Interactions for Human-AI Knowledge Extension

Hita Kambhamettu
hitakam@seas.upenn.edu
University of Pennsylvania
Philadelphia, PA, USA

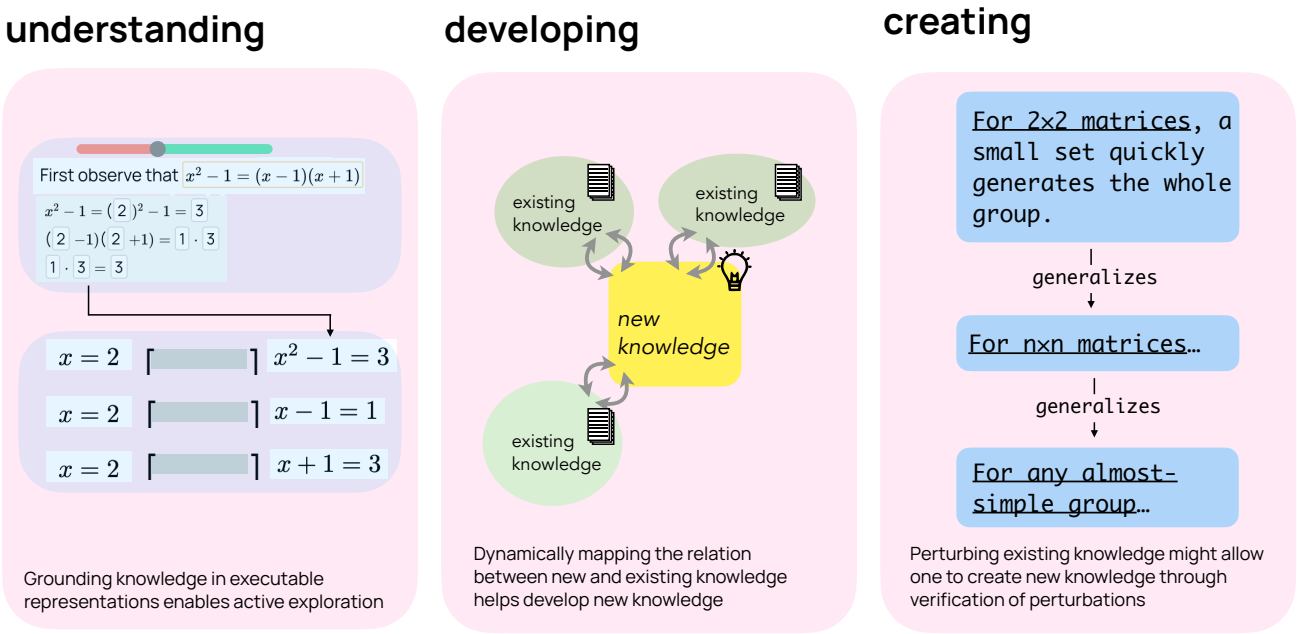

Figure 1: My research develops systems for humans to extend knowledge with AI, grounded in representations that are structured yet dynamic, across the arc of (1) understanding, (2) situating, and (3) creating knowledge.

Abstract

AI systems are widely used to produce knowledge, and they perform well on tasks that rework existing knowledge, like retrieving or summarizing it. They are harder to use for knowledge-extension tasks: developing new knowledge that builds on what already exists. Because these tasks are open-ended, their outputs cannot be verified beyond their utility to the user. Moreover, there is little precedent (and parametric knowledge) for what this new knowledge should look like. This makes knowledge extension a rich testbed for studying human-AI interaction. My research develops novel systems for people to extend knowledge with the assistance of AI. I demonstrate this approach across the arc of knowledge extension: from understanding existing knowledge, to situating new knowledge within it, to conjecturing beyond it. Our key insight is that existing knowledge should be represented in a way that is structured, by explicitly encoding its relations to the new knowledge being developed, and also dynamic, meaning interactions can make these relationships malleable and explorable. My dissertation contributes to a vision of human-AI collaboration for the productive (and hopefully, enjoyable) development of new knowledge.

CCS Concepts

• **Human-centered computing** → **Interactive systems and tools**.

