

DiscipLink: Unfolding Interdisciplinary Information Seeking Process via Human-AI Co-Exploration

Chengbo Zheng
cb.zheng@connect.ust.hk

Hong Kong University of Science and
Technology
Hong Kong, China

Yuanhao Zhang
yzhangiy@connect.ust.hk

Hong Kong University of Science and
Technology
Hong Kong, China

Zeyu Hunag
zhuangbi@connect.ust.hk

Hong Kong University of Science and
Technology
Hong Kong, China

Chuhan Shi
chuhanshi@seu.edu.cn

Southeast University
Nanjing, China

Minrui Xu
mxubh@connect.ust.hk

Hong Kong University of Science and
Technology
Hong Kong, China

Xiaojuan Ma
mxj@cse.ust.hk

Hong Kong University of Science and
Technology
Hong Kong, China

ABSTRACT

Interdisciplinary studies often require researchers to explore literature in diverse branches of knowledge. Yet, navigating through the highly scattered knowledge from unfamiliar disciplines poses a significant challenge. In this paper, we introduce DiscipLink, a novel interactive system that facilitates collaboration between researchers and large language models (LLMs) in interdisciplinary information seeking (IIS). Based on users' topic of interest, DiscipLink initiates exploratory questions from the perspectives of possible relevant fields of study, and users can further tailor these questions. DiscipLink then supports users in searching and screening papers under selected questions by automatically expanding queries with disciplinary-specific terminologies, extracting themes from retrieved papers, and highlighting the connections between papers and questions. Our evaluation, comprising a within-subject comparative experiment and an open-ended exploratory study, reveals that DiscipLink can effectively support researchers in breaking down disciplinary boundaries and integrating scattered knowledge in diverse fields. The findings underscore the potential of LLM-powered tools in fostering information-seeking practices and bolstering interdisciplinary research.

CCS CONCEPTS

- **Information systems** → **Users and interactive retrieval**; • **Human-centered computing** → **Interactive systems and tools**; • **Computing methodologies** → *Artificial intelligence*.

KEYWORDS

Information Retrieval, Human-AI Collaboration, Interdisciplinary Research

ACM Reference Format:

Chengbo Zheng, Yuanhao Zhang, Zeyu Hunag, Chuhan Shi, Minrui Xu, and Xiaojuan Ma. 2024. DiscipLink: Unfolding Interdisciplinary Information

Seeking Process via Human-AI Co-Exploration. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*, October 13–16, 2024, Pittsburgh, PA, USA. ACM, New York, NY, USA, 20 pages. <https://doi.org/10.1145/3654777.3676366>

1 INTRODUCTION

The knowledge of humankind is classified into disciplines, and many researchers are trained and specialize in a single discipline [1]. However, there is a growing trend towards adopting interdisciplinary approaches to research, especially for complex, real-world problems [72]. Although forming research teams composed of experts from various fields is one preferred way of carrying out interdisciplinary studies, resource constraints and budget limitations sometimes make it impractical to establish such teams [9]. Researchers trained in one discipline often have to independently seek out information from the literature from various, maybe unfamiliar, research fields to develop their project ideas and approaches, a process we refer to as interdisciplinary information seeking (IIS) in this paper. However, this process is challenging for several reasons. First, relevant literature to an interdisciplinary project is often **highly scattered** in various weak-connected research areas [6, 47]. It often requires multiple iterations for researchers to identify *what to search* [20, 52]. Second, different disciplines use **different terminologies**, which becomes an obstacle for researchers to form accurate queries for searching literature in their unfamiliar fields [20, 48, 66]. Also, interdisciplinary researchers face the challenge of interpreting **vast amounts of information** from various unfamiliar fields and **evaluating the relevance** to their interests during the search [20, 52].

While prior human-computer interaction (HCI) work has proposed various approaches for seeking scholarly information, few of them are targeted at the needs of IIS. For example, when a computer science researcher is looking to find existing work on computational approaches to combat misinformation, existing literature search tools can facilitate the search by expanding on citations and references [16, 29, 33], iterating search keywords [17], or tracking key actor activities [32]. However, when the researcher is curious about what knowledge in sociology and education, in particular, can be blended into research on this topic, these tools provide limited guidance within the vast search space.

UIST '24, October 13–16, 2024, Pittsburgh, PA, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*, October 13–16, 2024, Pittsburgh, PA, USA, <https://doi.org/10.1145/3654777.3676366>.

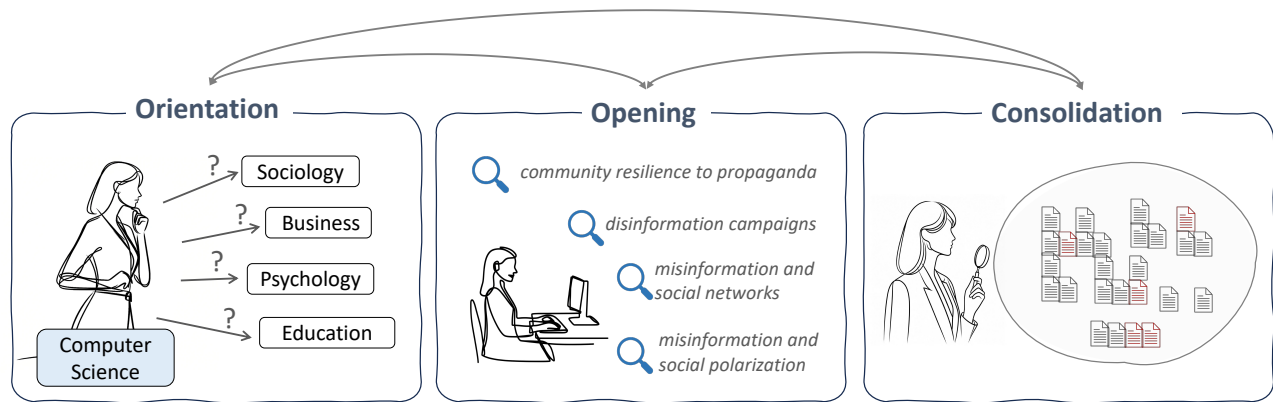


Figure 1: The nonlinear model of IIS behavior includes three components [20]. Researchers do not go through these components with a certain order, but intertwine the activities in these components during IIS.

In this paper, we aim to leverage the recent breakthroughs in Large Language Models (LLMs) to tackle the challenges of IIS. LLMs, such as GPT4¹ and Gemini², are pre-trained on vast amounts of internet-wide text data, enabling them to embed the knowledge from various academic fields and can potentially tackle challenges mentioned above [10, 39]. Various LLM-powered literature search tools have emerged, such as Consensus³, Perplexity⁴, and SciSpace⁵. These tools generally retrieve relevant papers based on user questions and provide answers grounded on the retrieved papers, which lowers the risks of hallucination [38]. However, current LLM tools might still not be a perfect solution for IIS for several reasons. First, interdisciplinary projects often target “ill-structured problem space” [52]. As a result, the search process is by nature exploratory, and the intents of researchers will change along with the search process [52]. Automating the search with LLMs does not account for the shifts and leaves users with the burden of iterating exploration. Second, existing workflows allocate the paper screening task entirely to LLMs. However, the paper screening process is critical as interdisciplinary researchers need to make sense of outside domains [52] continuously, have serendipitous discovery [20], and build cognitive contexts [20, 71] from the process. Moreover, it is found that the information seeking by LLMs might be biased, creating a “generative echo chamber” [60]. Thus, it is critical to engage researchers in the exploration process instead of solely relying on the responses provided by LLMs.

