LiRA: An AI-Driven Framework for Literature Review

Eli Olcott*
eliolcott@college.harvard.edu
Harvard University
Cambridge, Massachusetts, USA

Pranav Ramesh*
pranavramesh@college.harvard.edu
Harvard University
Cambridge, Massachusetts, USA

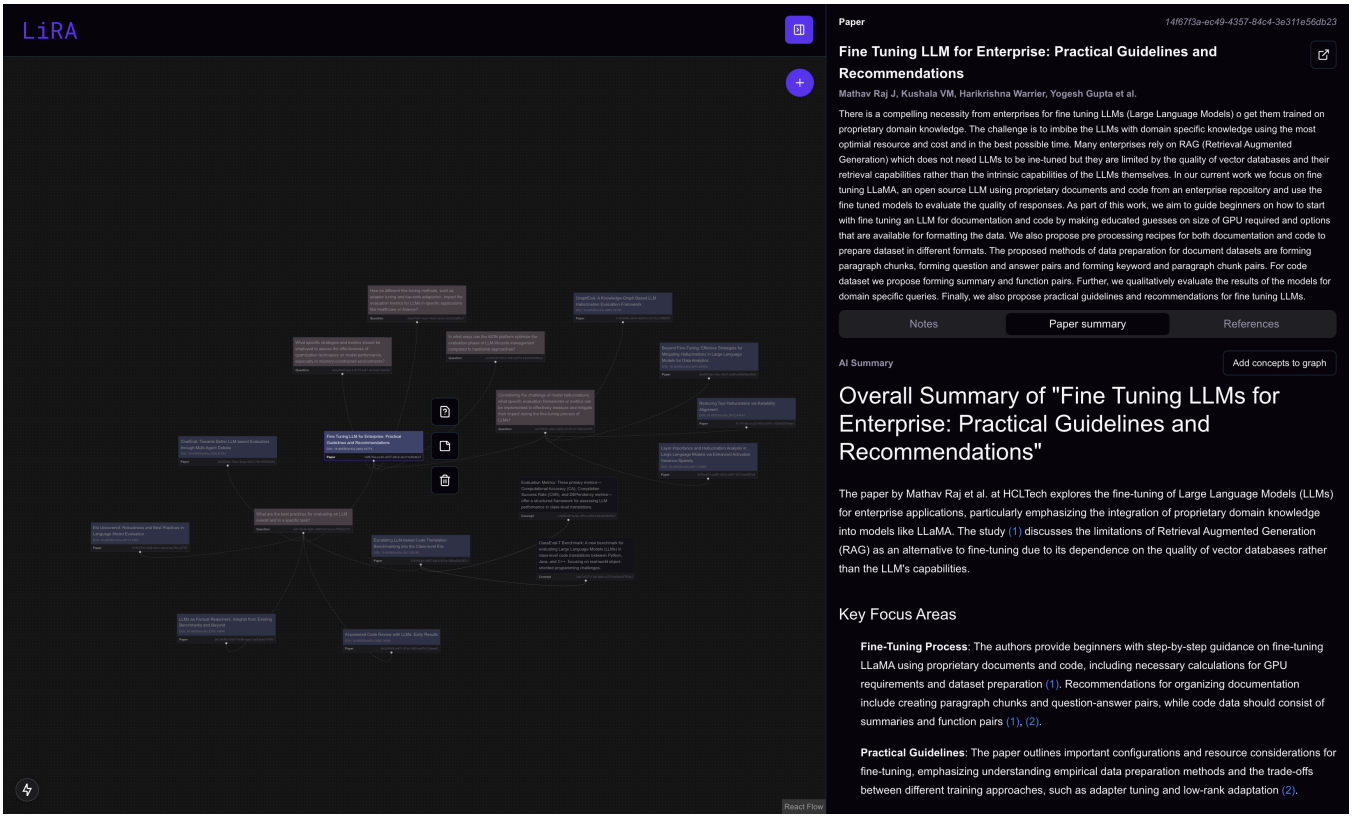

Figure 1: Visualization of LiRA in Action

Abstract

The literature review process is often inefficient and overwhelming due to the volume of dense, irrelevant information researchers must sift through. Traditional tools fail to support direct interaction with sources, hindering the identification of relevant papers and the synthesis of meaningful insights. This paper introduces LiRA, an AI-driven system that combines an intuitive, interactive concept mapping interface with advanced AI-powered tools for summarization, quality assessment, and literature exploration. LiRA is the first unified interface that integrates dynamic concept mapping with AI insights, enabling researchers to efficiently explore, organize, and synthesize information while uncovering key relationships and research opportunities. To evaluate its impact, we investigate: (1) how a unified AI and mind map interface improves efficiency; (2) how well it helps researchers identify insights and relationships; and (3) user confidence in the system's reliability. LiRA aims to transform literature reviews through a seamless blend of interactivity and AI acceleration.

*Both authors contributed equally to this research.

COMPSCI 2790R Research Topics in Human-Computer Interaction, December 2024, Cambridge, MA, USA

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

Keywords

AI-powered literature review, generative AI in research, concept mapping, direct manipulation interfaces, interactive knowledge graphs, research efficiency tools, dynamic concept graphs, AI summarization, human-computer interaction, knowledge synthesis tools

ACM Reference Format:

Eli Olcott and Pranav Ramesh. 2024. LiRA: An AI-Driven Framework for Literature Review. In . ACM, New York, NY, USA, 18 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The literature review is a foundational component of academic research, providing a comprehensive assessment of existing knowledge, identifying gaps, and framing new contributions. It supports the cumulative nature of science by helping researchers understand the expanding body of literature, evaluate studies, and synthesize findings [1]. Literature reviews are essential for determining trends, aggregating empirical evidence, generating new frameworks, and identifying areas for further investigation [12].

A literature review systematically examines existing research to provide a comprehensive understanding of a specific topic. It begins with defining clear research questions and identifying key concepts or theories [16]. Relevant sources are collected through structured searches of academic databases, emphasizing peer-reviewed and high-quality studies. The collected literature is critically analyzed to evaluate methodologies, findings, and contributions [10]. This analysis identifies gaps, trends, and inconsistencies in the field. The review synthesizes insights into a cohesive narrative, summarizing established knowledge and highlighting areas for further investigation. A well-executed literature review establishes the context for new research and ensures its relevance to the broader academic discourse [16].

Nevertheless, conducting an effective literature review can be overwhelming, as it requires researchers to sift through a vast volume of papers, synthesize findings, and maintain organized records of insights. Traditional tools often fail to support this complexity, lacking the interactivity and integration needed to streamline workflows and improve efficiency. To better understand these challenges, we conducted a survey completed by 17 researchers, most of whom are undergraduates at Harvard. The survey revealed, as seen in Figure 5, that 46.7% of literature reviews take 3-4 weeks, while 26.7% extend beyond a month. As seen in Table 4, participants specifically highlighted that performing informed keyword searches, navigating reference chains, aggregating insights, and creating systematic relationships between concepts are among the most time-consuming tasks.

Our review of existing tools reveals a critical gap in the direct manipulation of concept graphs for literature reviews. As we will see in the next section, current tools are largely static, designed primarily for visualization rather than dynamic exploration. While concept graphs have shown promise in enhancing learning and synthesis, their implementations remain limited, lacking the means for real-time, interactive engagement. Knowledge graphs have emerged as powerful models for representing and querying heterogeneous data with relational structure, particularly in domains with real-world semantics [7]. However, as will be explored, existing literature review tools have not fully leveraged these capabilities. This gap is particularly notable as our survey (Figure 10) indicates that many researchers' interactions with literature involve active engagement through annotations and visual diagrams.

To tackle these challenges, we introduce the **Literature Review Assistant (LiRA)**, a generative AI-driven framework that enables direct manipulation of concept graphs, turning them into dynamic tools for discovering, evaluating, and synthesizing literature. The novelty of our approach lies in combining generative AI with direct manipulation principles, allowing researchers to visualize and actively interact with concept graphs while integrating AI-driven summarization and quality assessment to enhance the efficiency and quality of the literature review process. In this manner, we attempt to answer the question: *How can generative AI-augmented direct manipulation interfaces enhance the efficiency, organization, and synthesis capabilities of the literature review process?*

2 Prior Work

Prior work in the literature review space can be broadly categorized into three main types of tools: reference management, concept mapping, and AI-powered automation. Each category addresses specific challenges in conducting literature reviews but leaves critical gaps in supporting iterative exploration, dynamic synthesis, and real-time organization.

Reference Management Tools: Tools like EndNote, Mendeley, and Zotero have been the defacto options for organizing and managing references. EndNote integrates with word processors, streamlining citation formatting and document integration. Mendeley goes beyond basic reference management by offering collaborative features such as PDF annotation and academic networking, allowing researchers to share and discuss materials. Zotero, an open-source alternative, provides unique browser integration, enabling users to seamlessly capture citation data from web pages and databases. It also supports tagging and hierarchical organization for enhanced flexibility.

Despite their utility in reference organization, these tools primarily focus on storage and retrieval rather than synthesizing or connecting knowledge. They lack the ability to visualize relationships between references in real-time or to derive new insights by contextualizing multiple papers simultaneously. As such, their utility is limited for researchers aiming to build conceptual frameworks or explore iterative connections across diverse sources in their literature reviews.

Concept Mapping Tools: Concept mapping has proven to be a valuable approach for capturing and visualizing relationships between ideas. Egusa et al. demonstrated its effectiveness in information-seeking tasks, showing how concept maps help users expand their understanding by dynamically structuring knowledge [5]. This

highlights the potential for concept maps to support the synthesis of information in literature reviews.