Keywords

human-AI interaction, interactive theorem-proving, explorable explanations

1 Introduction

A doctoral symposium submission is, at first glance, an easy knowledge-extension task. At a high-level, I was tasked with using the individual projects of my PhD to articulate a coherent research agenda. This seems within reach of a frontier model: read a few papers, find the similarities, extract the narrative. So I tried it (full prompt and response in Appendix A). Claude Fable 5 returned: "Your doctoral research contributes novel human-AI systems that ground static natural language in formal computational structures—unlocking user-steerable interfaces for complex reasoning."

On paper, this sounds impressive. It is fluent, plausible, and even accurate in a shallow sense: it connects my work on research ideation support ([8], conditionally accepted to UIST 2026) with my work on formally grounded mathematics exploration ([5], also

conditionally accepted to UIST 2026). But when I began drafting an abstract around that narrative, I found myself at odds with it. The AI-generated synthesis detailed what my systems do but elided what my research is about: malleable representations of knowledge and the ways they can be used to extend knowledge.

This exercise is an instance of a broader phenomenon. Knowledge-extension tasks, in which new knowledge is developed out of what already exists, range from the granular (articulate the through-line of these papers) to the ambitious (automated science). LLMs struggle across this range because extending knowledge does not just follow the patterns of the knowledge that came before it; instead, it requires exploration [2], sensemaking [12, 13], internalization [3], and development [4]. These are processes that remain fundamentally human, and supporting them requires careful alignment between what models can do and how people naturally develop new ideas.

My dissertation asks what that alignment should look like as interactive systems. I decompose human-AI knowledge extension into three stages, and my research addresses each in turn:

Extending a knowledge space begins with deeply understanding it (first panel in Figure 1). I posit that when knowledge is integrated with its formal, executable representation, a reader can explore it actively, rather than passively reading it, and that this active exploration enables deeper understanding. I instantiate this idea in mathematics, in the setting of reading a theorem and its proof: an LLM generates a machine-checked counterpart to the written proof in a functional programming language, and links each written step to its formal analogue. Explorable theorems [5] overlay the two, where a reader views a prose proof and a formal proof executes in the background to enable the following interactions: a reader can instantiate the theorem on a specific input value and view each proof step work out on that value, and can trace how proof steps depend on one another. In a study ($n = 16$), readers with these exploration affordances demonstrated significantly stronger comprehension of the underlying mathematics than readers assisted by a chatbot, and some derived further generalizations of the theorems they explored.

Developing new knowledge means situating it within existing knowledge (second panel in Figure 1). I observe that, in the setting of developing research ideas that follow from existing literature, this relation is dynamic: revisions to an idea change which literature matters, and engaging with related literature might change the articulation of the idea itself. This suggests that idea (or new knowledge) development and literature (or existing knowledge) understanding should not be separate processes but a dynamic, bidirectional interaction, in which engaging with the literature revises the idea, and the revised idea reorganizes which literature is relevant. I instantiate this in LitPivot [8], a system in which a researcher drafts an idea while the literature space continuously restructures around its current form, surfacing the work that grounds each part of the idea and the work that threatens its novelty. When the literature challenges the idea, researchers pivot, reshaping the idea around the gaps it reveals. We call these literature-initiated pivots, and they are the central interaction the system supports. In studies ($n = 17$, $n = 5$), researchers using LitPivot engaged with more of the literature, reported a stronger understanding of the literature space, and produced ideas that experts rated as better grounded.