To better assist researchers in IIS with LLM, we design and develop DiscipLink, a human-centered AI system to *co-explore* literature in various disciplines with users. DiscipLink integrates a mixed-initiative [26] and human-AI collaboration [63, 83] workflow. With users’ topic of interest as input, DiscipLink first generates *exploratory questions* (EQs) around the topic from the perspectives of various relevant disciplines. In this process, LLMs simulate domain experts from specific fields to provide EQs. Users can tailor the EQs or elicit DiscipLink to generate more EQs by either making

direct requests or sharing papers they find interesting, thus guiding the system to uncover new directions for exploration. Besides, when users want to delve into any specific EQs, DiscipLink surfaces the existing knowledge about the EQs for users. Specifically, we design an LLM-based query formulation strategy by including discipline-specific terminologies in the queries to extensively search papers around the EQs while maintaining relevance to the intended field(s). After retrieving papers from queries, DiscipLink extracts key themes that are related to users’ topics of interest as well as selected EQs to facilitate users in finding answers for the EQs. Additionally, DiscipLink annotates returned papers by highlighting their connection with users’ exploration focus to facilitate users in efficiently screening papers.

We evaluated DiscipLink through two complementary studies. First, we tested the usability of DiscipLink with 12 graduate students by comparing our system with a baseline condition on searching papers for specific interdisciplinary research topics. The quantitative findings indicate that DiscipLink enabled participants to complete their tasks more efficiently and to uncover a more comprehensive range of knowledge. In the second study, we invited seven experienced interdisciplinary researchers to use DiscipLink for literature exploration in their own projects. The qualitative feedback from participants highlighted their appreciation for DiscipLink’s co-exploration workflow. Participants also identified potential limitations of DiscipLink, particularly in meeting the unique needs of individual researchers.

In summary, the contribution of this paper is threefold:

- DiscipLink, an interactive system featuring a human-AI co-exploration workflow, empowering researchers in interdisciplinary information seeking.
- A comparative experiment and an open-ended exploratory study to evaluate the effectiveness of DiscipLink.
- Design implications derived from our design process and the user studies on how to support interdisciplinary researchers with LLMs.

¹<https://openai.com/gpt-4>

²<https://deepmind.google/technologies/gemini>

³<https://consensus.app/search/>

⁴<https://www.perplexity.ai/>

⁵<https://scispace.com/>

2 RELATED WORK

In this section, we first introduce the key processes and challenges of interdisciplinary information seeking. Then, we discuss the existing systems that support the literature review. Lastly, we discuss existing human-AI collaboration research on information-seeking and sensemaking.

2.1 Interdisciplinary Information Seeking (IIS)

Interdisciplinary research involves integrating knowledge from multiple disciplines to address research problems [18, 72]. A critical component of this type of research is interdisciplinary information seeking (IIS), the process of gathering relevant information across various fields. Compared with normal scholarly information seeking, IIS is typically more complex and time-consuming due to its highly dynamic process [20, 66].

Foster [20] describe IIS through a nonlinear model comprising three main stages: *orientation*, *opening*, and *consolidation*, which are often intertwined during the IIS process (see Figure 1). *Orientation* involves researchers determining their search direction. The key challenge in *orientation* is **the highly-scattered nature of the necessary knowledge for an interdisciplinary project** [6, 20, 48, 52]. Researchers aim to construct an interdisciplinary knowledge picture for their projects, yet they cannot be experts in every field. Consequently, they often seek out resources, including experts, to discern relevant knowledge [34]. In the absence of such guidance, they have to rely on their own general knowledge and intuition to determine the disciplines from which to gather information [20, 52]. This approach, however, introduces considerable uncertainty into their search process. Furthermore, researchers often need to revisit their search goals based on insights from the *opening* and *consolidation* stages, adding to the time required for IIS [20].

Opening is the stage where researchers explore literature in their chosen search directions. Interdisciplinary researchers tend to *deliberately increase the breadth of their search* and maintain an open-minded attitude towards all encountered information, considering it potentially valuable [20]. This approach is driven by the necessity to gather widely dispersed knowledge inherent to IIS. A significant challenge in this stage is the **terminology gap**: researchers often struggle to craft queries that use appropriate terminologies to effectively bridge their research interests with literature from unfamiliar disciplines [6, 52]. Previous work suggests that terminology may “have a different meaning” [48], and “not always appropriate or transferable” [20] across disciplines. Researchers might not be able to immediately find the papers for the search direction they identified in *orientation* due to the terminology gap. Thus, they typically need to gradually accumulate their knowledge in unfamiliar fields and iterative refining search terms to uncover relevant literature [66].

Consolidation is the stage where researchers decide whether and how to include found papers in their projects, typically following the *Opening* and *Orientation* stages. Two primary challenges emerge during consolidation. First, the *Opening* stage results in the accumulation of a large volume of information, much of which may be irrelevant noise, thereby imposing a significant **information load** on researchers [48, 52]. Second, the found papers have their own contexts in their belonging disciplines, and to sift, verify, and

incorporate the found papers, the interdisciplinary researcher must **make sense of the connection between the found papers and their own project**. This task is challenging for researchers who are not accustomed to reading literature from unfamiliar fields [52, 66].

In summary, IIS is challenging because of the highly scattered relevant knowledge, terminology gap, information load, and the difficulty in identifying relevant papers. These challenges identified by prior work motivate our system design (see section 3).

2.2 Interactive Systems Supporting Literature Review

Getting a comprehensive understanding of prior works is generally challenging for researchers [11, 56]. Although survey papers synthesizing previous works into themes are commonly used by researchers to grasp fields [27], it is uncertain whether existing surveys can effectively cover all fields and capture rapidly emerging research topics. To overcome the obstacles associated with literature reviews, some studies have aimed to simplify the process of digesting research papers [15, 24], while others have focused on aiding researchers in compiling relevant literature through mechanisms like citation networks [29, 30, 70] and identification of key authors [32].

