative sampling, integrate qualitative comments with tap metrics, and protect feedback authenticity. This leads to more reliable, actionable usability insights.

5. Discussion and Limitations

This paper introduces the Revisited Socratic Knowledge Interaction Framework and a proof-of-concept implementation that routes turns across Elenchus, Aporia, Maieutics, and Dialectic. The central thesis is that reviving the Socratic method provides a practical protocol for augmented intelligence. The interaction model shifts from user-command to co-inquisitive dialogue. The AI acts as a catalytic partner that provokes articulation, while the human remains the final arbiter of purpose, ethics, and context. This also safeguards against LLM fallibility, since justification and counter-questioning are built into the process.

The contributions are threefold. First, the framework maps classical Socratic elements onto agentic roles, showing how questioning, impasse, guidance, and synthesis can be operationalized. Second, it delineates complementarity: AI excels at breadth, recall, and disciplined probing, while humans contribute depth, values, and creative synthesis. Third, an open-source implementation provides a replicable artifact for extension and comparative study.

The framework carries implications for theory and practice. It specifies a class of catalytic or maieutic systems whose purpose is to improve human thinking rather than automate answers. For design, three guidelines follow: challenges should be transparent with rationales or evidence, user agency should be preserved with a mode shift to summarization, and dialogue graphs should be persisted for audit and reuse.

There are important limitations to acknowledge. The analysis draws on scenarios rather than controlled or longitudinal studies, and constructs such as internalization and aporia are difficult to measure with conventional methods. LLM fallibility and hallucinations remain a risk if questions are not grounded. The process also imposes time and cognitive costs, which may deter users in fast-paced settings. Finally, adoption in high-stakes or affective domains cannot be assumed, since tone and consent are crucial.

Future research should therefore test the framework more rigorously and in real-world settings. Controlled experiments are needed to compare its outcomes with strong baselines on assumption surfacing, problem quality, and delayed transfer. Field studies that embed Socratic agents into collaborative platforms can assess reuse, cycle time, and decision reversals.

Technical work should explore retrieval-augmented questioning, adaptive tone and consent, and role-diverse multi-agent extensions. Equally important are studies of socio-technical integration, including incentives and governance patterns, as well as fairness and safety to ensure respectful and psychologically safe questioning.

In sum, the framework treats Socratic dialogue as a design protocol for human-in-the-loop AI. AI provides scale in questioning and recall, while humans supply judgment, goals, and ethics. With safeguards for grounding, tone, and provenance, Socratic interaction can enhance transparency, trust, and the quality of co-created outcomes while respecting organizational and time constraints.

6. Acknowledgements

We would like to acknowledge the use of Google's Gemini 2.5 language model for its assistance in refining and enhancing the clarity of this paper.

References

- Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science*, 228(4698), 456–462.
- Brasse, J., Broder, H. R., Förster, M., Klier, M., & Sigler, I. (2023). Explainable artificial intelligence in information systems: A review of the status quo and future research directions. *Electronic Markets*, 33(1), 26.
- Córdova-Esparza, D.-M. (2025). Ai-powered educational agents: Opportunities, innovations, and ethical challenges. *Information*, 16(6), 469.
- Cress, U., & Kimmerle, J. (2018). Collective knowledge construction. In *International handbook of the learning sciences* (pp. 137–146). Routledge.
- Delić, H., & Bećirović, S. (2016). Socratic method as an approach to teaching. *European Researcher Series A*, (10), 511–517.
- Farnese, M. L., Barbieri, B., Chirumbolo, A., & Patriotta, G. (2019). Managing knowledge in organizations: A nonaka's seci model operationalization. *Frontiers in psychology*.
- French, A. M., & Shim, J. (2024). From artificial intelligence to augmented intelligence: A shift in perspective, application, and conceptualization of ai. *Information Systems Frontiers*, 1–22.
- Georgescu, A.-A. (2018). Chatbots for education—trends, benefits and challenges. *Conference proceedings of eLearning and Software for Education (eLSE)*, 14(02), 195–200.
- Hanke, M. (1990). Socratic pragmatics: Maieutic dialogues. *Journal of pragmatics*, 14(3), 459–465.
- Ho, Y.-R., Chen, B.-Y., & Li, C.-M. (2023). Thinking more wisely: Using the socratic method to develop critical thinking skills amongst healthcare students. *BMC medical education*, 23(1), 173.
- Jarrahi, M. H., Askay, D., Eshraghi, A., & Smith, P. (2023). Artificial intelligence and knowledge management: A partnership between human and ai. *Business Horizons*, 66(1), 87–99.
- Lara, F., & Deckers, J. (2020). Artificial intelligence as a socratic assistant for moral enhancement. *Neuroethics*, 13(3), 275–287.
- Lee, M., Kim, H., & Kim, M. (2014). The effects of socratic questioning on critical thinking in web-based collaborative learning. *Education as Change*, 18(2), 285–302.
- Liu, Y., & Shen, L. (2025). Consolidating human-ai collaboration research in organizations: A literature review. *Journal of Computer, Signal, and System Research*, 2(1), 131–151.
- Mercer, L. E., & McDonagh, D. (2021). The pedagogy of discomfort: Transformational experiential learning. *DISCERN: International Journal of Design for Social Change, Sustainable Innovation and Entrepreneurship*, 2(2), 22–35.
- Meyer, M. (2018). Dialectic and questioning: Socrates and plato. *Revue internationale de philosophie*, 284(2), 113–129.
- Polit, D. F., & Beck, C. T. (2004). *Nursing research: Principles and methods*. Lippincott Williams & Wilkins.
- Politis, V. (2006). Aporia and searching in the early plato. *Remembering Socrates: Philosophical Essays*, 88, 109.
- Seeskin, K. (1987). *Dialogue and discovery: A study in socratic method*. Suny Press.
- Shoop, B. L. (2014). Developing critical thinking, creativity and innovation skills. *2014 IEEE Innovations in Technology Conference*, 1–6.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, 17(2), 89–100.
- Zaks, N. (2018). Socratic elenchus in the sophist. *apeiron*, 51(4), 371–390.
- Zhang, J., Wang, Z., Zhu, H., Liu, J., Lin, Q., & Cambria, E. (2025). Mars: A multi-agent framework incorporating socratic guidance for automated prompt optimization. *arXiv preprint arXiv:2503.16874*.