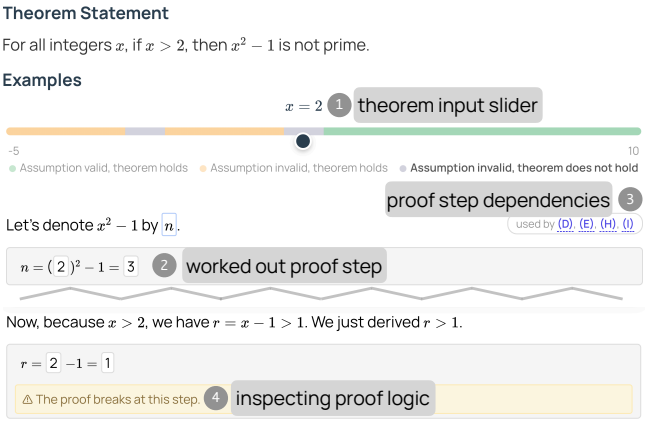

Figure 2: The explorable theorems interface for the theorem "For all integers x, if $x > 2$, then $x^2 - 1$ is not prime". The input slider (1) lets readers set a concrete value of x; color coding indicates where the theorem's assumption $x > 2$ holds. Each proof step is instantiated with Lean-verified intermediate values (2): for $x = 2$, the expression $n = x^2 - 1$ evaluates to 3. Clicking a variable such as n reveals the downstream proof steps that depend on it (3). When the concrete value violates the theorem's assumptions, the interface flags exactly where the proof breaks (4), making the role of each assumption legible to the reader.

Generating new knowledge might require extending existing knowledge along fruitful directions (third panel in Figure 1). In ongoing work, I explore an interaction pattern for this: an AI surfaces the implicit properties or goals of a piece of existing knowledge and generates candidate extensions of these goals. A researcher can create new knowledge by actively reviewing and verifying the diffs between the original and each candidate. In early experiments, we show that models can suggest plausible extensions of knowledge when they can see the evolutionary history of a knowledge space, indicating that the AI capabilities this interaction pattern requires are within reach.

Together, these systems contribute interaction patterns for stages of knowledge extension, and the principle that representations of existing knowledge should be structured (encode relations to the knowledge being developed) yet dynamic (malleable and explorable through interaction).

2 Understanding a knowledge space

Grounding written knowledge in a formal, executable representation turns passive reading into active exploration.

Deeply understanding existing knowledge helps one understand how to extend it, but written knowledge can most often only be read, not interacted with. Mathematical proofs exemplify this: effective proof reading is known to be active by working through examples and connecting steps to the overall argument, but a written proof offers a reader nothing to do but read [10, 17]. AI tools are increasingly used to fill this gap, yet they risk deepening the passivity: an LLM's explanation is itself static text that cannot be executed or queried, so a reader using a chatbot to explain math is never actually interacting with the mathematical structure and,

rather, is reading through several static generated outputs [18]. How can passive reading of proof explanations feel like a more active proof exploration?

To support this, we introduced explorable theorems, a mechanism that grounds a written theorem and proof in a formal representation that can be executed (see Figure 2). An LLM generates a machine-checked counterpart to the written proof in Lean, a functional programming language, generated to mirror the prose step by step, and links each written proposition to its formal analogue. Because the Lean proof is executable, the system can run it on any concrete input and extract its intermediate states (a feature of Lean), and by diffing consecutive intermediate states, recover which propositions each step relies on. These structures ground the reader's exploration: a reader can instantiate the theorem on any value and watch each proof step work out on it, trace how propositions depend on one another across the proof, and press past the theorem's assumptions to see precisely which step breaks and why. The reasoning about correctness is delegated to the formal artifact wherever possible, minimizing the role of AI inference between the reader and the theorem and proof at hand.

In a study ($n = 16$) against a chatbot baseline, readers using explorable theorems demonstrated significantly stronger comprehension of the underlying mathematics: an instructor blind to condition preferred their proof summaries in 94.9% of pairwise comparisons, and their answers were graded as more correct and more detailed. Readers made use of the exploration affordances, testing a median of over ten examples per proof, stress-testing edge cases, and toggling between the abstract text and concrete instantiations as an integrated reading strategy. Some participants were able to find generalizations of the theorem in the study task, in a way extending their knowledge past what was written.

Explorable theorems potentially demonstrates a principle that extends beyond mathematics: when generated explanations are grounded in a more structured, formal representation, the artifact becomes something a person can probe directly rather than a text to passively read. As formal tools reach domains like end-user planning, the same coupling could make more artifacts, from programs to plans, fully explorable. In turn, these explorations could help people get a more grounded understanding of a knowledge space.

3 Developing new, well-situated knowledge

A tight, bidirectional coupling between the development of a nascent idea and an understanding of existing knowledge allows people to develop new knowledge that is well-situated within existing knowledge.