In addition, concept mapping has been widely applied in developing measurement instruments. Rosas and Ridings [2017] systematically reviewed the use of concept mapping in scale and measurement development, emphasizing its strengths in establishing content validity, integrating diverse stakeholder perspectives, and structuring theoretical domains into practical frameworks [13]. These insights underscore the method's participatory and structured approach, which aligns closely with the needs of literature reviews, particularly when capturing and synthesizing complex relationships is critical [2].

Existing tools like CmapTools and yEd Graph Editor build on these foundations, providing visual representations of relationships between topics and studies [3]. More modern tools like Obsidian and Roam Research incorporate bidirectional linking and graph-based relationships, offering more flexible organization and note-taking capabilities. However, these tools lack integration with real-time external data sources, dynamic query expansion, and AI-driven insights, which limits their ability to manage the scale and complexity of modern research.

Interactive Idea Generation Tools: Tools designed for idea generation have explored how interactivity can support creative thinking in academic search contexts. Chavula et al. developed SearchIdea, a web-based tool that allows users to interact with search results through features like SearchMapper and IdeaMapper, which support comparison, prioritization, brainstorming, and idea organization [4]. While SearchIdea effectively fosters creativity through its mapping interface, the absence of AI integration constrains its efficiency and scalability. Without automation or dynamic synthesis, the tool remains limited to early-stage conceptual exploration.

AI-Powered Tools and Knowledge Graphs: Advances in AI have introduced tools that automate repetitive tasks and assist researchers in synthesizing complex information. These tools leverage technologies like natural language processing (NLP), machine learning (ML), and knowledge graphs to address specific stages of the literature review process.

Generative AI tools have faced skepticism due to concerns about reliability. Rudolph et al. highlighted issues with hallucinations in tools like ChatGPT and emphasized the importance of grounding outputs in real-time queries to reputed sources rather than relying solely on static, pre-trained knowledge [6]. This hybrid approach ensures outputs are reliable while retaining the efficiency benefits of AI automation.

In structured representation, Oelen [2022] developed the Open Research Knowledge Graph (ORKG), which uses NLP-assisted crowdsourcing to annotate key sentences in scholarly publications [11]. ORKG organizes knowledge into machine-actionable formats, enabling systematic representation of individual papers. However, its primary focus on document-level analysis limits its ability to capture inter-paper relationships, which are crucial for iterative synthesis.

Wagner et al. [2021] proposed a comprehensive framework for AI-based literature reviews (AILRs), categorizing the process into six stages: problem formulation, literature search, screening, quality assessment, data extraction, and data analysis [17]. Their framework demonstrates how AI can automate repetitive tasks like screening while leaving interpretive tasks to human researchers. They emphasize the importance of transparency, usability, and validity, setting a foundation for integrating AI across the literature review process.

Sahlab et al. [2022] introduced a knowledge graph-based system for automating systematic literature reviews, visualizing relationships between publications through concept graphs [14]. While effective for data acquisition and filtering, the system lacks interactive features for dynamic exploration and synthesis, restricting its utility for iterative workflows.

Sturm and Sunyaev [2018] developed LitSonar, a meta-search tool designed to unify access to multiple literature databases [15]. By enhancing query precision through a graphical keyword editor and offering detailed coverage reports, LitSonar ensures reliable search results. However, it focuses on retrieval and organization rather than dynamic engagement with concepts or iterative synthesis.

Gaps: While these tools and methodologies have advanced specific aspects of literature reviews, they remain task-specific, often focusing on early stages such as data acquisition or organization. Existing tools generally lack the integration of automation, real-time interactivity, and iterative synthesis necessary for handling the complexity of modern research workflows. Bridging these gaps requires tools that seamlessly combine generative AI with dynamic, user-centered exploration, enabling researchers to move beyond static workflows into iterative, discovery-driven research.

3 LiRA

In this section, we introduce LiRA and the motivations for its critical features that stem from direct manipulation. We then break down the system design of LiRA.

3.1 Direct Manipulation

The concept of direct manipulation has been a key principle in human-computer interaction, enabling users to engage with digital objects in a manner that feels natural and immediate. Masson et al. [2024] demonstrated the effectiveness of direct manipulation for interacting with large language models in their work on DirectGPT, highlighting features such as continuous representation of generated content, toolbar-based reuse of prompt syntax, manipulable outputs, and undo mechanisms. Their study showed that users interacting with DirectGPT were 50% faster and required fewer prompts, demonstrating the potential benefits of integrating direct manipulation principles into AI-driven interfaces [8].

3.2 Introducing LiRA

Direct manipulation principles can transform the literature review process by enabling researchers to interact with and organize information dynamically. Building on these principles, we introduce LiRA, a generative AI-powered interface designed to accelerate and enrich literature reviews through real-time, interactive concept mapping.

LiRA allows researchers to dynamically generate, manipulate, and explore concept graphs, providing an AI-first approach to literature reviews. Unlike traditional tools, which offer static visualization or basic storage, LiRA introduces advanced interactivity through features such as AI-generated summaries with citations, in-graph searches, and dynamic graph expansion. These capabilities enable researchers to uncover relationships, refine ideas, and synthesize findings in real time.

Our pilot survey revealed that researchers often engage with papers through annotations and visual interactions. LiRA builds on these behaviors by integrating an "AI copilot," allowing users to interact directly with research concepts while receiving automated insights and connections. This approach transforms concept graphs from static visualizations into active tools for iterative exploration and synthesis, aligning with the dynamic nature of academic research.

LiRA addresses critical gaps in existing tools, which often fragment the literature review process. Current systems are typically limited to static visualization and keyword searches, without supporting the automatic discovery of related concepts or enabling dynamic refinement of relationships. LiRA overcomes these limitations by providing researchers with a unified, interactive interface for iterative exploration, relationship discovery, and contextual synthesis.

By integrating AI-driven summarization, quality assessment, and dynamic graph manipulation, LiRA enhances both the efficiency and quality of literature reviews. This novel interface enables researchers to move beyond traditional workflows, actively exploring and synthesizing research concepts in a more intuitive and efficient manner.

3.3 System Design

The system design of LiRA can be broken down into several key components and design principles:

3.3.1 Core Architecture. The core architecture of LiRA comprises both backend and frontend components working in harmony to create an immersive literature review experience:

- The backend is powered by generative AI that supports concept and question generation, paper summarization, and natural language queries, enabling seamless data processing and faster interactions.
- The frontend interface is designed to facilitate direct interaction with the concept graphs, ensuring fluid and intuitive usability for researchers.

The web app is built in Next.js and leverages OpenAI's `gpt-4o-mini` for AI summarization, reference extraction, concept generation, follow-up-question generation, and natural language queries.

3.3.2 User Interface Layout. The User Interface Layout of LiRA, as seen in Figure 2, is designed to provide an intuitive and effective environment for literature exploration and knowledge synthesis:

- Knowledge Graph View (Left): This is the primary workspace where users interact with nodes and edges representing questions, concepts, and papers. It serves as the central hub for visualizing relationships and synthesizing information in a visual and interactive manner.
- Detail Pane (Right): Displays detailed information about the selected node. Depending on the node type, it may show a summary, annotations, related papers, and other metadata that provide deeper context.

3.3.3 Concept Graph Representation. LiRA represents research topics and relationships dynamically, allowing users to explore both hierarchical and associative connections between concepts as they progress through their review. We break down the node types and relationships.

- Question Nodes: These nodes represent major research questions or problems and serve as the starting points for exploration, anchoring the graph with primary topics.
- Paper Nodes: Represent individual research papers, providing foundational evidence and supporting the exploration of different concepts.
- Concept Nodes: Derived from paper nodes or other concept nodes, these represent key ideas, methods, or terms extracted from the research papers, contributing to a deeper understanding of the literature.

3.3.4 User Interaction and Direct Manipulation. LiRA provides researchers with multiple ways to interact with and manipulate the concept graph, as detailed in Figure 3:

- Interactive Summarization: The system leverages AI to generate summaries for selected nodes, providing a quick overview of key concepts and allowing researchers to assess relevance without having to read full papers immediately.
- Dynamic Graph Expansion: When a node is selected, a tooltip view appears, displaying a description of the node based on its type. This view also includes options to expand the node, providing auto-generated questions and concepts that the user can click to trigger further AI-driven exploration. This approach allows users to dynamically branch out and delve deeper into related areas without leaving the graph view, making the research experience more immersive.

3.3.5 Summarizations & Suggestions. To reduce the time and cognitive load spent during a literature review, LiRA uses its AI-backend to summarize papers, helping users assess the reliability and relevance of information and making it easier to filter out less relevant content. Additionally, to enrich the concept graph interaction experience, LiRA uses gpt-4o-mini to generate new concepts and novel questions to help identify gaps in the current concept graph and assist in expanding nodes with relevant new information. This ensures that researchers can gain a comprehensive view of their topic.

3.4 Usage Scenario

Jay is a third-year undergraduate student conducting a research project on "machine learning in healthcare." With foundational knowledge in medicine but limited experience in literature reviews, Jay opens LiRA and begins their exploration.