Specifically, Kang et al. [33] outline two primary strategies employed by prior interactive tools to assist researchers in synthesizing literature reviews from extensive bodies of work: the *bottom-up approach*, which supports researchers in developing key themes through improved navigation at the paper level, as seen in tools like Threddy [29] and Apollo [16], and the *top-down approach*, which offers an overarching view of the literature, exemplified by Metro Maps [59]. Building on these concepts, Kang et al. [33] propose for a mixed-initiative approach that combines clustering papers via citation networks to provide an overview for researchers while also allowing them to customize this overview from the bottom up. Inspired by Kang et al. [33], our system incorporates a mixed-initiative workflow once a collection of relevant papers has been assembled. But differently, our design focuses on a more expansive search process across various disciplines to meet the unique demands of IIS. Therefore, the system we propose takes users’ descriptions of their research interests as input, instead of using seed papers as in Kang et al. [33]’s work. Moreover, we specifically tailor the discovery and synthesis workflow to preserve the link between discovered papers and the user’s contexts, considering the potential divergence between them in IIS.

Previous studies have also endeavored to enable researchers to uncover relevant knowledge from other disciplines, such as developing analogical literature search engines that facilitate the discovery of papers from distant fields sharing similar research aims [13, 31], directions or challenges [35]. Nevertheless, the analogical search has a clear goal in performing literature-based discovery [64, 68, 69] to uncover solutions that could be transferred for solving the target research problem Kang et al. [31]. Differently, our work targets a more expansive search process (i.e., IIS), and focuses on supporting researchers oriented to relevant research in various disciplines. Additionally, Shi et al. [62] introduce a visual analytics system designed to automatically collate and synthesize research data from several disciplines, aiding in the selection of research directions

in experimental medicinal chemistry. While their visual analytics method is tailored to a specific interdisciplinary domain, our study explores the broader potential of LLMs to facilitate the general processes involved in IIS.

2.3 AI-powered Information Seeking and Sensemaking

Numerous initiatives have been undertaken to develop AI systems aimed at enhancing information-seeking and sensemaking activities. For instance, Qiu et al. [55] introduced DocFlow, a visual analytics system that facilitates document retrieval and categorization through natural language questions within the biomedical field. Palani et al. [51] design in-context positioning of query suggestions and found it improves the exploratory search process.

The advent of large language models (LLMs) has positioned them as potential knowledge bases for information retrieval [2, 81]. There is a growing body of research focused on refining the user experience in sourcing information from LLMs. For example, Suh et al. [67] argue that the typical linear conversational workflow found in many LLM-based tools falls short in aiding users to synthesize and comprehend information generated by LLMs. To address this, they advocate for the adoption of interactive hierarchical views to enrich information interaction practices. Jiang et al. [28] present a novel approach by organizing LLM responses into interactive graph-based diagrams, thereby streamlining the process for users to pinpoint the information they need.

While LLMs in the above works mainly serve by directly generating answers for information seekers, hallucination remains a significant issue. To address this, some research focuses on using LLMs to extract and summarize information from search results. Selenite, an LLM-powered system recently proposed by Liu et al. [40], employs LLMs to generate comprehensive overviews of options and criteria grounded in search results, guiding users through complex decisions. Recent emerging works also see opportunities to leverage potential hallucinated LLM generations. For example, Gao et al. [22] use LLMs to generate “hypothetical documents” for input queries to improve dense retrieval. Similarly, Son et al. [65] utilize LLM generation for visual search, but to help users better express their intents and refine their searches. Lee et al. [37] connect retrieved papers with users’ existing collections using LLMs, which may include hallucinations, to aid in sensemaking. They found that in their design, hallucinations had minimal negative impacts. In our work, we also explore how potentially hallucinated generations can support information seeking. We use LLMs to produce high-level exploratory questions, instead of facts, from various disciplinary perspectives. These questions enable users to co-explore with LLMs. Additionally, we leverage LLMs to aggregate discipline-specific terminology by generating pseudo-answers, thereby expanding search queries with relevant terminologies to help users bridge terminology gaps in IIS.

3 DESIGN GOALS

Drawing on insights from prior research on interdisciplinary information seeking (refer to subsection 2.1) and reflections from our design iterations (detailed in the supplementary material), we

have identified four design goals (DG) essential for designing a human-AI collaborative system to aid in IIS.

DG1 Support orienting the exploration in various disciplines.

Engaging in interdisciplinary research often necessitates acquiring knowledge from various disciplines, which challenges researchers in determining relevant search areas in unfamiliar fields [6, 20, 48, 52]. Although LLMs can assist researchers in delving into topics across various disciplines, their tendency to generate hallucinations may lead to confusion. Additionally, prior studies [20, 58] and our preliminary design experiments underscore the importance of maintaining high self-agency for researchers who prefer to be actively involved in the exploration process. Accordingly, the system should not solely provide answers for IIS with LLMs but should utilize LLMs as a knowledge resource, guiding users in their exploration and enabling a collaborative exploration experience.

DG2 Support formulating search queries with domain-specific terminologies.

Formulating effective search queries, especially with terminology spanning multiple fields, poses a significant challenge for researchers in IIS [48, 52]. This often involves iterative refinement of search keywords, informed by ongoing discoveries from the literature [20]. Given that LLMs excel in linking concepts and generating diverse query sets [75, 76], the system should leverage LLMs to assist users in crafting queries, thereby saving time and enhancing search efficiency.

DG3 Organize retrieved papers around exploration focus.

The sheer volume of papers encountered in IIS can overwhelm researchers [20, 48]. Moreover, search results often include irrelevant noises that detract from the exploration focus, as revealed in our design iterations. Previous HCI research suggests the aggregation of similar information and providing overviews is helpful in such cases to alleviate user burden [16, 33, 54]. Therefore, the system should categorize retrieved papers into overviews, enabling users to swiftly pinpoint relevant groups of literature.

DG4 Provide Information Scent for Efficient Paper Sifting.

In the IIS process, researchers often need to meticulously review papers to verify their relevance to specific aspects of their research [20, 52]. This labor-intensive task necessitates a mechanism for quickly identifying relevant information within papers. To facilitate this process, the system should highlight critical paper details and offer “information scents” that guide researchers in their review process [53].

4 THE DISCIPLINK SYSTEM

We introduce DiscipLink, a human-AI collaborative system designed to enhance interdisciplinary information seeking (IIS). Unlike previous efforts in scholarly information seeking, DiscipLink focuses on enabling researchers to effectively collaborate with LLMs for broader exploration across disciplines, thereby linking their own projects with diverse knowledge.