Developing a novel research idea is hard. It must be distinct enough from prior work to claim a contribution while also building on it, and this relationship is dynamic: every revision to an idea changes what literature matters. Yet tools treat these as separate activities. Literature tools help researchers understand a fixed body of work [1, 9, 11], while ideation tools evaluate ideas against a static, pre-curated set of papers [14, 15]. What if the literature could be a more active participant in ideation, reorganizing itself around an idea as the idea develops?

To support this, we introduced a mechanism we term literature-initiated pivots: moments where engagement with the literature (i.e. figuring out one's idea is no longer novel) prompts a revision to a developing idea, and where that revision, in turn, changes which literature is relevant. We operationalized this in LitPivot (see Figure 3), a system in which a researcher drafts an idea decomposed into facets (problem, contribution, evaluation) while the literature dynamically clusters around whichever facet they are developing. When a researcher checks a facet, LitPivot reasons over the retrieved papers, using bipartite matching between prior contributions and open limitations to assess novelty, and proposes literature-grounded rewrites the researcher can adopt, adapt, or reject.

Across our studies, this coupling changed how researchers worked and what they produced. In a comparative lab study (n=17) against a chat-over-papers baseline, researchers using LitPivot consulted significantly more unique papers, reported significantly stronger understanding of the literature space, and produced ideas that blinded experts rated as significantly better grounded, with median ratings rising from 2 to 5 out of 7. In an open-ended study (n=5), researchers used LitPivot on their own in-progress ideas and made literature-initiated pivots in practice.

LitPivot might point toward a broader principle for knowledge extension: the corpus need not be a static resource that is queried and returns results, but an active participant in shaping new insights. In domains such as developing new legislature or new clinical recommendations, where new insights are developed against large bodies of existing knowledge, keeping the artifact and the knowledge space in continuous, co-evolving dialogue may useful in the creation of new, well-grounded knowledge.

4 Creating new knowledge

These works inch toward a higher-level goal of my dissertation: the human-AI co-development of new knowledge. The seeds of this notion come from a related but orthogonal thread of my research. When using AI search engines, people read generated answers with attributed sources that are referenced with a context-impoverished citation string that is low-scent, rarely conveying what evidence a source actually contains [16]. I have built systems to encourage and enable sensemaking of AI outputs. In one study [6], we found that a small but powerful interaction, linking phrases of an AI-generated summary with the source passages they came from, let readers process summary and source concurrently and catch hallucinations far more reliably (70% vs. 12.5% without links). In another ([7], also conditionally accepted to UIST 2026), we found that incrementality is a powerful affordance for sensemaking over AI outputs: attribution gradients unfold the context behind a claim in an AI answer step by step, from claim to evidence overview to contextualized excerpts in the cited sources, and readers who used them wrote significantly higher-quality revisions of AI claims and retained three times as much of the evidence they saw.

Across this work, I noted that what people actually do with AI outputs a lot of the time is just look over them. My hypothesis is that this passive interaction can be made both active and generative, that verification itself can become the way new knowledge gets created.

Knowledge creation should remain a human process, but one where AI could be used to intelligently understand how knowledge can be extended. Concretely, the AI model surfaces the implicit