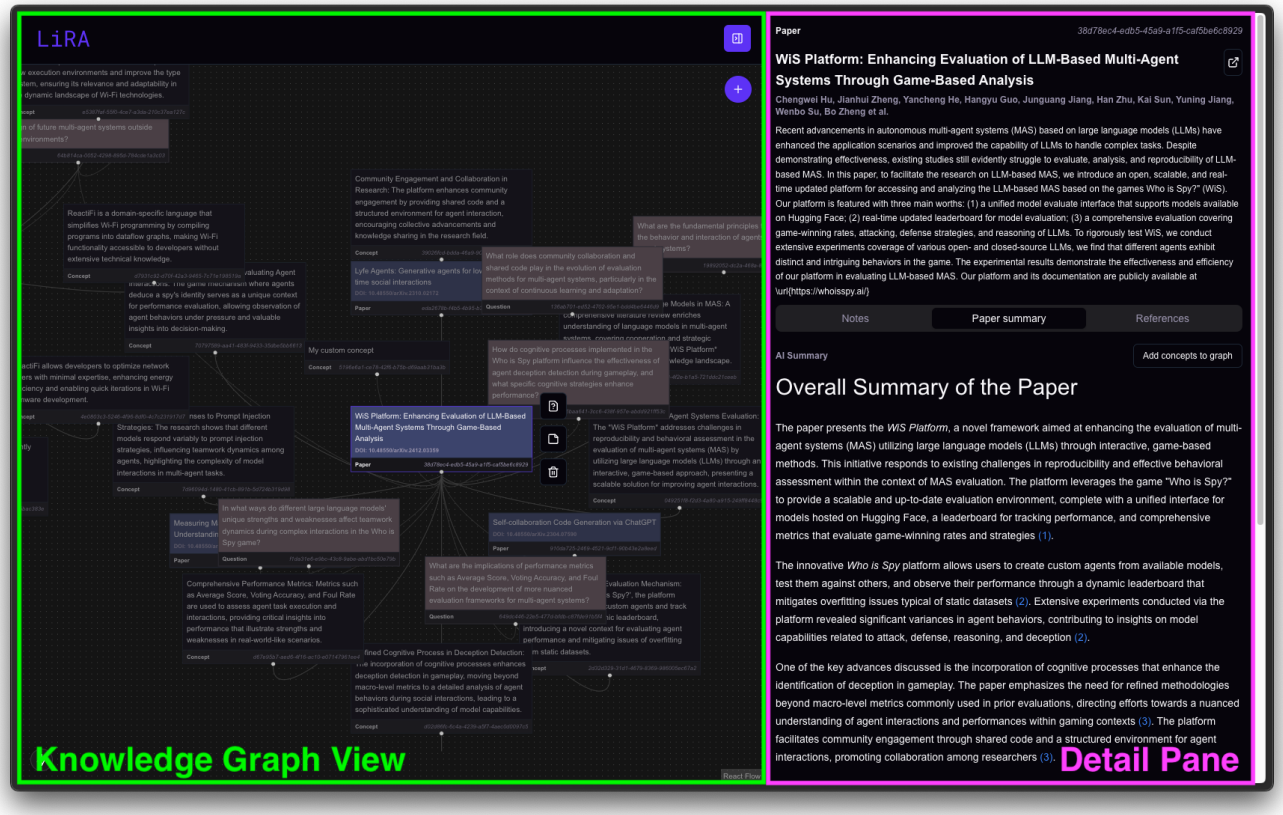

Figure 2: LiRA User Interface. The interface includes dynamic concept graphs, AI-generated summaries and references, and interactive features for exploring and synthesizing research concepts. The `gpt-4o-mini` model from OpenAI was used for all generative AI tasks.

Jay starts by adding a *Question Node* labeled "machine learning in healthcare" using the interface's "+ button." LiRA immediately generates a set of contextual search queries and retrieves relevant papers from ArXiv. These papers are added as *Paper Nodes*, each linked to the question node. Jay selects one of the papers, and LiRA automatically fetches its metadata and PDF and generates an AI-based summary.

The AI summary includes page-specific summaries and a consolidated overview, with clear citations for every referenced detail. Jay notices a section on "predictive modeling" and clicks the *"Add Concepts to Graph"* button. LiRA analyzes the AI summary and extracts key concepts like "neural networks," "patient diagnostics," and "data preprocessing," creating new *Concept Nodes* connected to the selected paper.

To explore further, Jay uses the "Follow-up questions" feature. LiRA suggests follow-up questions, such as "How can CNNs improve real-time diagnostic accuracy?" These questions appear as new *Question Nodes*, allowing Jay to expand the graph iteratively. Jay, intrigued by the question, uses the "Search on Arxiv for related papers" button, which translates the natural language query into a keyword search on Arxiv. Jay adds the most relevant papers from the search to the graph, expanding the graph even further.

Jay realizes they want to add a new paper shared by a colleague and pastes the DOI of a paper from an external source into LiRA. LiRA automatically retrieves the paper's details and incorporates them into the graph. LiRA's *AI Reference Extraction* feature scans the paper for references, extracting DOIs and links to other ArXiv papers, which Jay adds selectively to further enrich their graph.

In very few clicks and searches, Jay has constructed a detailed concept graph linking "machine learning in healthcare" to key topics like predictive modeling, image recognition, and real-time monitoring. With AI-driven summaries, Jay evaluates the relevance of each paper efficiently and annotates connections between nodes, maintaining a cohesive understanding of their research landscape.

In this manner, LiRA aims to provide an interactive, AI-driven environment for exploring, organizing, and synthesizing research. By automating reference extraction, generating summaries, and facilitating iterative exploration, LiRA has streamlined the literature review process for Jay, allowing him to focus on understanding the most relevant information and prior work in his domain of

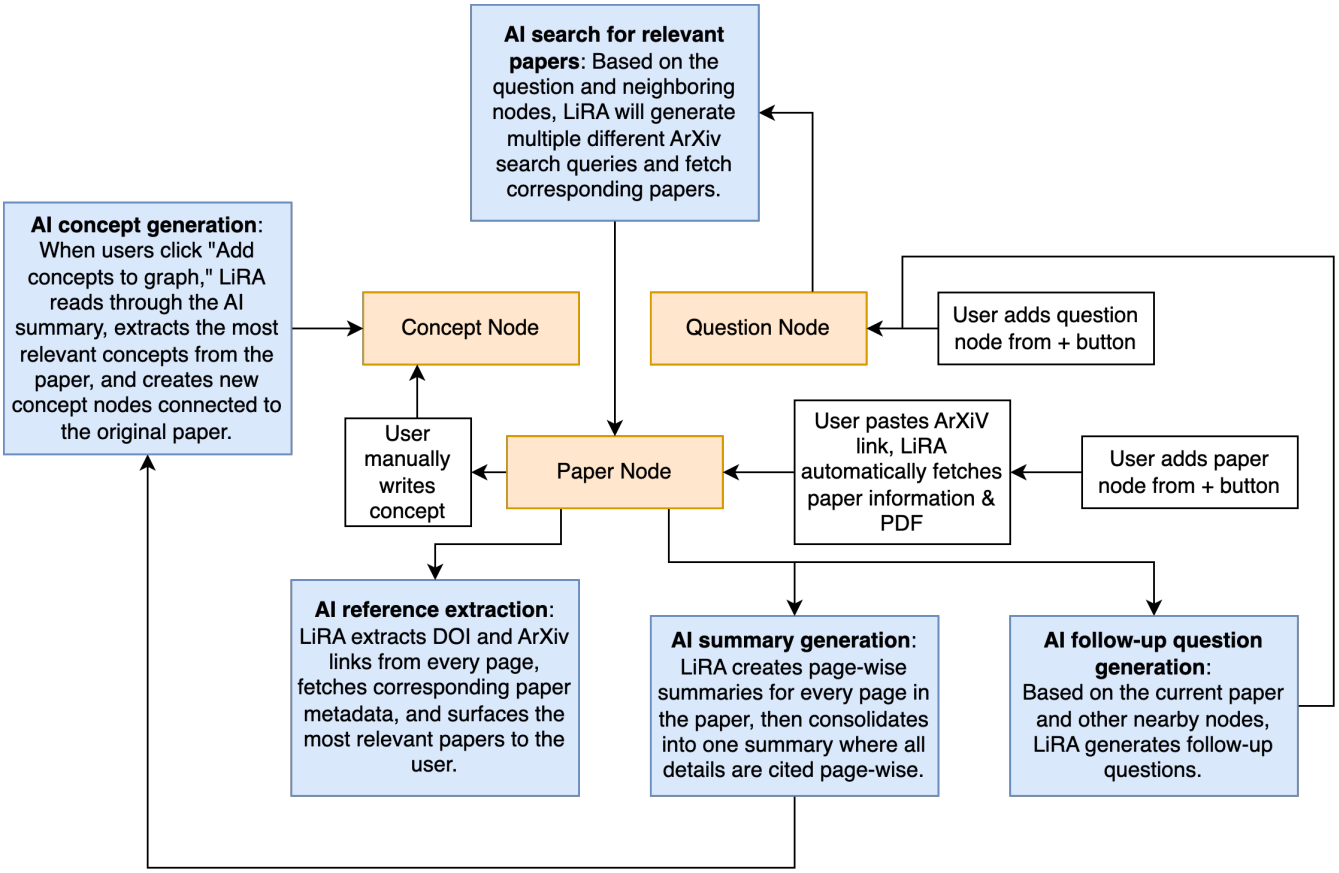

Figure 3: System architecture of LiRA. The architecture includes the backend for generative AI processing, the frontend for interactive concept graph manipulation, and integration with external data sources for real-time updates.

interest and generating actionable insights rather than spending significantly more time while navigating fragmented workflows. By the end, he reaches the same understanding of his research domain in less time and more efficiently.

4 Study

To evaluate the effectiveness of LiRA compared to traditional tools, we conducted a user study measuring performance, user experience, and learning outcomes in literature review tasks. The study compared a traditional academic search and note-taking setup (Control) with LiRA's integrated concept graph-powered interface (Treatment). Using structured tasks, surveys, and interviews, we assessed how LiRA impacts efficiency and comprehension during the literature review process.

4.1 Informed Consent

All participants were informed about the purpose and structure of the study prior to their participation. They were briefed on the tasks they would perform, the data that would be collected, and the methods of analysis. Participation was entirely voluntary, and participants could withdraw from the study at any time without providing a reason. All data was anonymized to ensure privacy and confidentiality. While Institutional Review Board (IRB) approval was not sought for this study, ethical research principles were adhered to, including transparency, voluntary participation, and data protection.

4.2 Methodology

4.2.1 Study Design. The study utilized a within-subjects crossover design, randomly assigning participants into two groups:

- **Group A:** Used the Control interface first, followed by the Treatment interface.
- **Group B:** Used the Treatment interface first, followed by the Control interface.

This design controlled for ordering effects by reversing the sequence of interface exposure for the two groups.

4.2.2 Participants. We recruited 9 participants, comprising graduate students and early-career researchers familiar with academic

tools and literature review processes. Participants were selected based on their engagement with academic research in their fields.

4.2.3 Procedure. The study consisted of the following phases:

Overview and Pre-Task Survey. Participants received an overview of the study and completed a 5-minute background survey. The survey collected information on:

- Academic and professional background.
- Familiarity with academic search tools and literature review processes.
- Frequency of tool usage and research strategies.

Research Tasks. Participants completed two structured research tasks. Each task involved exploring a specific research question, one using the Control interface (e.g., Google Scholar and Google Docs) and the other using the Treatment interface (LiRA).

Task Setup:

- Research questions addressed real-world topics:
  (1) **Task 1:** Strategies for optimizing LLMs for culturally relevant responses.
  (2) **Task 2:** Best practices for designing adaptive user interfaces using LLM-generated interactions.
- Participants were provided a starting paper to anchor their exploration and spent 5 minutes skimming the abstract and introduction.

Task Activity:

- Participants spent 15 minutes exploring academic papers, using the assigned interface to gather and synthesize information.
- They verbalized their thought process, strategies, and interaction with interface features.

Post-Task Survey. After each task, participants filled out a 5-minute survey to reflect on:

- Key insights gained from the task.
- Confidence in understanding the research question.
- Challenges encountered during the process.
- Usefulness of the assigned interface for the task.

Cumulative Feedback and Final Interview. Following both tasks, participants completed a 10-minute survey and a structured exit interview. Topics included:

- Detailed feedback on each interface.
- Preferences between the interfaces and reasons for their choices.
- Suggestions for improving LiRA's user experience and functionality.
- Clarifications or expansions on specific survey ratings.

4.2.4 Controls for Confounding Variables. To control for confounding factors, the research tasks and starting papers were kept identical for all participants. The sequence of interfaces (Control vs. Treatment) was reversed between groups to account for order effects. This approach ensured a balanced evaluation of the two interfaces.

4.2.5 Evaluation Metrics. The study focused on the following metrics:

- **Efficiency:** Time is taken to identify and synthesize relevant insights.
- **Comprehension:** Confidence and depth of understanding based on post-task surveys.
- **Usability:** Participant ratings on ease of navigation, organization, and feature utility.
- **Engagement:** Observations of participant interaction strategies and verbalized feedback.
- **Novel Contributions:** Identification of areas for potential new research contributions.

This methodology provided a robust framework for comparing LiRA's performance against traditional tools, offering valuable insights into how LiRA can enhance literature review processes and improve research workflows.

5 Results

In this section, we present the results obtained from our experiments and surveys.

5.1 Analysis of Arxiv and LiRA Through Self-Perceived Performance Scores

Table 1 presents the results of a one-sided paired t-test ($\alpha = 0.05$) comparing Arxiv and LiRA across several self-evaluated metrics related to the literature review process. Each score reflects participants' self-perceived performance on a 1–10 scale, where higher values indicate stronger agreement or satisfaction with the corresponding metric. For each metric, the table reports the mean scores for both interfaces, the mean difference (LiRA − Arxiv), the p-value, and whether the difference is statistically significant.

Overall Preference for LiRA. Participants reported significantly higher self-perceived performance using LiRA compared to Arxiv across most metrics. Notable differences included:

- **Average Recommendation:** Participants were significantly more likely to recommend LiRA (Difference = 3.4, $p = 0.0005$).
- **Clarity of Connections:** LiRA enabled significantly greater perceived clarity in connecting ideas (Difference = 5.1, $p = 0.0001$).
- **Research Efficiency:** Participants felt LiRA significantly improved their efficiency in conducting literature reviews (Difference = 4.9, $p < 0.0001$).

These results suggest that LiRA is perceived as a more effective and user-friendly tool for facilitating key aspects of the literature review process.

Ease of Navigation. LiRA scored slightly lower than Arxiv in **Ease of Navigation** (Difference = −0.7), though the difference was not statistically significant ($p = 0.1660$). This indicates that participants found both tools comparable in navigation, but LiRA may require further refinement to optimize its user interface for this aspect. In

Table 1: Results of a One-Sided Paired t-Test ($\alpha = 0.05$) Comparing Arxiv and LiRA Across Metrics

Metric	Arxiv	LiRA	Difference (LiRA - Arxiv)	P-Value	Statistical Significance
Avg. Recommendation	4.6	8.0	3.4	0.0005	YES
Avg. Confidence Software Helped	3.8	6.6	2.8	0.0012	YES
Avg. Clarity of Connections	2.1	7.2	5.1	0.0001	YES
Avg. Information Organization Quality	4.2	7.3	3.1	0.0070	YES
Avg. Research Efficiency	3.4	8.3	4.9	0.0000	YES
Avg. Ease of Navigation	8.0	7.3	-0.7	0.1660	NO
Avg. Insight Extraction Effectiveness	2.9	6.7	3.8	0.0006	YES
Avg. Comprehension Confidence Level	4.1	5.6	1.4	0.0447	YES
Avg. Research Question Understanding	4.1	5.4	1.3	0.0669	NO

the next section our qualitative survey results will reveal potential directions for user interface optimization.

Mixed Results for Comprehension and Understanding. Participants reported significant improvements in **Comprehension Confidence Level** when using LiRA (Difference = 1.4, $p = 0.0447$). However, the improvement in **Research Question Understanding** (Difference = 1.3) was not statistically significant ($p = 0.0669$). These results suggest that while LiRA aids in building confidence, it may not consistently enhance deeper understanding of research questions. This is a reasonable limitation: by design of our study, participants are only given 15 minutes to learn about a subject matter, and most participants were at least almost entirely unfamiliar with both subject domains. Given that there was not a statistically significant difference in research question understanding across usage of the control and treatment, we can extrapolate that the treatment at least maintains similar comprehension of the research domain while increasing overall research efficiency (as we reveal in our earlier discussion of other statistics from our survey).

5.2 LiRA Usage: Observations and Participant Feedback

Participants demonstrated diverse engagement with LiRA, with specific patterns emerging in their usage of key features such as concept generation, interaction with referenced papers, and summary evaluation. This section synthesizes aggregate findings from observed participant behaviors and qualitative post-task survey feedback to highlight both strengths and opportunities for improvement.

5.2.1 Concept Generation and Management. Concept generation and management were integral to participants' workflows. Across all sessions, participants engaged with both AI-generated and manually created concepts, though the level of reliance on these features varied:

- **AI-Generated Concepts:** While all participants utilized AI-generated concepts, they often found the output excessive or irrelevant. Many concepts required significant deletion or manual adjustment to align with the research context. Participants suggested the inclusion of a preview or selective addition mechanism to address this issue.
- **Manual Concept Creation:** Manual concept creation played a critical role in participants' workflows. Many participants preferred creating their own concepts, viewing them as more reliable and tailored to their needs. Participants emphasized the importance of maintaining flexibility in manual concept creation.
- **Editing and Customization:** Editing AI-generated concepts was a common activity. Participants frequently adjusted these concepts to fit their organizational preferences or research goals. They noted that improving the relevance of AI outputs would reduce the need for extensive editing.

Participants provided specific feedback to improve concept generation and management:

- Introduce a preview mechanism to allow selective addition of AI-generated concepts.
- Enable participants to directly create concepts by highlighting relevant text in summaries or papers.
- Improve the relevance of AI-generated concepts through better prompts or contextual awareness.
- Provide easier ways to connect concepts manually, such as linking related papers or concepts in the graph.

5.2.2 Interaction with Papers and Summaries. Interaction with referenced papers and summaries was another prominent aspect of LiRA usage. Participants actively explored these features to gather insights and organize information:

- **Exploration of Referenced Papers:** Most participants added referenced papers to the canvas and explored their contents. This feature was widely used to extend the breadth of their research and identify relevant connections.
- **Use of Summaries:** Summaries were a key resource for evaluating papers. Participants often relied on them to quickly assess the relevance of a paper and determine whether to explore it further. However, some participants expressed a desire for greater depth in summaries, including the integration of visuals and sentence-level exploration features.
- **Notes as an Alternative to Concepts:** While the notes feature was used by some participants as a supplementary tool for organizing ideas, its overall utility was limited. Participants often found manually created concepts to be more effective for structuring their workflows.

Feedback from participants suggests enhancements to these features:

- Expand summaries to include visual aids and more granular details.
- Add functionality for sentence-level exploration within summaries, allowing participants to focus on specific sections of interest.
- Clarify or integrate the notes feature more effectively with concept management to streamline its use.

5.2.3 Aggregate Behavioral Patterns. Certain behavioral patterns emerged across participants:

- Many participants prioritized manual concept creation over AI-generated concepts, using it as their primary method of structuring information.
- Referenced papers and their summaries were heavily utilized, often forming the foundation of participants' exploration and organization activities.
- While most participants engaged deeply with LiRA's features, a few initially struggled with navigation or understanding the purpose of certain tools, such as the notes feature or follow-up questions.