Generally, inspired by prior research [26, 33], DiscipLink facilitates both top-down and bottom-up mixed-initiative literature discovery workflows (Figure 2). In the top-down workflow, users

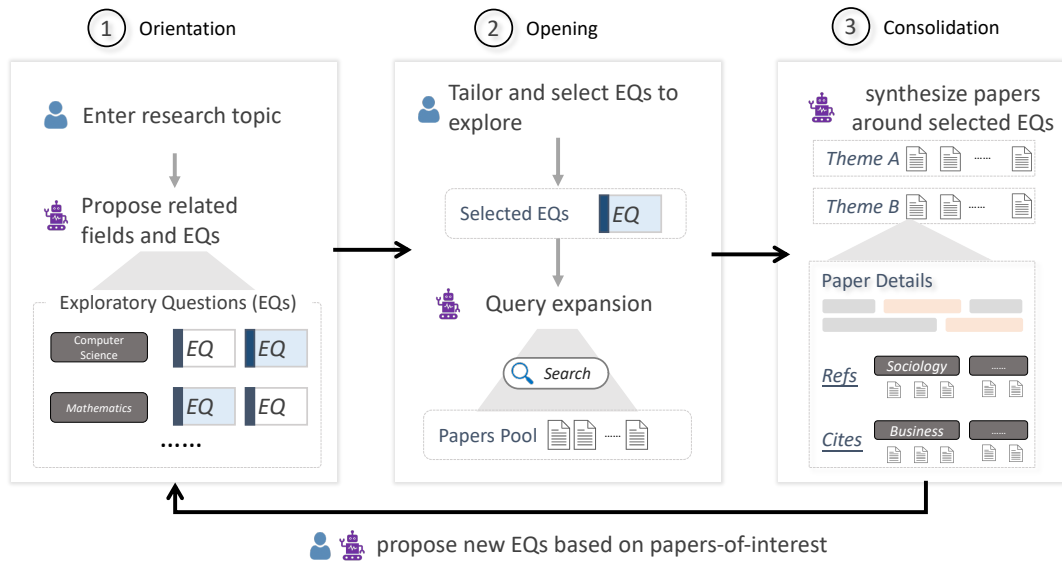


Figure 2: The workflow of DiscipLink. We harness LLMs to support the three stages of the IIS process.

decide the exploration direction they are interested in first, avoiding being lost in the details of a number of papers. To aid this process, DiscipLink generates diverse EQs from various disciplines based on the user’s research topic (Figure 2-1), crafts comprehensive queries for each EQ (Figure 2-2), and summarizes key themes from retrieved papers to streamline screening (Figure 2-3). On the other hand, the bottom-up approach empowers users to spontaneously navigate through papers, citations, and references. Users can organize the papers of interest and formulate new EQs based on these papers, thereby starting new research directions. DiscipLink enhances this exploration by organizing citations and references based on user exploration history and suggesting EQs informed by user-selected papers.

As a research prototype, DiscipLink uses GPT-4 as the underlying LLM, Semantic Scholar API service for searching papers, and “text-embedding-3-small” OpenAI [49] for text embedding. The subsequent subsections detail a usage scenario and describe DiscipLink’s key features in addressing our design goals.

4.1 Usage Scenario

Zoey, a computer science researcher, is tackling a new project aimed at raising awareness about social media misinformation among older adults. While she possesses expertise in misinformation detection through natural language processing, she finds herself stumped by the challenges older adults face in navigating social media and the human factors influencing misinformation awareness. To bridge these gaps, Zoey turns her attention to unfamiliar disciplines like sociology and psychology for insights. Beginning with search terms such as “older adults misinformation,” through Google Scholar, she finds a mountain of papers that seem relevant across disciplines, including education, sociology, and psychology, each offering unique perspectives on the subject. Determined not to miss out on anything potentially useful, Zoey digs into each paper, one by one.

However, the scattered insights across different fields make her research feel like assembling a vast puzzle without a guiding image. Moreover, she needs to understand different perspectives during her search, each requiring a shift in mindset. Struggling to manage the information overload and the constant shifts in disciplinary perspectives, Zoey considers seeking help from DiscipLink.

Upon describing her research project to DiscipLink, Zoey receives a list of exploratory questions (EQs) organized by specific disciplines. Zoey is particularly interested in the perspectives of psychology, education, and sociology. These EQs, framed with easy-to-understand language, introduce some key factors, such as “What can older adults be motivated to learn digital literacy?” (education), “What cognitive strategies reduce belief in false information?” (psychology), that has never come to her mind before, possibly due to her unfamiliarity with these disciplines. Among these, the EQ “How effective are peer-led education programs for seniors?” catches her interest. However, aiming to align it more closely with her focus on misinformation, she refines it to “How can peer-led education programs combat misinformation among older adults.”

After tailoring several EQs across psychology, education, and sociology to her preferences, Zoey prompts DiscipLink to fetch relevant papers. After searching, DiscipLink organizes the findings into themes for each selected EQ in the Exploration view (Figure 3-middle). Delving into the search results on cognitive strategies to counter misinformation, Zoey discovers themes such as “Critical Thinking and Cognitive Improvement” and “Inoculation and Overconfidence Strategies to Combat Misinformation”. This overview allows her to navigate the themes without getting bogged down in the details of individual papers. Attracted by the concept of “psychological inoculation,” a term new to her, she explores a linked wiki page, uncovering its relevance in preemptively addressing misinformation. Through annotated key phrases in paper titles

and abstracts, such as “misperception” and “credibility,” Zoey confirms the relevance of these papers, collecting them by dragging the theme to the Collection view (Figure 3-right).

Later, Zoey’s exploration leads her to a highly cited paper, and habitually, she checks its citations. DiscipLink categorizes these citations by discipline, with “Medicine” surprisingly ranking in the top three. Out of curiosity, she examines the citations under Medicine, and based on the highlighted terms, she finds a paper that discusses the impacts of health misinformation on older adults. She realizes the strong connection between health misinformation and the older adults, but she hasn’t explored this direction. To extend her exploration, Zoey drags the paper to the orientation view (Figure 3-left). Based on the paper, DiscipLink recommends three new EQs around health misinformation’s impact on older adults in the context of social media, thus broadening her exploration horizon.

4.2 System Design

4.2.1 [DG1] Elicit Exploration across Disciplines. To support users in expanding their exploration to relevant information scattered in various academic fields (DG1), DiscipLink encapsulates potential literature search objectives into interactive elements named *exploratory questions* (EQs). In other words, DiscipLink proposes EQs from different disciplines’ perspectives to elicit users’ breath exploration. For example, when exploring a research topic of facilitating older adults to learn about AI tools, possible EQs could be “*what cognitive challenges older adults may face in understanding complex concepts?*”, which concerns psychology, and “*how to design interactive tools to support learning for older adults?*”, which is more related to HCI literature. Below, we introduce the enabled interactions around EQs in DiscipLink, the design rationale, and the EQ-generation process.

Supported Interactions around Exploratory Questions. Upon users entering a research topic (Figure 3-A1), DiscipLink generates a list of EQs tailored to the *research topic of interest*. In this process, we leverage the inherited knowledge of LLMs derived from its large-scale training data to ensure the coverage of diverse disciplines. Users have the flexibility to modify their topic description anytime, prompting DiscipLink to produce new EQs based on the updated description or to initiate a new exploration session by resetting the current ones (Figure 3-A2). The EQs are categorized by their relevant disciplines (Figure 3-D1), with a navigation bar allowing users to easily filter EQs by research field (Figure 3-B). By selecting EQs, users can specify their exploration intentions for DiscipLink to identify literature closely aligned with their interests. Further interactions include editing EQs (Figure 3-D2), adding or removing selected EQs (Figure 3-C1), and generating additional EQs by specifying disciplines and particular needs or by manually entering questions (Figure 3-C2).