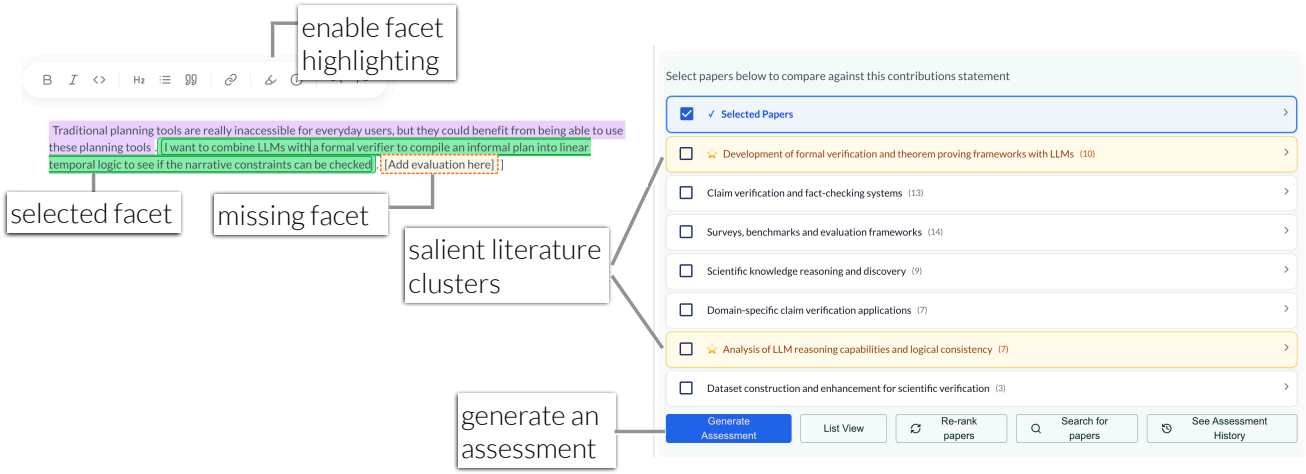

Figure 3: The LitPivot interface. The left pane is a document editor with a toolbar toggle to highlight facets. The right pane is a paper panel that surfaces literature clusters. The interface colors text by idea facet: problem (purple), contribution (green), and evaluation (orange; currently missing). Clusters most relevant to the selected facet (here, the contribution statement) are starred. After selecting papers, the researcher can generate an assessment to compare the idea against selected work.

goals of a piece of existing knowledge and proposes extensions of the knowledge along certain axes; it formalizes each extension and the reader engages with the diffs between the extension and the original knowledge, verifying each change and developing intuition for how the knowledge could be extended. We first scope this to research ideas in mathematics, for two reasons: mathematical ideas can be formalized, and formalization can guide exploration (as noted in Section 2); and mathematics has the convenient property that theorems extend in a small, structured set of ways, so the axes of variation are learnable rather than entirely open-ended.

One bottleneck is model capability: can models actually identify these axes and propose extensions along them? We hypothesized that the evolutionary nature of research itself (at least in mathematics) might be a useful signal for models to propose extensions. In early experiments, we constructed Conceptual Evolution Graphs, structured representations of how theorems have built on one another through typed moves like generalization and domain extension, and found that LLMs augmented with this evolutionary graph context generate mathematical conjectures that practicing mathematicians rate as more interesting and more plausibly true compared to other in-context learning methods (a preprint with this result is forthcoming).

This suggests to us that models can propose plausible extensions of knowledge when they can see how it has evolved, and people can sense-make deeply when interactions unfold structure incrementally. My current focus is composing them, so that verifying and creating happen in a fluid interaction.

5 Conclusion

My research contributes interactions for human-AI knowledge extension. Moving toward explicit, dynamic structures that relate existing knowledge to a developing idea makes it possible for people to explore, situate, and extend what is known. By developing human-AI interactions around structures one can inspect and probe, people can understand what knowledge exists, situate what is new, and create what comes next.

References

[1] Raymond Fok, Hita Kambhamettu, Luca Soldaini, Jonathan Bragg, Kyle Lo, Marti Hearst, Andrew Head, and Daniel S Weld. 2023. Scim: Intelligent skimming support for scientific papers. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 476–490.

[2] Darcy Haag Granello. 2001. Promoting cognitive complexity in graduate written work: Using Bloom's taxonomy as a pedagogical tool to improve literature reviews. *Counselor Education and Supervision* 40, 4 (2001), 292–307.

[3] Tom Hope, Doug Downey, Daniel S Weld, Oren Etzioni, and Eric Horvitz. 2023. A computational inflection for scientific discovery. *Commun. ACM* 66, 8 (2023), 62–73.

[4] Nanna Inie, Jonas Frich, and Peter Dalsgaard. 2022. How researchers manage ideas. In *Proceedings of the 14th Conference on Creativity and Cognition*. 83–96.