5.3 Post-Task Survey Feedback

In addition to observing participant interactions with LiRA, qualitative feedback collected through post-task surveys provided further insights into participants' experiences. This feedback highlighted both the effectiveness of LiRA and areas for enhancement.

5.3.1 Comprehension of Research Domain. One of our goals was to test the difference in comprehension of a research domain based on a participant's 15 minute interaction with the research task and assigned software. As part of their post-task surveys, participants were asked to submit a response to the following question:

"Provide a summary of your current understanding of the research question and what research has already been done based on your review. Be as detailed as possible."

Responses were graded according to the rubric in Table 2. Grading was conducted by a single evaluator with expertise in both subject matters covered in the research tasks. While this ensured a consistent and informed evaluation process, the lack of multiple graders and inter-rater reliability (IRR) is admittedly a limitation of this study. Future work should address this by involving multiple evaluators to enhance the objectivity and robustness of the grading process. Due to limited scope and the complexity of aligning multiple graders, a single expert was used in this study.

We then plotted box-and-whisker plots for each rubric item, partitioned into separate box-and-whisper plots for the control and treatment, seen in Figure 4.

The results revealed that LiRA performed comparably to the traditional Google Scholar + Google Docs setup across all rubric categories, with slight advantages observed in comprehension, critical thinking, and research direction.

For **Basic Literature Coverage**, participants using LiRA demonstrated a higher median score, reflecting its ability to help participants identify and synthesize key papers and approaches more effectively. The broader interquartile range (IQR) for LiRA suggests it supported a wider variety of participant strategies while maintaining strong upper-end performance.

For **Problem Scope**, LiRA enabled participants to develop a deeper understanding of the core challenges and considerations within the research domain. The contextual insights provided by LiRA's dynamic concept graph and AI-driven summarization likely contributed to this improved comprehension of complex relationships in the research landscape.

LiRA performed slightly better than the control in **Technical Insight**, particularly for tasks requiring deeper analysis, such as addressing regional challenges or understanding participant interaction patterns. Features like dynamic graph expansion and in-graph search appear to have helped participants identify nuanced connections between concepts.

In **Critical Thinking**, participants using LiRA were more effective at identifying gaps and proposing potential research directions compared to the control setup. LiRA's structured organization and contextual suggestions may have facilitated the development of more coherent and actionable insights.

For **Research Direction**, LiRA participants demonstrated similar or slightly better performance in suggesting next steps or refining their research focus. The system's iterative graph exploration and AI-generated follow-up questions encouraged participants to delve deeper into research possibilities.

Overall, the aggregated scores across all categories suggest that LiRA is at least as effective as the traditional tools in supporting literature reviews. While the edge over the control was not consistently large, these findings highlight that LiRA's dynamic, AI-driven features can match and occasionally exceed the capabilities of traditional approaches, providing an alternative that supports iterative and exploratory research processes.

5.3.2 Comparison to Traditional Methods. Participants frequently noted that LiRA was faster and more efficient compared to traditional tools like Google Scholar or Google Docs. Several participants emphasized the ease of identifying relevant papers and synthesizing information:

- **Speed and Efficiency:** Participants highlighted the ability to explore multiple papers and concepts quickly, which minimized the time spent searching for connections between ideas.
- **Organization:** LiRA's mind map approach offered a novel way to structure information, though some participants expressed a preference for more traditional, linear formats for certain tasks.
- **Depth of Exploration:** While LiRA was effective for broad exploration and identifying connections, some participants noted that their usual methods provided greater depth when focusing on individual papers.

5.3.3 Suggested Improvements. Participants provided specific feedback on features and functionalities they wished to see improved:

- **Enhanced Concept Management:** Participants suggested decluttering AI-generated concepts and adding features for manually connecting and organizing concepts.
- **Improved Summaries:** Many participants desired longer and more detailed summaries, with integrated visuals and the ability to clarify specific terms or sections.
- **Enhanced Graph Interactivity:** Suggestions included spatially grouping related papers in the graph and enabling secondary connections (e.g., connecting levels of related papers automatically).
- **Export and Integration:** Participants requested the ability to export their mind maps or notes into external tools like linear note-taking software.
- **Undo/Redo Functionality:** A common request was for simple undo/redo options to streamline workflows and correct errors efficiently.
- **AI Assistance:** Several participants suggested integrating a chatbot to answer follow-up questions about papers, summaries, or concepts in real-time.

6 Discussion

The findings from this study provide a comprehensive understanding of how LiRA compares to traditional literature review tools in terms of efficiency, usability, and user experience. By combining quantitative performance metrics with qualitative survey feedback and participant observations, we gain a nuanced perspective on LiRA's strengths and areas for improvement.

6.1 Synthesizing Key Insights

Our results demonstrate that LiRA significantly improves several critical aspects of the literature review process:

- **Efficiency and Organization:** LiRA's interactive concept graphs and AI-powered summarization enabled participants to explore multiple papers and synthesize insights more efficiently than traditional tools. This aligns with our quantitative results, which showed statistically significant improvements in metrics like research efficiency ($p < 0.0001$) and clarity of connections ($p = 0.0001$).
- **Broad Exploration vs. Depth:** While participants praised LiRA for its ability to provide an overarching view of research topics, qualitative feedback highlighted that traditional methods still offer advantages for deeply analyzing individual papers. This suggests that LiRA excels in facilitating broad, exploratory tasks but may need enhancements for tasks requiring deep dives into specific content.
- **Actionable Insights:** The interactive features of LiRA, such as dynamic graph expansion and in-graph searches, were particularly effective in helping participants uncover relationships between concepts. These tools allowed participants to discover novel connections, supporting the generation of actionable insights—a key challenge in traditional workflows.

Despite these strengths, our results also highlight areas where LiRA could be improved. For instance, the lack of real-time integration of visuals in summaries and the occasional irrelevance of AI-generated concepts were recurring points of critique. Addressing these issues could further enhance LiRA's effectiveness.

6.2 Implications for Literature Review Processes

LiRA's novel approach to integrating generative AI with concept graphs addresses long-standing challenges in the literature review process. Unlike traditional tools that rely on static note-taking or keyword-based searches, LiRA offers a dynamic, participant-driven exploration of research materials. This study underscores the potential for AI-powered tools to not only enhance efficiency but also transform how researchers interact with and synthesize information.

However, our findings also reveal the importance of balancing automation with user control. Participants frequently emphasized the need for features like manual concept creation, undo/redo functionality, and customizable graph layouts, reflecting a desire for tools that align with individual workflows rather than imposing rigid structures.

6.3 Broader Context and Future Directions

LiRA's ability to facilitate exploratory research and uncover relationships between concepts aligns with broader trends in human-computer interaction and AI research. The integration of direct manipulation principles with AI-powered insights represents a significant advancement in interactive systems. However, there remains an opportunity to push these capabilities further:

- Incorporating features like sentence-level exploration and real-time visualization of inter-paper relationships could make LiRA even more effective for comprehensive literature reviews.
- Expanding LiRA's dataset coverage beyond Arxiv to include other databases could enhance its applicability across disciplines.
- Exploring how LiRA can integrate with other research tools, such as citation managers or linear note-taking software, could streamline workflows and improve adoption.

Future research should also investigate how LiRA performs in larger-scale studies with diverse participant groups to better understand its impact across various academic contexts.

7 Limitations

While this study demonstrates LiRA's potential to enhance the literature review process, several limitations in the software and study design must be acknowledged. Addressing these limitations can inform future development and research, ensuring broader applicability and improved user experiences.

7.1 Software Limitations

LiRA's reliance on Arxiv as its primary data source is a notable constraint. While Arxiv provides extensive coverage for fields such

as computer science and physics, it is less comprehensive in disciplines like the social sciences or medicine, where peer-reviewed journals and other databases such as PubMed or Scopus play a more prominent role. Expanding LiRA's integration with additional data sources would significantly enhance its versatility and adoption across disciplines.

Participants also highlighted issues with the contextual relevance of AI-generated concepts. Many found these concepts excessive or irrelevant, often requiring significant manual intervention to delete or refine them. Although this feedback underscores the value of manual customization, it also suggests a need for improved contextual understanding in the AI generation pipeline. Enhancing the specificity and accuracy of generated concepts would reduce the cognitive load on participants and improve workflow efficiency.

Another recurring theme in participant feedback was the absence of real-time visuals within summaries. For research papers heavily reliant on figures, tables, or graphical data, the lack of visual integration in LiRA's summaries limited their utility. Incorporating visuals directly into summaries could improve comprehension and make the tool more appealing to users dealing with visual-heavy content.

The usability of the concept graph interface also warrants attention. Several participants noted that clutter caused by AI-generated concepts occasionally overwhelmed the interface. Introducing features such as hierarchical layouts, decluttering options, or spatial grouping of related nodes could address these challenges. Additionally, the absence of robust undo/redo functionality emerged as a common source of frustration, particularly when errors occurred during graph manipulation.

7.2 Study Design Limitations

The study design also presents several limitations that may have influenced the results. First, the sample size was relatively small, comprising nine participants, primarily graduate students and early-career researchers. While this group provided valuable insights, a larger and more diverse sample would be necessary to generalize findings to broader academic audiences. Including researchers from varied disciplines and career stages could reveal additional use cases and challenges.

Second, the 15-minute task duration may not fully capture the complexities of real-world literature reviews, which often span days or weeks. Participants noted that the limited time required them to focus on immediate usability rather than assessing LiRA's performance in longer, more iterative workflows. Future studies should explore longer task durations to better simulate authentic literature review scenarios.

The choice of research domains—machine learning and human-computer interaction—also limits the generalizability of findings. These topics are highly relevant to LiRA's design focus but may not reflect the diverse challenges faced by researchers in fields like biology, sociology, or history. Expanding the scope of research tasks in future studies would ensure a more comprehensive evaluation.

Finally, while the within-subjects crossover design minimized order effects, participants' familiarity with traditional tools (e.g., Google Scholar) may have introduced bias. This familiarity could have influenced their efficiency and confidence when using the control interface, potentially affecting the comparison with LiRA.

Additionally, the grading of post-task survey responses was conducted by a single evaluator with expertise in the subject matter. While this ensured consistency, the absence of multiple graders and inter-rater reliability (IRR) is a limitation. Because our study represents a scenario where we are using pre-defined codes in the form of rubric items, aim to describe results quantitatively, and enable replicability, a replication of this study would best be augmented with IRR [9]. Future studies should involve multiple evaluators to enhance the robustness and objectivity of the grading process.

7.3 Future Directions

To address interface limitations, future implementations of LiRA should consider:

- Recruiting a larger and more diverse participant pool to capture varied academic contexts and workflows.
- Designing tasks with varying durations and complexities to better reflect the iterative nature of literature reviews.
- Expanding LiRA's integration with additional databases such as PubMed, IEEE Xplore, and Scopus to improve its applicability across disciplines.

Iterative design cycles incorporating user feedback should also remain central to LiRA's development. For example, enhancing the AI generation pipeline for concepts and integrating visuals into summaries were common participant requests that could significantly improve the tool's usability. Furthermore, enabling seamless integration with external tools, such as citation managers like Zotero or linear note-taking software like Obsidian, would make LiRA a more holistic solution for researchers.

8 Conclusion

This study introduced LiRA, a generative AI-powered interface designed to enhance the efficiency, organization, and synthesis capabilities of the literature review process. By integrating AI-driven summarization, dynamic concept mapping, and direct manipulation principles, LiRA offers an interactive alternative to traditional literature review.

Through a participant study comparing LiRA to traditional tools, we demonstrated that LiRA performs comparably, with modest improvements in areas such as comprehension, critical thinking, and research direction. These findings suggest that LiRA's interactive features and iterative workflows effectively support researchers in navigating complex research domains.

While the study highlighted the potential of LiRA, it also underscored several limitations in both the software and study design, including a small sample size and the use of a single evaluator for post-task surveys. Future work should address these limitations by expanding participant diversity, exploring broader research domains, and incorporating more robust evaluation methods.

Overall, this work provides a foundation for future exploration of AI-augmented, direct manipulation tools for academic research, bridging gaps in traditional literature review workflows and paving the way for more dynamic and efficient research processes.

References

[1] Jan vom Brocke, Alexander Simons, Björn Niehaves, Kai Riemer, Ralf Plattfaut, and Anne Cleven. 2009. Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. *http://www.alexandria.unisg.ch/Publikationen/67910.*

[2] Alberto J. Cañas, Roger Carff, Greg Hill, Marco Carvalho, Marco Arguedas, Thomas C. Eskridge, James Lott, and Rodrigo Carvajal. 2005. *Concept Maps: Integrating Knowledge and Information Visualization.* Springer Berlin Heidelberg, Berlin, Heidelberg, 205–219. https://doi.org/10.1007/11510154_11

[3] Alberto J Cañas, Greg Hill, Roger Carff, Niranjan Suri, James Lott, Gloria Gómez, Thomas C Eskridge, Mario Arroyo, and Rodrigo Carvajal. 2004. CmapTools: A knowledge modeling and sharing environment. (2004).

[4] Catherine Chavula, Yujin Choi, and Soo Young Rieh. 2023. SearchIdea: An Idea Generation Tool to Support Creativity in Academic Search. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval.* ACM, Austin TX USA, 161–171. https://doi.org/10.1145/3576840.3578294

[5] Yuka Egusa, Masao Takaku, and Hitomi Saito. 2014. How concept maps change if a user does search or not?. In *Proceedings of the 5th Information Interaction in Context Symposium* (Regensburg, Germany) *(IIiX '14).* Association for Computing Machinery, New York, NY, USA, 68–75. https://doi.org/10.1145/2637002.2637012

[6] Michael Haman and Milan Školník. 2024. Using ChatGPT to conduct a literature review. *Accountability in Research* 31, 8 (2024), 1244–1246. https://doi.org/10.1080/08989621.2023.2185514 arXiv:https://doi.org/10.1080/08989621.2023.2185514 PMID: 36879536.

[7] Mayank Kejriwal. 2022. Knowledge Graphs: A Practical Review of the Research Landscape. *Information* 13, 4 (2022). https://doi.org/10.3390/info13040161

[8] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. 2024. DirectGPT: A Direct Manipulation Interface to Interact with Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24).* ACM, 1–16. https://doi.org/10.1145/3613904.3642462

[9] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (Nov. 2019), 23 pages. https://doi.org/10.1145/3359174

[10] Christoph Müller-Bloch and Johann Kranz. [n. d.]. A Framework for Rigorously Identifying Research Gaps in Qualitative Literature Reviews. ([n. d.]). https://core.ac.uk/outputs/301367526/?source=oai

[11] Allard Oelen. 2022. Leveraging human-computer interaction and crowdsourcing for scholarly knowledge graph creation. (Nov. 2022). https://doi.org/10.15488/13066

[12] Guy Paré, Marie-Claude Trudel, Mirou Jaana, and Spyros Kitsiou. 2015. Synthesizing information systems knowledge: A typology of literature reviews. *Information Management* 52, 2 (2015), 183–199. https://doi.org/10.1016/j.im.2014.08.008

[13] Scott R. Rosas and John W. Ridings. 2017. The use of concept mapping in measurement development and evaluation: Application and future directions. *Evaluation and Program Planning* 60 (2017), 265–276. https://doi.org/10.1016/j.evalprogplan.2016.08.016

[14] Nada Sahlab, Hesham Kahoul, Nasser Jazdi, and Michael Weyrich. 2022. A Knowledge Graph-Based Method for Automating Systematic Literature Reviews. https://doi.org/10.48550/arXiv.2208.02334

[15] Benjamin Sturm and Ali Sunyaev. 2019. Design Principles for Systematic Search Systems: A Holistic Synthesis of a Rigorous Multi-cycle Design Science Research Journey. *Business Information Systems Engineering* 61, 1 (Feb. 2019), 91–111. https://doi.org/10.1007/s12599-018-0569-6

[16] Mathieu Templier and Guy Paré. 2018. Transparency in literature reviews: an assessment of reporting practices across review types and genres in top IS journals. *European Journal of Information Systems* 27, 5 (2018), 503–550. https://doi.org/10.1080/0960085X.2017.1398880 arXiv:https://doi.org/10.1080/0960085X.2017.1398880

[17] Gerit Wagner, Roman Lukyanenko, and Guy Paré. 2022. Artificial intelligence and the conduct of literature reviews. *Journal of Information Technology* 37, 2 (2022), 209–226. https://doi.org/10.1177/02683962211048201 arXiv:https://doi.org/10.1177/02683962211048201

Appendix

The appendix below contains supplementary materials to support the main text, including pilot study results, additional methodological details, evaluation rubrics, and visualizations of user study results. These materials provide further context for the study design, grading criteria, and performance comparisons between LiRA and traditional tools. The source code can be found at this GitHub link.

Received 20 November 2024; revised XXXX; accepted XXXX

Table 2: (Post-Task Survey) Evaluation Criteria for Research Understanding and Analysis

Subcategory	Scoring Criteria
Basic Literature Coverage (20 points)	• **20 points**: Identifies 2-3 key papers/approaches and their main findings. • **15 points**: Mentions 1-2 relevant papers/approaches. • **10 points**: General awareness of existing work without specific examples. • **5 points**: Very limited awareness of existing work.
Problem Scope (20 points)	• **20 points**: Clear understanding of core challenges and key considerations. • **15 points**: Good grasp of main challenges. • **10 points**: Basic understanding of the problem. • **5 points**: Unclear problem understanding.
Technical Insight (20 points)	• **For Task 1:** – **20 points**: Identifies key aspects (e.g., region detection, cultural embeddings, bias). – **15 points**: Understands some technical aspects. – **10 points**: Basic technical awareness. – **5 points**: Minimal technical understanding. • **For Task 2:** – **20 points**: Identifies key aspects (e.g., user modeling, personalization, interaction patterns). – **15 points**: Understands some technical aspects. – **10 points**: Basic technical awareness. – **5 points**: Minimal technical understanding.
Critical Thinking (20 points)	• **20 points**: Identifies gaps and potential research directions. • **15 points**: Some analysis of limitations or opportunities. • **10 points**: Basic critical thinking. • **5 points**: Minimal analysis.
Research Direction (20 points)	• **20 points**: Clear suggestion for next steps or research focus. • **15 points**: General idea of potential direction. • **10 points**: Vague suggestions. • **5 points**: No clear direction.