Note that if users encounter an interesting paper on a yet-to-be-explored topic, they can obtain new EQs from DiscipLink based on the paper by dragging it from the exploration or collection view to the orientation view. Users have the option to pinpoint specific aspects of a paper, such as particular keywords, that pique their interest. Upon specifying these details, they can direct DiscipLink to

suggest new EQs tailored to these focal points. We set DiscipLink to provide three new EQs each time. Users can choose one or several EQs to refine or opt to craft new EQs from scratch. Once these EQs are finalized and submitted, DiscipLink initiates a new round of search for papers concerning these EQs immediately.

Design Rationale. Our design of EQs draws inspiration from human-centered design principles, including the creation of a *shared representation* to balance automation with user agency [25], and the *reification of actions into objects* for direct manipulation [7]. EQs act as a bridge between users and the system, offering an intuitive means for users to convey their interests and for DiscipLink to interpret and further translate these interests into complex queries, thereby fetching a broad spectrum of relevant papers. We have explored other forms of LLM assistance in the design process for guiding and expanding user exploration in IIS. Initially, our prototype leveraged LLMs to generate research topic statements for users. However, this approach led to two main issues: the topics often contained inaccuracies due to LLM hallucinations, and they served more as end products of exploration instead of inspirations for users to explore. Another alternative we considered was utilizing LLMs to generate simple queries, similar to those users might input into search engines like Google Scholar. This method, while straightforward, can be inefficient in IIS. Effective queries for IIS necessitate the use of discipline-specific terminology, which could be challenging for users to understand. To address the problems in the above two methods, EQs in DiscipLink are framed as questions rather than facts, aiming to expand the exploration space, and users can flexibly customize EQs to *co-explore* with LLMs. Also, EQs are crafted in user-friendly language as a high-level abstraction of research directions over low-level typical search queries.

Prompts Design. DiscipLink takes two main types of inputs from users to generate EQs: the *research topic of interest* and/or the *paper of interest*, as illustrated in Figure 4. To generate EQs from diverse disciplinary perspectives based on these seeds of exploration, we employ prompt-chaining and persona-prompting techniques [78, 79].

The process involves chaining prompts to guide LLMs in generating EQs. More specifically, given a research topic, we provide the list of disciplines used by Semantic Scholar for LLMs to determine which of them might be relevant. The list includes primary disciplines, such as “computer science” and “psychology”. LLMs then determine specific subfields within these disciplines (e.g., “developmental psychology” within “psychology”) that closely align with the topic of interest, enabling it to propose more targeted questions. Then, for each identified field and subfield, our prompt to LLMs begins with, “*You are an expert in field. A researcher is developing a research topic: topic. What knowledge from your field could assist in developing this idea? Present your advice in the form of questions.*” After collecting EQs across different disciplines, we instruct LLMs to refine this list by removing any duplicates, resulting in the final EQs shared with users.

Preliminary experiments indicate that adopting specific expert personas enables LLMs to produce questions that closely match the disciplinary focus. For instance, when exploring how AI tools can support older adults learning in science without employing

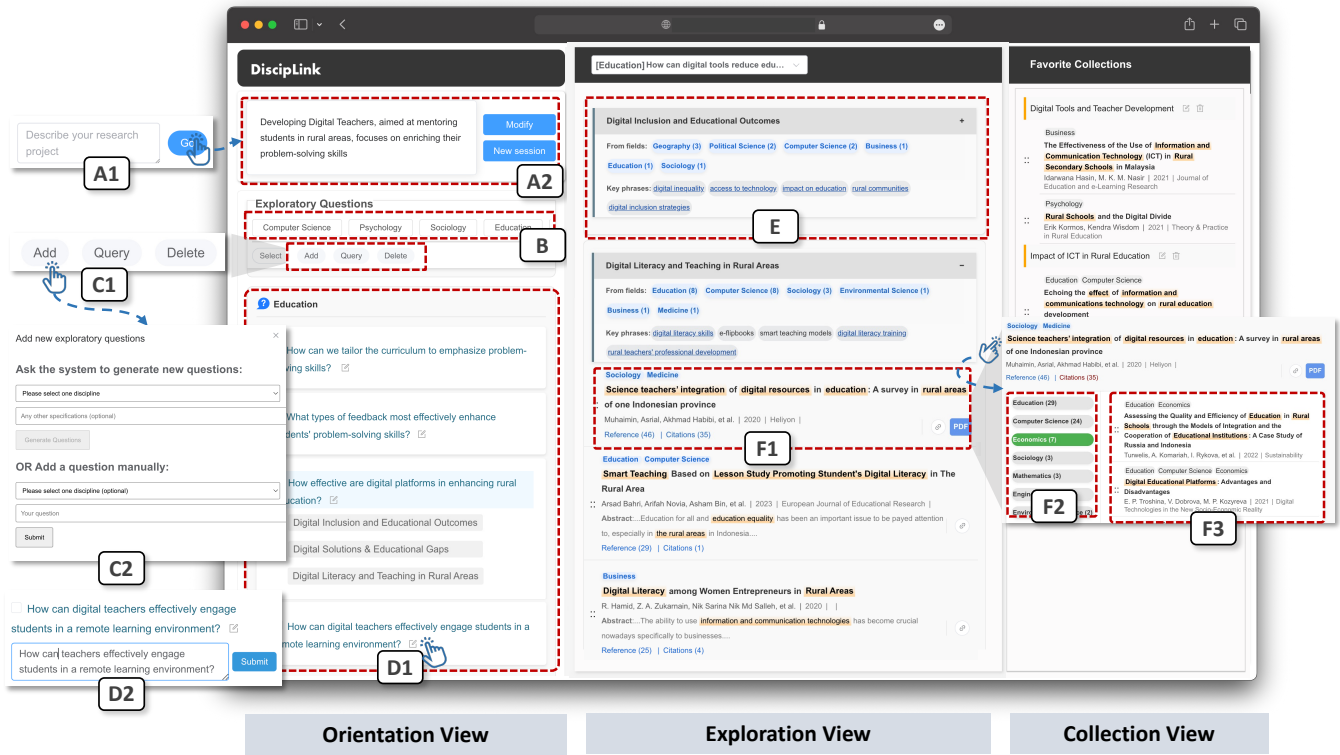


Figure 3: The user interface of DiscipLink, including *Orientation View* for creating and tailoring EQs, *Exploration View* for explore retrieved papers for each EQ, and *Collection View* for organizing papers users found useful.

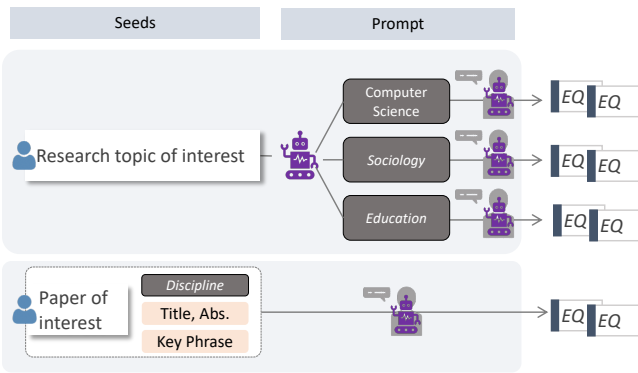


Figure 4: DiscipLink provides diverse EQ suggestions to discover highly-scattered relevant information based on two types of seeds given by users: research topic of interest and paper of interest. We prompt LLMs to simulate as experts in specific academic fields to provide EQs. Users can also create EQs by themselves.

a persona, language models tend to produce questions centered on “AI for learning.” These questions focus predominantly on “AI,” steering the subsequent discovery process toward computer science literature rather than education. Conversely, utilizing persona

prompting can generate EQs that lean more heavily towards aspects of older adults learning from an education perspective, such as “What motivates older adult learners in science education?”

For generating EQs based on a paper of interest, perhaps discovered through navigating the references and citations of papers returned in the previous rounds of exploration, DiscipLink adopts a similar persona-prompting strategy. Here, the meta-information of the paper, including its title, abstract, designated fields, and keywords, is integrated into the prompt context, allowing users to specify keywords that highlight their areas of interest. Following this, we prompt LLMs to identify relevant fields within the primary disciplines and simulate experts in these fields to recommend EQs related to the provided context.

4.2.2 [DG2] Query Expansion to Retrieve Relevant Papers.

When users express interest in a particular EQ, DiscipLink furthers the exploration by expansively generating numerous queries related to the EQ. Next, it conducts searches for papers in Semantic Scholar using these queries concurrently. In EQ generation, we guide LLMs to employ straightforward, easy-to-understand language so that users can quickly decide whether to dive into an EQ or not. However, when it comes to searching for papers, we instruct LLMs to use precise terminology and field-specific jargon relevant to the EQ to craft the search queries. This aims to fill the terminology gap of IIS introduced in subsection 2.1. Moreover, we intend to retrieve a diverse set of papers using the output queries to allow DiscipLink

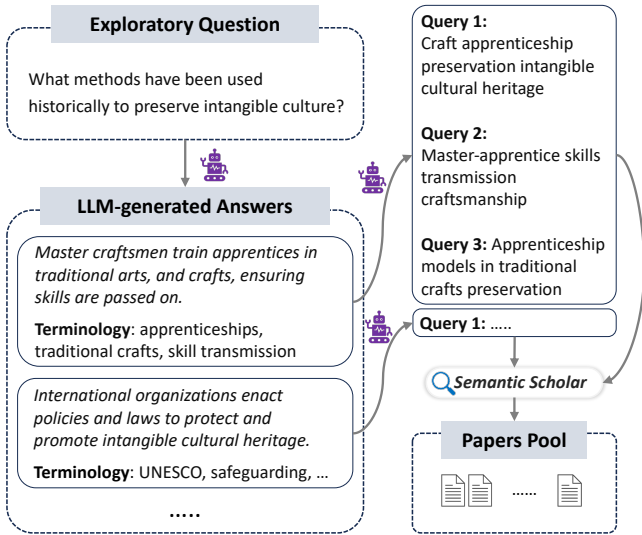


Figure 5: DiscipLink generates a diverse set of queries to retrieve relevant papers for an EQ.

to map out the knowledge landscape around the EQ. We design a prompt to craft queries according to these requirements, as shown in Figure 5.

To be more specific, we begin by requesting LLMs to provide (pseudo-)answers from various perspectives on the current EQ in the form of bullet points. Although there is a risk of LLMs hallucination, these responses, even when flawed, are found to contain keywords and concepts beneficial for formulating effective search queries, as noted by Wang et al. [75]. Subsequently, LLMs compiles a list of relevant terms associated with each bullet point. These terms lay the groundwork for creating a set of targeted search queries, which are later fed into paper search engines. This process, inspired by Wang et al. [76], helps to enrich the conceptual breadth and depth of inquiries into varied branches of knowledge.

4.2.3 [DG3] Organize the Retrieved Papers of around EQs.

It is challenging for users to go through the extensive papers retrieved from the queries, especially considering that many papers are likely from unfamiliar fields of study. To alleviate this information overload, DiscipLink distills key themes from the papers, focusing on the connection to the user’s current exploration context (i.e., the research topic and the selected exploratory question).

Themes Extraction. While the papers retrieved from the expansive queries could contain many themes, users would be more interested in those more related to their current exploration context, i.e., the research topic and the selected EQ. A LLM-based clustering pipeline (Figure 6) is deployed to sift through papers and extract themes pertinent to the user’s interests. Several LLM-based clustering techniques have been proposed by researchers, while trade-offs exist. For example, LLOOM [36] and ClusterLLM [82] produce high-quality clusters but need to operate at the data instance level – LLOOM involves distilling info from each input data

instance and ClusterLLM uses triplet comparisons – leading to increased time costs. To balance quality and time costs and enable interactive theme exploration in DiscipLink, our pipeline begins by generating contextualized embeddings for papers, proceeds to cluster these embeddings, and concludes with LLMs refining the clustering outcomes.

In the first step, we are inspired by the keyphrase expansion technique proposed by Viswanathan et al. [74] to construct the contextualized paper embedding. This technique concatenates original text embedding with keyphrase embedding related to user intents. Similarly, in DiscipLink, the contextualized embedding comprises text embedding of the paper’s metadata (title and abstract) and embeddings of the paper’s discipline, if matching the EQ’s discipline, as well as key phrases from the paper’s metadata resonating with the research topic and EQ. By considering whether the paper’s discipline matches the EQ’s discipline in the paper embedding, we add a soft constraint on clustering papers that fall into the same field of study. To identify relevant phrases from academic papers, we first convert key concepts $C = \{c_1, c_2, \dots, c_m\}$ from the research topic and EQs, along with phrases $P = \{p_1, p_2, \dots, p_n\}$ from the paper’s metadata, into embeddings. A phrase p_i is considered relevant if its cosine similarity with any concept c_j , defined as $\text{sim}(v_{c_j}, v_{p_i}) > 0.6$, surpasses a predetermined threshold of 0.6⁶.

In the clustering phase, we employ DBSCAN [57] to group the contextualized embeddings. While, through constructing the contextualized embeddings, the algorithm is more likely to put papers relevant to the exploration context together, it can produce other clusters less related to the contexts. Besides, the resulting clusters might still have many papers that making it challenging for users to quickly screen them. To this end, we harness LLMs to curate the clustering results. Specifically, we prompt LLMs with two questions for each cluster: “Is this cluster related to the research topic and the selected question?” and “Is there a clear division between papers that a subclustering would be helpful?” Based on the results, we remove irrelevant clusters and decide whether to divide any existing clusters further. We still keep the papers under the removed clusters in a “possibly relevant paper” list in DiscipLink for participants to browse if they want. This design decision is driven by the observation that researchers would not only search on purpose but also be interested in serendipitous information discovery [20].

Supported Interactions. Upon selecting an EQ for which papers have already been queried, DiscipLink presents the extracted research themes. It displays the disciplinary distribution and key phrases for papers within each theme (Figure 3-E) and links the key phrases to their respective wiki pages via DBpedia Spotlight [44] when possible. This feature aids users in demystifying unfamiliar terms. Users can delve into themes to review the papers classified under them.

4.2.4 [DG4] Support Sifting Papers. DiscipLink presents retrieved papers with their disciplines, titles, and meta information such as authors, published year, venue, abstract, and references as well as citations (Figure 3-F1). Users can also click a link button

⁶This threshold is determined through empirical testing during development.

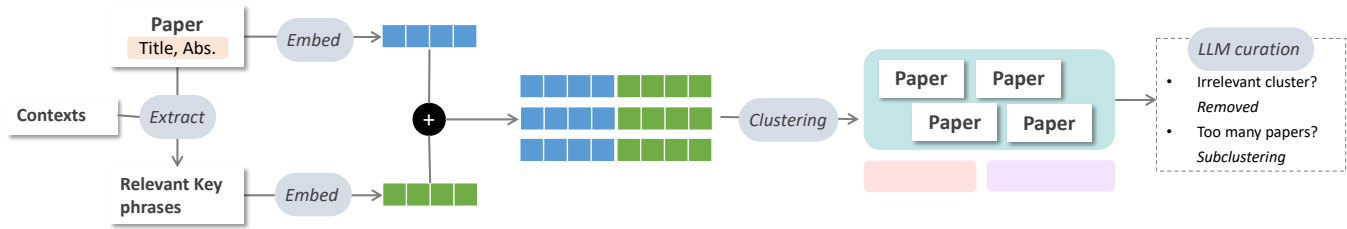


Figure 6: To organize the retrieved papers, DiscipLink first clusters the papers by embedding them based on the meta information and their connection to the exploration contexts (i.e., the research topic of interest and the associated EQ); then, DiscipLink uses LLMs to curate the clusters by removing less relevant clusters and doing subclustering if too many papers exist in one cluster.

to navigate to the Semantic Scholar page of the papers for further investigation.

Besides, DiscipLink highlights the relevant phrases in the title and abstract that are aligned with the current exploration contexts, i.e., research topic as well as the selected EQ, as shown in Figure 3-F1. The method of extracting relevant keyphrases is the same as the one in the clustering process described in subsection 4.2.3. Additionally, to prevent information overload in the user interface, only the sentence from the abstract deemed most relevant to the exploration context is displayed. Formally, given an abstract represented as a document consisting of a set of sentences $D = \{s_1, \dots, s_n\}$ and the set of concepts in the exploration contexts $C = \{c_1, \dots, c_m\}$, the key sentence s^* is identified by: $s^* = \arg \max_{s_i \in D} \{ \{c_j \in C \text{ exists a phrase relevant to } c_j \text{ in } s_i\} \}$. Users can also expand the key sentence to see the full abstract at will.

4.2.5 Other Features. Besides the aforementioned four features that directly target the four design goals, DiscipLink also includes some other features to facilitate the IIS process.

Support Organization of Useful Papers. DiscipLink allows users to drag and drop themes or papers from the *exploration view* to the *collection view*. When dropping themes, a collection is created in DiscipLink with the same title as the themes and containing the same group of papers. When papers are dropped, they would be added to collections based on their sources: from a theme or from citations and references of other papers. Users can edit the title of the collections, move papers from one collection to another also through dragging and dropping, and delete collections. This organization feature, together with the themes extraction feature, supports the mixed-initiative workflow proposed by Kang et al. [33].

Ranking Citations and Reference. DiscipLink displays the citations and references of a paper upon users' requests, and these citations and references are organized based on their disciplines (Figure 3-F2). In particular, we ranked the disciplines to prioritize the ones that contain relevant information to the exploration contexts but are less explored according to users' interaction history with the system. This ranking aims to balance exploration and exploitation [3] and support **DG1**. We define users' engagement level in each discipline as $U_d = p_d + q_d$, where p_d is the number of papers collected and q_d is the number of EQs queried in discipline d . The exploration score, E_d , is set to $1/(U_d + 1)$ to encourage exploring

lesser-engaged disciplines. The relevance of each discipline, V_d , is computed as the average cosine similarity between the paper embeddings in d and the user-specified research topic embedding. The combined score C_d for ranking disciplines is simply computed as $C_d = \beta V_d + E_d$, where β is set to 1 in our prototype. For papers under each discipline, we rank them by their cosine similarity score with the users' topic of interest.

5 STUDY 1: USABILITY EVALUATION

To evaluate DiscipLink, we first conducted a usability study on DiscipLink with a focus on how useful and easy it is to use DiscipLink in helping researchers search papers regarding an interdisciplinary research topic. Specifically, we explore the following research questions:

- RQ1** Can DiscipLink help researchers gather more comprehensive literature?
- RQ2** Does DiscipLink improve user efficiency in the IIS process?
- RQ3** How do users perceive the usefulness and experience of interacting with DiscipLink??

5.1 Participants

We recruited 12 participants (six females, five males, and one prefer not to say) from social media. They are all postgraduate research students. Eight participants have 1-3 years of research experience, three have 3-5 years, and one have 5-10 years. Their research directions include HCI (6), visualization (2), machine learning (2), programming language (1) and cognitive science (1). All but one participant confirmed that they have experience in interdisciplinary research. Additionally, five participants suggested they often use AI tools, including Consensus⁷, Elicit⁸, Perplexity⁹ and Paper Digest¹⁰, for searching literature.

5.2 Protocol

We adopted a within-subject experiment design. Participants were asked to complete IIS tasks under two conditions in a counter-balanced order: the system condition (using DiscipLink) and the baseline condition (using tools they usually use). We prepare two interdisciplinary research topics for participants to perform IIS:

⁷<https://consensus.app/>

⁸<https://elicit.com/>

⁹<https://www.perplexity.ai/>

¹⁰<https://www.paperdigest.org/>

- Develop digital teachers to mentor high school students in rural areas, focusing on training their problem-solving capabilities. We refer to this task as **AI for education** in later texts.
- Help designers use AR to creatively document intangible cultural heritage, ensuring its preservation for future generations. We refer to this task as **AR for ICH** in later texts.

Participants were instructed to widely survey papers from various academic fields for each topic. In both conditions, participants were required to complete an outline in Google Docs as the task outcomes. In the outline, participants needed to list the research papers they found relevant and group them into topics. Participants were also asked to write a short explanation of how the topics related to the given project.

Before the task starts in the system condition, we first provide a 5-minute tutorial on DiscipLink and give 5 minutes for participants to try DiscipLink. Then, they are given 30 minutes to search papers and complete the outline for each task. After each task, they are asked to complete a corresponding questionnaire. Upon completing both tasks, we ask participants to share their experience in using DiscipLink.

5.3 Measurement

To answer **RQ1**, we compare participants' outlines in the two conditions from three aspects: statistics of the outlines, participants' subjective ratings, and ratings from external experts. For the statistics of the outlines, we computed the number of topics and papers in each outline. For the participant's ratings, in the questionnaire after each task, participants self-rated the outline's helpfulness, comprehensiveness, organization, and relevance using the 7-point Likert Scale. Moreover, before each task started, we asked participants to tell us what they planned to search for during the task. Subsequently, upon completion of the assigned task, we directed the participants to provide ratings indicating the extent to which they identified relevant literature *they had planned to search*, as well as the degree to which they discovered relevant literature that *had not been part of their initial search strategy*. These ratings help to assess whether DiscipLink achieves the goal of expanding researchers' exploration scope. For the expert ratings, we recruited two experts to score the outlines. One is a final-year PhD student, and another is an assistant professor. Both are HCI researchers and have experience in multiple interdisciplinary research projects widely covering AI, XR, education, and design. They evaluated the participants' outlines while blind to the condition using a 10-point Likert scale by rating the following statements:

- Overall outline helpfulness: *"I found the outline is overall helpful for developing the research project."*
- Outline comprehensiveness: *"I believe this outline covers a comprehensive set of relevant topics."*
- Topic relevance for each topic: *"I found this topic relevant to the project."*
- Topic support for each topic: *"I found the topic to be well-supported by papers under it."*

We refer to Kang et al. [33] in designing these statements except for the outline comprehensiveness. We add the outline comprehensiveness as a measurement as an important design goal of DiscipLink is

to assist researchers in gathering highly-scatter knowledge. Before their evaluation, experts first engaged in a session to discuss how to give specific points. Experts' given ratings were averaged as the final score.

To answer **RQ2**, we measure the task workload using the NASA task load index [23]. We also measure participants' knowledge gain during the search process, as researchers typically need to learn knowledge from other disciplines on the fly in IIS. Participants rated their knowledge on the given research topic using a seven-point Likert scale before and after the task, and we take the change in self-rated knowledge as the knowledge gain [51]. They also rate their self-efficiency during the task.

To answer **RQ3**, we adopt the technology acceptance model [73] to assess the perceived usefulness, ease to use, ease to learn, workflow compatibility of DiscipLink, and participants' intention to use it. Participants also rate their perceived helpfulness of DiscipLink to key stages of IIS, including orienting to various academic fields, breadth searching papers, filtering irrelevant papers, and making sense of the connection between the papers and the focused research project.

5.4 Findings

For statistical results presented in this section, we used the paired samples t-test for comparison between conditions, unless the Shapiro-Wilk test of normality indicated a deviation from normality. In cases of non-normal data, we utilized the Wilcoxon Signed Rank test.

5.4.1 [RQ1] Quality of the outline. Participants in the system condition, on average, collected 5.08 topics (SD=2.28), more than the baseline condition (M=3.92, SD=1.51), but there are no significant differences ($p=0.057$). There are also no significant differences regarding the number of papers (the system condition: M=20.58, SD=14.39; the baseline condition: M=17.83, SD=10.31; $p=0.58$).

The experts ratings indicated significantly higher outline helpfulness of the system condition (M=6.67, SD=1.09) than the baseline condition (M=6.04, SD=0.96, $p=0.003$). Similarly, experts' ratings of outline comprehensiveness are significantly higher in the system condition (M=6.63, SD=1.23) than in the baseline condition (M=5.83, SD=0.69, $p=0.02$). Nevertheless, there are no significant differences regarding averaged topic relevance and averaged topic support for the two conditions.

For participants' subjective ratings (shown in Figure 7), there are no significant differences between the self-report outline helpfulness ($p=0.16$), organization ($p=0.31$), and relevance ($p=0.18$). However, participants reported higher comprehensiveness of the outline in the system condition (M=5.00, SD=1.60) compared to the baseline condition (M=3.08, SD=1.31, $p=0.002$). Besides, there is no significant difference between how well participants consider they found literature they plan to search ($p=0.07$). But regarding how well they found relevant literature they originally did not plan to search, the system condition (M=6.08, SD=1.44) is significantly better than the baseline condition (M=4.67, SD=1.44, $p=0.05$).

These findings suggest that DiscipLink is beneficial in helping researchers to gather comprehensive literature from various disciplines as intended.

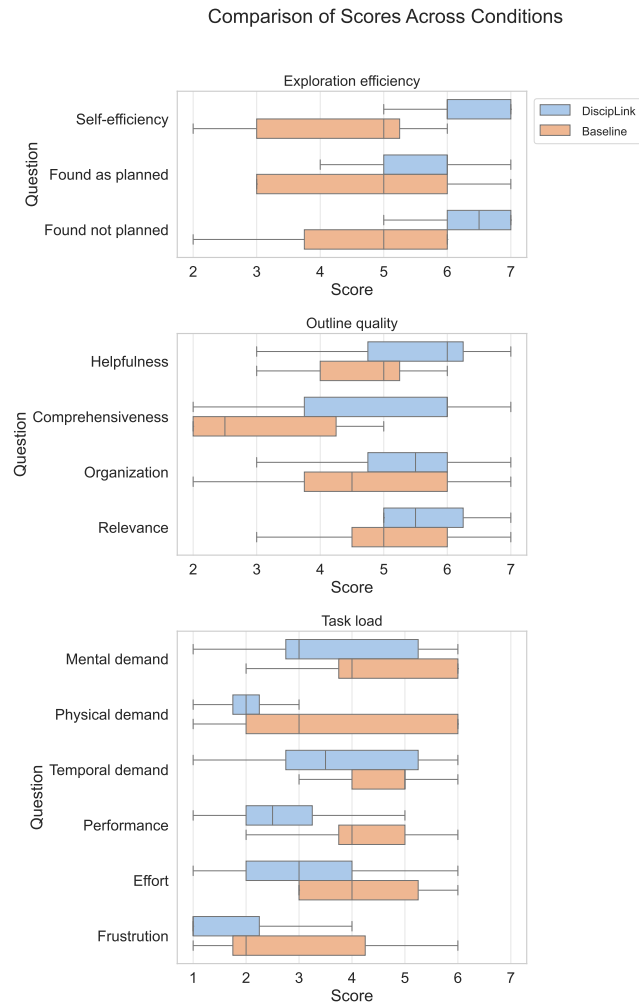


Figure 7: Comparison of participants' self-ratings between the two conditions. In particular, 1 point of "Performance" means success, and 7 means failure.

5.4.2 [RQ2] Search process. