AI as a socratic dialogue partner: A conceptual discussion on generative artifical intelligence as a cognitive scaffold for academic enquiry.

Philip Dennett & Syeda Binish Zahra
London South bank University – UAE
Ajman, United Arab Emirates
philip.dennett@lsbuac.ae

https://doi.org/10.63962/PIYB4717

Abstract— The rapid emergence of generative artificial intelligence (AI) tools has prompted significant debate regarding their role in academic knowledge production. While concerns about authorship, academic integrity, and epistemic reliability are widely discussed, less attention has been paid to how AI might function as a dialogic cognitive scaffold within scholarly inquiry. This conceptual/position paper proposes a framework in which generative AI, when used through a structured Socratic questioning process, operates not as an autonomous content generator but as a reflective dialogue partner that supports the development of academic thinking. Drawing on theories of Socratic pedagogy, Vygotskian scaffolding, and extended cognition, the paper argues that AI-mediated dialogue can potentially enhance conceptual exploration, reflexive reasoning, and intellectual synthesis when used transparently and critically. The paper introduces the concept of AI-supported Socratic inquiry, outlines a conceptual theoretical model of human–AI co-cognition, and discusses the ethical and epistemological conditions necessary for responsible use. The paper concludes by suggesting that generative AI should be understood not simply as a writing tool but as a potential cognitive infrastructure for scholarly dialogue.

Keywords—generative AI, academic integrity, socratic dialogue

I. Introduction

Generative artificial intelligence has rapidly entered the academic landscape, prompting both interest and concern. Universities and scholarly communities are currently debating whether AI tools undermine academic integrity, threaten authorship norms, or offer new forms of intellectual augmentation. Much of this discussion has focused on the risks associated with automated text generation, including plagiarism, hallucinated citations, and the potential erosion of critical thinking[1].

However, these debates often assume that generative AI functions primarily as a content production system. This assumption overlooks an alternative mode of engagement in which AI is used not to generate finished text but to support iterative intellectual dialogue. This conceptual paper proposes that generative AI can function as a Socratic dialogue partner in academic inquiry.

We argue that when used through structured questioning and critical engagement, AI can support processes of conceptual clarification, theoretical exploration, and argumentative development. The argument developed here is not that AI replaces scholarly reasoning, but that it can serve as a cognitive scaffold – an interactive system that supports human intellectual processes[2].

Three research questions guide the discussion:

1. How can generative AI function as a dialogic partner in academic inquiry?
2. What theoretical frameworks help explain human–AI collaborative cognition?
3. Under what ethical conditions can such practices be considered academically legitimate?

To address these questions, the paper draws on three theoretical traditions:

- Socratic pedagogy
- Vygotskian scaffolding
- Extended cognition theory

Together, these perspectives provide a framework for understanding AI-mediated inquiry as a form of distributed intellectual practice.

II. Socratic Dialogue as a Method of Knowledge Construction

The Socratic method represents one of the earliest formalised approaches to dialogic knowledge development. In Platonic dialogues, Socrates engages interlocutors through structured questioning designed to expose contradictions, clarify definitions, and stimulate deeper reflection[3].

The core features of Socratic inquiry include:

- iterative questioning
- critical examination of assumptions
- conceptual clarification
- dialogic reasoning

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

- intellectual humility

Socratic dialogue encourages participants to develop understanding through reflective interrogation, rather than transmitting knowledge directly. Modern educational theory has widely recognised the value of such dialogic processes[4]. Importantly, the Socratic method does not depend on the authority of the interlocutor. The dialogue partner does not need to possess superior knowledge; rather, the dialogue itself facilitates intellectual discovery.

This observation becomes significant when considering AI-mediated dialogue. If the value of Socratic exchange lies in the process of questioning rather than the authority of the responder, then generative AI may serve as a viable interlocutor within structured intellectual dialogue.

To test this contention the primary author conducted a Socratic Dialogue using ChatGPT (Appendix 1).

III. SCAFFOLDING AND COGNITIVE SUPPORT

The concept of scaffolding originates from Vygotsky's theory of the zone of proximal development (ZPD). Vygotsky proposed that learning occurs most effectively when individuals engage in tasks that are just beyond their current capability but achievable with guidance[5]. Educational scaffolding refers to temporary support structures that assist learners in performing complex cognitive tasks. These supports may include prompts, guiding questions, conceptual frameworks, or collaborative dialogue. Importantly, scaffolding does not replace the learner's intellectual effort. Instead, it supports the development of higher-order reasoning by structuring the cognitive environment.

Generative AI systems can potentially function as scaffolding mechanisms in academic inquiry by:

- prompting alternative perspectives
- helping clarify conceptual distinctions
- identifying gaps in reasoning
- suggesting theoretical connections

When used interactively, we argue that AI can therefore assist researchers in navigating complex intellectual terrain.

However, the scaffolding analogy also highlights an important limitation: scaffolding must be temporary and critically engaged with. Over-reliance risks reducing independent cognitive development.

IV. DISTRIBUTED THINKING

Clark and Chalmers[6] extended mind thesis argues that cognition is not confined to the brain but can extend into external artefacts that support thinking. Examples include notebooks, diagrams, calculators, and digital tools.

From this perspective, cognition becomes a distributed process involving interactions between individuals and their environments. Generative AI systems may represent a new form of cognitive extension. Unlike passive tools, they can provide dynamic responses that shape ongoing reasoning processes.

In AI-mediated Socratic dialogue, the cognitive process can become a human–machine interaction loop where:

1. Human poses question
2. AI generates response
3. Human evaluates response
4. Human reformulates question
5. Dialogue iterates

The intellectual labour remains human-driven, but the cognitive environment is expanded through interaction with the AI system. This model aligns with contemporary theories of human–AI co-cognition[7], in which AI systems function as thinking partners rather than autonomous agents. This process was tested empirically using ChatGPT as a dialogic partner (see Appendix 1).

V. AI-SUPPORTED SOCRATIC INQUIRY: A CONCEPTUAL MODEL

The interaction between Socratic dialogue, scaffolding, and extended cognition can be synthesised into what may be termed AI-supported Socratic inquiry (Figure 1).

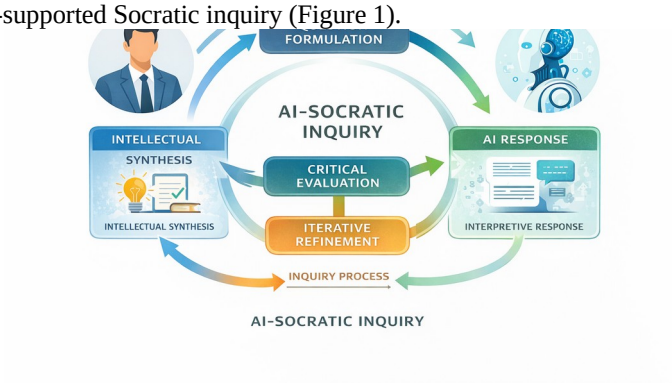


Figure 1: AI–Socratic Dialogue

In this model, generative AI operates as a reflective dialogue partner within a structured questioning process.

The interaction follows several stages:

Stage 1: Question Formulation

The researcher initiates inquiry by posing a conceptual question.

Stage 2: AI Response

The AI system produces an interpretive response based on its training data.

Stage 3: Critical Evaluation

The researcher critically assesses the response, identifying strengths, weaknesses, and gaps.

Stage 4: Iterative Refinement

The researcher reformulates questions to probe deeper conceptual issues.

Stage 5: Intellectual Synthesis

The researcher integrates insights into an independent scholarly argument.

Crucially, the AI system does not determine the outcome of the inquiry. The human researcher remains responsible for interpretation, verification, and synthesis.

VI. ETHICAL CONSIDERATIONS

The use of AI in academic inquiry raises important ethical concerns. These concerns centre on authorship, intellectual ownership, and epistemic reliability:

- Scholarly authors remain accountable for the arguments they present. AI outputs must be critically evaluated rather than accepted uncritically.
- Transparency regarding AI use is increasingly recommended in academic publishing.
- Responsible practice requires maintaining the AI system as a dialogic stimulus, not an intellectual substitute.

VII. DISCUSSION

In the scaffolding tradition, instructional support is typically understood as contingent, adaptive assistance designed to improve task performance and gradually fade as learner competence increases[8]. Recent applications of generative AI largely extend this paradigm, positioning AI as a tool to enhance self-regulated learning, metacognitive monitoring, and higher-order task execution[9-10].

In contrast, this framework, rather than treating AI primarily as a temporary support for performance, positions AI as a dialogic partner within a Socratic–experiential process aimed at developing situated human judgement. The focus shifts from task completion to the formation of a higher-order human capability encompassing cognitive discernment, ethical reasoning, identity, and contextual action. This extends beyond the predominantly cognitive and metacognitive orientation of existing AI scaffolding models.

Furthermore, while emerging work recognises the value of AI-supported dialogue and Socratic questioning[11-12] such approaches are typically framed as pedagogical techniques to enhance learning outcomes. The present framework instead treats dialogue as constitutive, arguing that judgement cannot be transmitted but must be constructed through sustained interrogation and reflection. Finally, it integrates experiential application as a necessary condition for capability formation, emphasising that judgement is stabilised only through enactment in complex, AI-augmented contexts. In this sense, the framework moves beyond scaffolded learning toward a developmental model of human capability.

VIII. IMPLICATIONS FOR PRACTICE

If AI-supported Socratic inquiry is accepted as a legitimate scholarly method, several implications follow. First, academic training may need to incorporate AI literacy that emphasises critical engagement rather than passive use. Second, scholarly writing may increasingly involve iterative dialogue between researchers and computational systems. Third, research methods may expand to include human–AI collaborative inquiry as a recognised mode of intellectual exploration.

Rather than replacing traditional scholarship, AI-mediated dialogue may become another tool within the broader ecosystem of academic research practices.

IX. CONCLUSION

Generative artificial intelligence presents both challenges and opportunities for academic knowledge production. While concerns about academic integrity are legitimate, framing AI solely as a threat to scholarship overlooks its potential role as a cognitive support system.

This paper has argued that when generative AI is used through structured Socratic questioning, it can function as a dialogic scaffold that supports intellectual exploration. Drawing on Socratic pedagogy, Vygotskian scaffolding, and extended cognition theory, the paper conceptualized AI-mediated dialogue as a form of distributed cognition in which human researchers remain the primary agents of interpretation and synthesis.

The key ethical condition is that AI must remain a reflective partner rather than an autonomous author. When used critically and transparently, AI-supported Socratic inquiry may enrich scholarly practice by expanding the cognitive environment in which academic thinking occurs.

Future research should further empirically investigate how scholars interact with AI systems during conceptual development and how such interactions influence research creativity, critical reasoning, and knowledge production.

Authors disclosure: We acknowledge that generative AI was used during the ideation and drafting process while emphasising that we retain responsibility for the final content. The Socratic dialogue used to test our proposed model is reproduced in Appendix 1.

REFERENCES

[1] Zhai, C., Wibowo, S. & Li, L.D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environment,* (11)28. https://doi.org/10.1186/s40561-024-00316-7

[2] Setyawan, Cahya & Zulaeha, Zulaeha & Muhammad, Mahdir & Qurrotaa'yun, Yumna. (2025). Cognitive Scaffolding Through AI Writing Assistants: Mixed Methods Evidence from Arabic Language Education Students. Uktub: *Journal of Arabic Studies*. 5. 312-325. 10.32678/uktub.v5i2.48.

[3] Plato. (Trans.1967). *Plato in twelve volumes, vol 3* (W. Lamb Trans.). Cambridge, MA: Harvard University Press.

[4] Altorf, Hannah. (2016). Dialogue and discussion: Reflections on a Socratic method. *Arts and Humanities in Higher Education*. 18. 10.1177/1474022216670607.

[5] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* Cambridge, Mass.: Harvard University Press.

[6] Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, *58*(1), 7–19. http://www.jstor.org/stable/3328150

[7] Mochizuki, Ichitaro (2024). Co-evolutionary Intelligence: Rethinking Human–AI Interaction. https://philpapers.org/rec/MOCCIR, viewed 14/03/2026.

[8] van de Pol, Janneke & Volman, Monique & Beishuizen, Jos. (2010). Scaffolding in Teacher–Student Interaction: A Decade of Research. *Educational Psychology Review*. 22. 271-296. 10.1007/s10648-010-9127-6.

[9] Keyang Qian, Shiqi Liu, Tongguang Li, Mladen Raković, Xinyu Li, Rui Guan, Inge Molenaar, Sadia Nawaz, Zachari Swiecki, Lixiang Yan, Dragan Gašević (2026). Towards reliable generative AI-driven scaffolding: Reducing hallucinations and enhancing quality in self-regulated learning support. *Computers & Education*, 240, https://doi.org/10.1016/j.compedu.2025.105448.

[10] Tsakeni, M., Nwafor, S. C., Mosia, M., & Egara, F. O. (2025). Mapping the Scaffolding of Metacognition and Learning by AI Tools in STEM Classrooms: A Bibliometric–Systematic Review Approach (2005–2025). *Journal of Intelligence*, *13*(11), 148. https://doi.org/10.3390/jintelligence13110148

[11] Linjin Xi, Yi Zhang, Qiyun Wang (2026). Investigating the effects of an LLM-based Socratic conversational agent on students' academic performance and reflective thinking in higher education, *Computers & Education*, 241. https://doi.org/10.1016/j.compedu.2025.105494.

[12] Wegerif, R., & Casebourne, I. (2025). A dialogic theoretical foundation for integrating generative AI into pedagogical design. *British Journal of Educational Technology*, 00, 1–16. https://doi.org/10.1111/bjet.70026

Appendix 1: A Socratic Dialogue with ChatGPT

The topic is Core skills that will be required by employers in 2030.

Ground rules:

- ChatGPT will primarily ask questions, not give answers
- ChatGPT will challenge assumptions and contradictions
- PD will respond, refine, and sometimes revise my position
- Periodically, ChatGPT will synthesise emerging insights.

ChatGPT: When we say "core skills for 2030," are we assuming that:

the nature of work is fundamentally changing, or

that the nature of skills is being redefined?

What is changing more significantly by 2030, the tasks people perform, or the capabilities humans must possess to remain valuable?

PD: Capabilities humans must develop. It is impossible to accurately predict what specific tasks people will perform but the pace of change will continue to accelerate.

ChatGPT: When you say *capabilities*, do you mean:

Stable, enduring human attributes (e.g., critical thinking, creativity), or

Adaptive meta-capabilities (e.g., learning how to learn, sense-making, navigating ambiguity), or a hybrid of both?

PD: I think a hybrid of both.

ChatGPT: If *everyone* claims creativity, critical thinking, and collaboration as "core skills," do they actually function as *differentiators* in 2030? How would you categorise capabilities (stable vs adaptive vs hybrid)? Do you believe traditional "core skills" are still distinctive, or merely baseline? If baseline—what sits *above* them?

PD: Hybrid. Traditional core skills I believe are merely baseline. I propose that "humanness" will sit above them.

ChatGPT: When you say *humanness sits above baseline skills*, are you referring to:

Emotional depth (empathy, ethical judgement, care)

Cognitive uniqueness (originality, intuition, sense-making)

Relational capacity (trust-building, influence, social intelligence)

Existential qualities (purpose, meaning-making, identity)

or something else?

If an employer in 2030 said, *"We hire for humanness,"* what would they actually be observing, measuring, or rewarding?

There is a potential contradiction embedded in your idea:

AI is increasingly capable of: mimicking empathy generating creative outputs simulating conversation

So…

What aspects of "humanness" are:

Non-replicable by AI, and

Valuable enough to be economically rewarded?