Key Elements to Look For: **For Task 1, assess understanding of regional computing challenges, bias in language models, and privacy/ethical implications. For the Task 2, evaluate awareness of personalization basics, user interaction patterns, and LLM capabilities.**

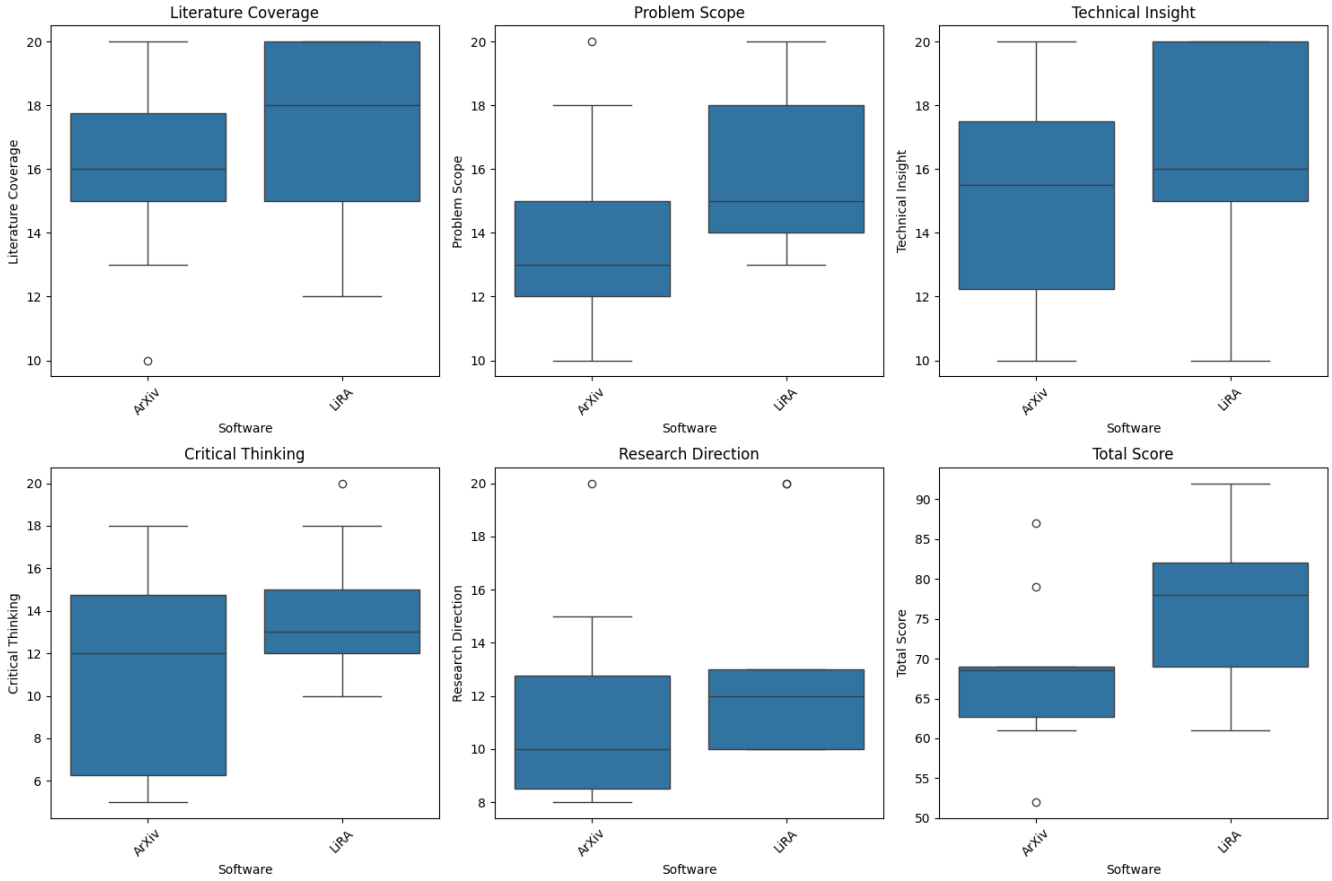

Figure 4: (Post-Task Survey) User research comprehension results. The figure illustrates differences in users' understanding of the research domain based on their interaction with assigned software during the study.

Table 3: (Pilot Study) Aggregate Summary of Research Tools Used by Pilot Study Participants

Search Tool	Frequency
Google Scholar	15
Library Catalogues/Databases	8
Hollis	6
PubMed	5
ResearchGate	4
JSTOR	2
Elicit.ai	2
ArXiv	1
Semantic Scholar	1
ACM Digital Library (ACM DL)	1
SciSpace	1
Exa AI	1
ResearchRabbit	1

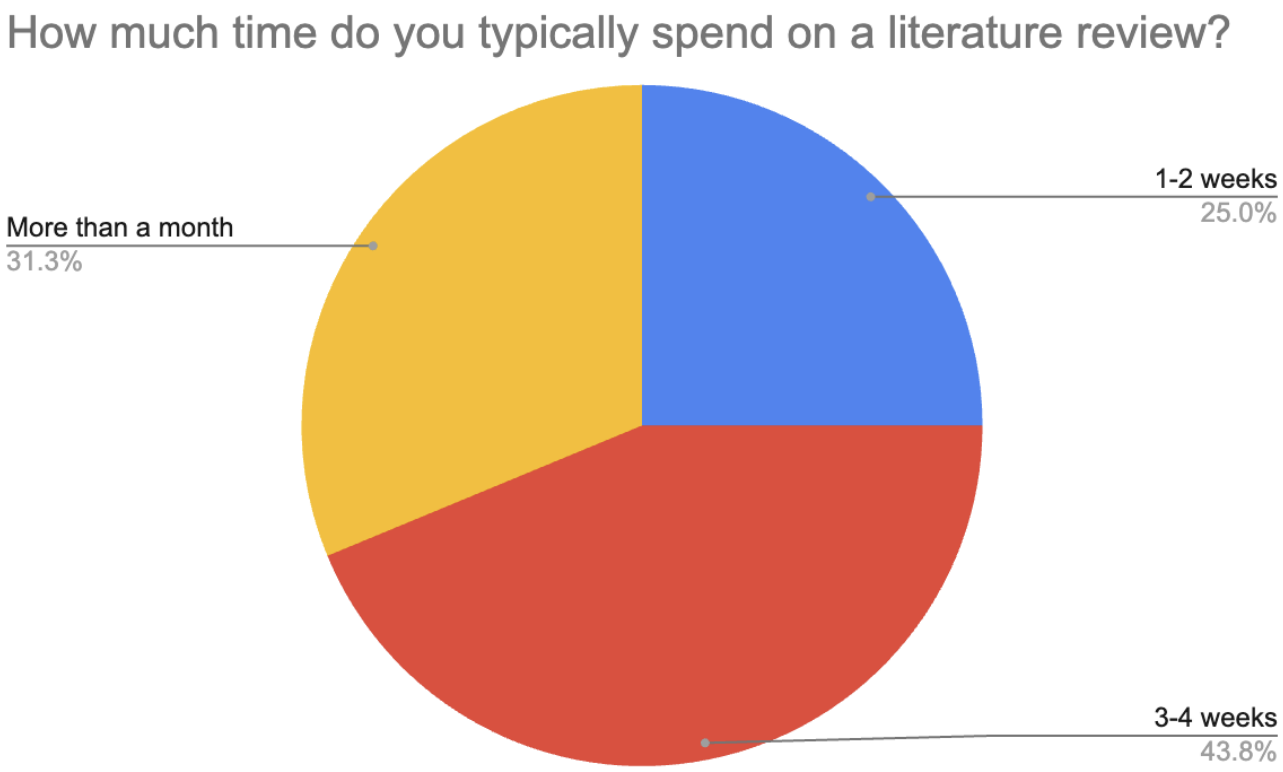

Figure 5: (Pilot Study) Responses to the question, "How much time do you typically spend on a literature review?"

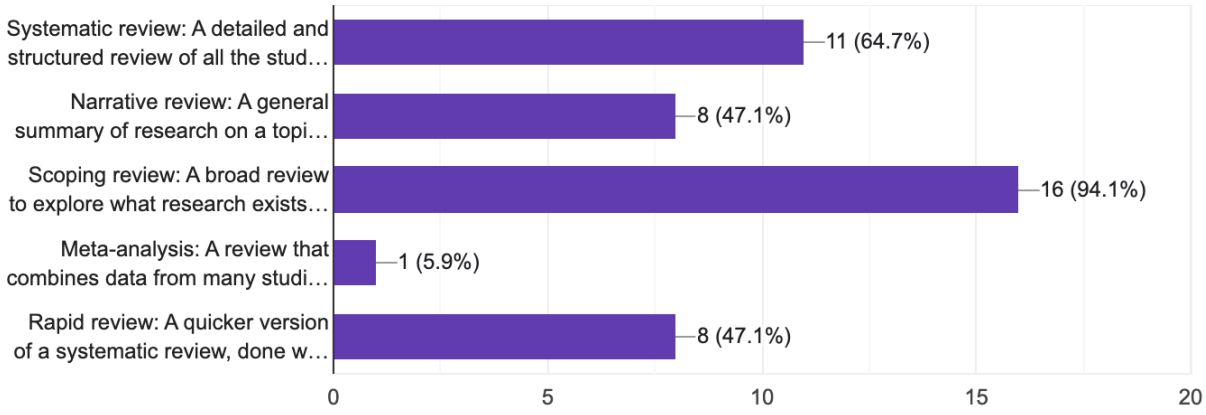

Figure 6: (Pilot Study) Responses to the question, "What types of literature reviews do you typically conduct for your research?"

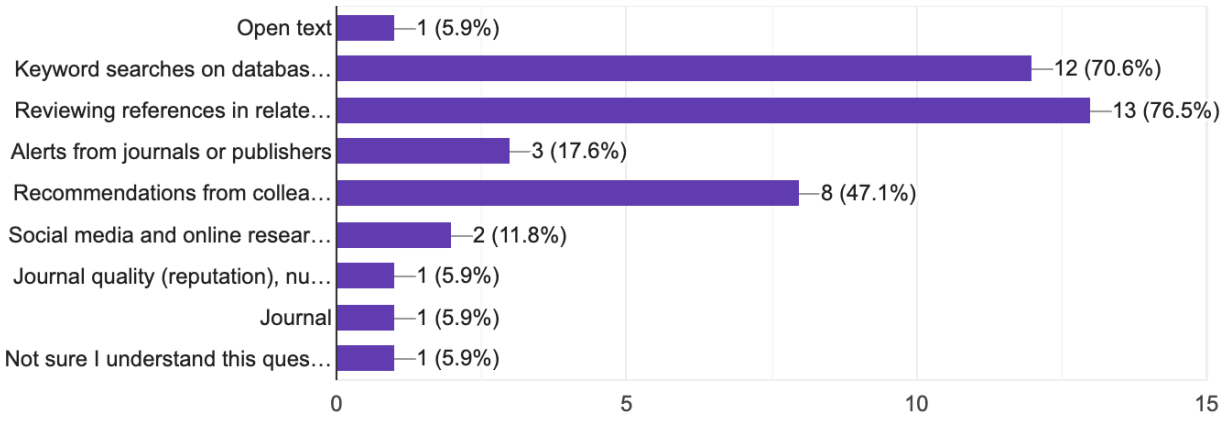

Figure 7: (Pilot Study) Responses to the question, "How do you determine the quality and relevance of the literature you find? (e.g., impact factor, number of citations, author credibility)."

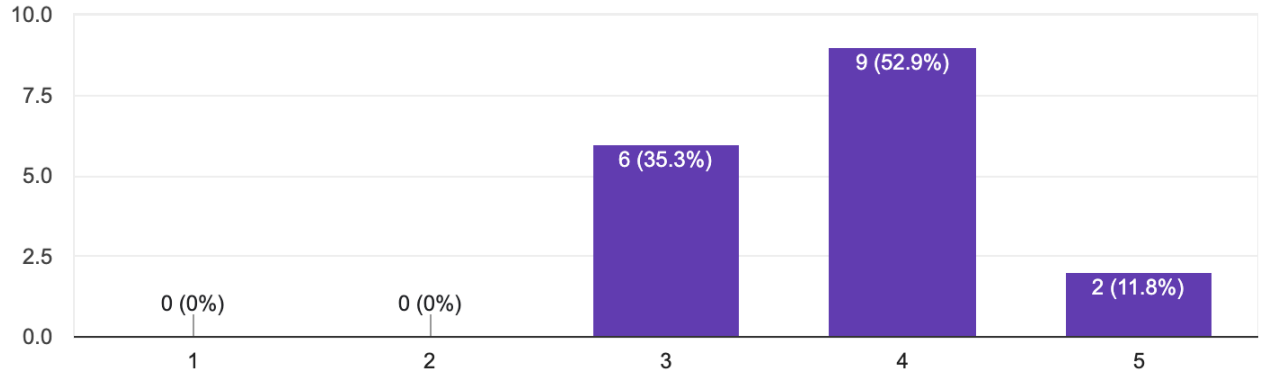

Figure 8: (Pilot Study) Responses to the question, "How confident are you in your ability to find comprehensive and relevant literature?"

Table 4: (Pilot Study) Aggregate Summary of Search Processes Used by Pilot Study Participants During Literature Reviews

Process	Frequency
Use keywords for searching papers	9
Follow references in papers	6
Explore papers that cite a relevant work	5
Save papers to tools like Zotero for further analysis	3
Start with existing reviews on the topic	3
Iterate on search terms (refine based on findings)	3
Open multiple papers in tabs and browse abstracts	2
Ask colleagues or GPT for recommendations	2
Systematic reviews with formal annotation processes	1
Prioritize high-impact journals or annual reviews	1
Break research into subquestions for targeted searches	1

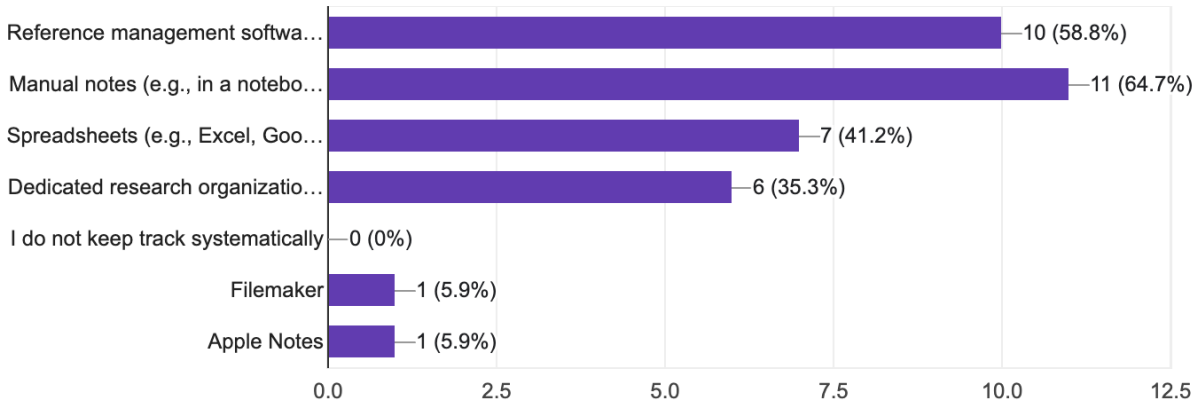

Figure 9: (Pilot Study) Responses to the question, "What kind of note-taking software do you use, if any?"

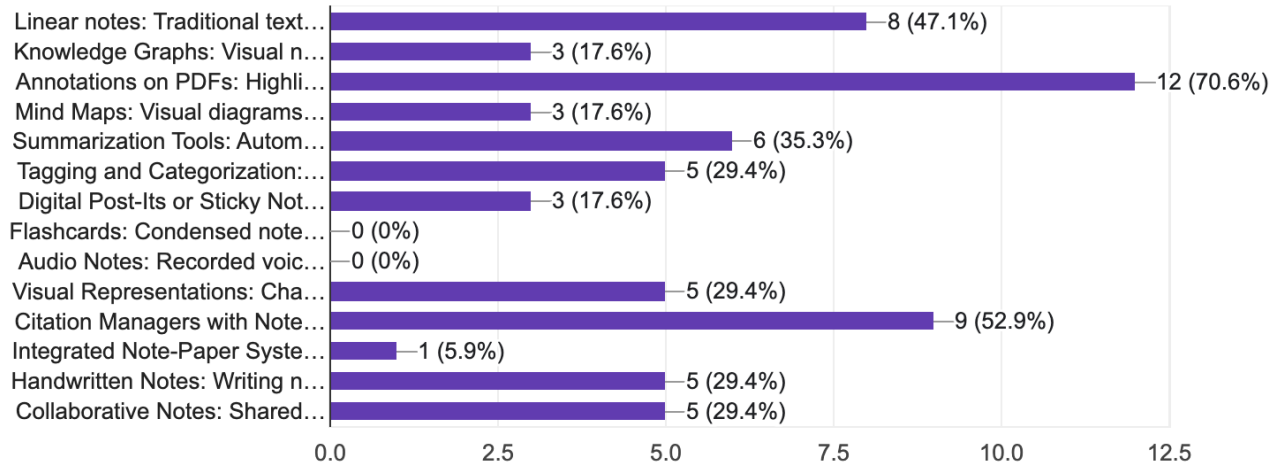

Figure 10: (Pilot Study) Responses to the question, "What kinds of note-taking modalities do you find to be the most useful?"

Table 5: (Pilot Study) Aggregate Summary of How Software Tools Are Used in the Research Process

Usage	Frequency
Organizing references (e.g., Zotero, EndNote)	10
Taking notes on papers (e.g., Notion, Obsidian, Word)	9
Searching for literature	7
Annotating papers (e.g., creating summaries, reflections)	6
Collaboration (e.g., Google Docs, spreadsheets)	4
Tracking citations	3
Brainstorming new research ideas	2
Maintaining annotated bibliographies	2
Using spreadsheets for formal reviews	2

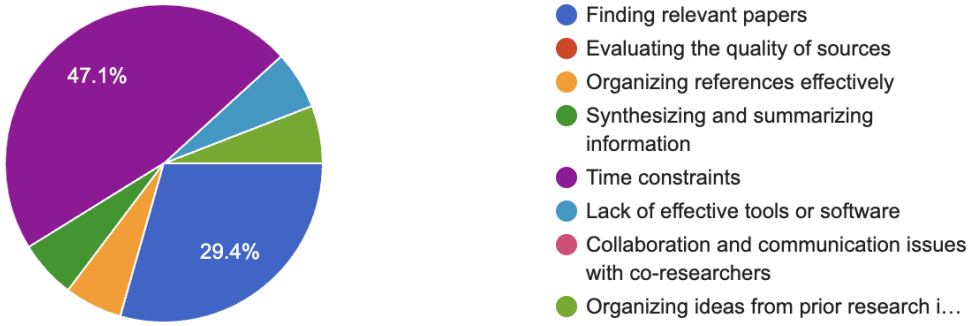

Figure 11: (Pilot Study) Responses to the question, "What is the main bottleneck in your current literature review process?"

Table 6: (Pilot Study) Aggregate Summary of Desired Improvements to the Literature Review Process

Desired Improvement	Frequency
Improved organization of ideas, papers, and notes	6
Time efficiency (e.g., faster processes)	5
Accurate and comprehensive summaries of papers	4
Enhanced search tools (e.g., filtering by methods, finding influential papers)	4
Support for interdisciplinary or question-driven research	3
AI-generated structured literature surveys	3
Better access to non-English or international research	2
Long-term storage and systematic retrieval of previous reviews	2
Ability to critically evaluate and synthesize arguments	2
Accurate related text or citation recommendations	1