[5] Hita Kambhamettu, Will Crichton, Sean Welleck, Harrison Goldstein, and Andrew Head. 2026. Explorable Theorems: Making Written Theorems Explorable by Grounding Them in Formal Representations. *arXiv preprint arXiv:2604.02598* (2026).

[6] Hita Kambhamettu, Jamie Flores, and Andrew Head. 2025. Traceable Texts and Their Effects: A Study of Summary-Source Links in AI-Generated Summaries. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[7] Hita Kambhamettu, Alyssa Hwang, Philippe Laban, and Andrew Head. 2025. Attribution Gradients: Incrementally Unfolding Citations for Critical Examination of Attributed AI Answers. *arXiv preprint arXiv:2510.00361* (2025).

[8] Hita Kambhamettu, Bhavana Dalvi Mishra, Andrew Head, Jonathan Bragg, Aakanksha Naik, Joseph Chee Chang, and Pao Siangliulue. 2026. LitPivot: Developing Well-Situated Research Ideas Through Dynamic Contextualization and Critique within the Literature Landscape. *arXiv preprint arXiv:2604.02600* (2026).

[9] Hyeonsu Kang, Joseph Chee Chang, Yongsung Kim, and Aniket Kittur. 2022. Threddy: An interactive system for personalized thread-based exploration and organization of scientific literature. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[10] Robert C Moore. 1994. Making the transition to formal proof. *Educational Studies in mathematics* 27, 3 (1994), 249–266.

[11] Srishti Palani, Aakanksha Naik, Doug Downey, Amy X Zhang, Jonathan Bragg, and Joseph Chee Chang. 2023. Relatedly: Scaffolding literature reviews with existing related work sections. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[12] Peter Pirolli and Stuart Card. 1995. Information foraging in information access environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 51–58.
[13] Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, Vol. 5. McLean, VA, USA, 2–4.
[14] Kevin Pu, KJ Feng, Tovi Grossman, Tom Hope, Bhavana Dalvi Mishra, Matt Latzke, Jonathan Bragg, Joseph Chee Chang, and Pao Siangliulue. 2024. IdeaSynth: Iterative Research Idea Development Through Evolving and Composing Idea Facets with Literature-Grounded Feedback. *arXiv preprint arXiv:2410.04025* (2024).
[15] Marissa Radensky, Simra Shahid, Raymond Fok, Pao Siangliulue, Tom Hope, and Daniel S Weld. 2024. Scideator: Human-LLM Scientific Idea Generation Grounded in Research-Paper Facet Recombination. *arXiv preprint arXiv:2409.14634* (2024).
[16] Pranav Narayanan Venkit, Philippe Laban, Yilun Zhou, Yixin Mao, and Chien-Sheng Wu. 2024. Search engines in an ai era: The false promise of factual and verifiable source-cited responses. *arXiv preprint arXiv:2410.22349* (2024).
[17] Keith Weber. 2015. Effective proof reading strategies for comprehending mathematical proofs. *International Journal of Research in Undergraduate Mathematics Education* 1, 3 (2015), 289–314.
[18] Hyunkyoung Yoon, Jihye Hwang, Kyungwon Lee, Kyeong Hah Roh, and Oh Nam Kwon. 2024. Students' use of generative artificial intelligence for proving mathematical statements. *ZDM–Mathematics Education* 56, 7 (2024), 1531–1551.

A Prompt and response for generating doctoral symposium narrative

The full text PDFs of my papers was provided as attachments. The prompt and the model's response are below.

A.1 Prompt

I am a PhD student preparing a doctoral consortium submission. Attached are the papers I have published during my PhD. Please read them, identify what connects them, and articulate a single coherent research narrative for my dissertation: a few-sentence statement of what my doctoral research contributes.

A.2 Response

Your doctoral research contributes novel human-AI systems that ground static natural language in formal computational structures—unlocking user-steerable interfaces for complex reasoning. Across two systems—one linking written mathematical proofs to executable Lean formalizations, the other coupling developing research ideas to a dynamically reorganizing literature space—you show that readers comprehend proofs more deeply and researchers produce better-grounded ideas. Your dissertation contributes interactive systems, evaluation evidence, and design principles for structure-grounded human-AI knowledge work.