Are relational and emotional skills truly "baseline"… or do they have advanced forms that remain highly differentiating?

PD: Of the 4 things you first suggested I would say relational capacity and emotional depth are baseline skills, but cognitive uniqueness and existential qualities are not potentially replicable by AI.

ChatGPT: If AI can already: generate novel combinations, simulate creativity, produce strategic options – then… What kind of "cognitive uniqueness" remains distinctly human? Is it:

Judgement under ambiguity?

Taste (aesthetic or strategic discernment)?

Framing the right problem rather than solving given ones?

Moral or contextual reasoning?

Existential Qualities: are they valuable or abstract? How do "purpose," "meaning," or "identity" translate into:

organisational value?

hiring decisions?

performance outcomes?

PD: Yes, I still believe that relational/emotional skills are baseline but am open to them containing elite, differentiating factors that so far have not been specifically identified by the WEF (and other) surveys.

I would say that cognitive uniqueness is the human ability to "think outside the square", especially in situations where AI relies on rational arguments.

An employer would value people who have a strong sense of identity that aligns with company values and goals.

ChatGPT: You've defined cognitive uniqueness as: *"thinking outside the square… where AI relies on rational arguments."*

Let me challenge this directly:

AI is already capable of: generating unconventional ideas recombining domains producing "creative" outputs at scale

What actually differentiates *human cognition* may not be novelty alone, but something more precise. Consider this: Is the real distinction not *novelty*, but judgement about novelty? For example: deciding **which ideas matter** knowing **when to break rules** sensing **contextual appropriateness** integrating **ethical, social, and temporal consequences.**

PD: Yes, judgement about novelty sums up the difference (all your examples apply). Alignment is the ability to navigate tension between personal identity and organisational demands – the 2030 employer will want someone who constructively reshapes the organisation when needed.

ChatGPT: You are implicitly moving toward a Model something like:

Layer 1: Baseline (Functional Human Skills)

Communication

Collaboration

Emotional literacy

Layer 2: Advanced Cognitive Capability

Judgement

Sense-making

Problem framing

Layer 3: Existential Capability

Identity

Meaning-making

Ethical positioning

PD: Agreed. But the real value lies in how these layers *interact under conditions of uncertainty and AI augmentation.* What if the true "core skill" for 2030 is something like:

The capacity to exercise situated human judgement in complex, AI-augmented environments.

Where:

Cognitive uniqueness = generating and framing possibilities

Judgement = selecting what matters

Existential grounding = anchoring decisions in values/identity

Relational capability = enacting change socially.

ChatGPT: So, to summarise your emerging framework. Your argument now appears to be this:

By 2030, traditional core skills will remain necessary but insufficient.

Communication, collaboration, and emotional/relational competence will likely function as baseline expectations rather than true differentiators.

What will matter more is a higher-order human capacity:

The capacity to exercise situated human judgement in complex, AI-augmented environments.

This is not a rejection of traditional skills, but a reframing of them. In your model, they are the foundation upon which a more distinctively human layer sits.

The hybrid structure you are proposing:

Layer 1: Baseline human skills

communication

collaboration

emotional literacy

relational competence

Layer 2: Higher-order human differentiation

cognitive uniqueness, understood not as novelty alone, but as **judgement about novelty**

the ability to frame problems, discern what matters, and assess contextual appropriateness

Layer 3: Existential-human grounding

identity

values

meaning

the capacity to navigate tension between personal convictions and organisational demands

Integration mechanism

These layers come together through: **situated human judgement**

That is the mechanism that activates the model in practice. It is what enables a person not merely to perform, but to respond wisely, ethically, and constructively in fluid environments shaped by AI, uncertainty, and organisational change.

And your implied ideal employee for 2030 is not someone who simply “fits” an organisation, but someone who can:

- interpret complexity
- make context-sensitive judgements
- remain grounded in identity and values
- and constructively reshape the organisation when required.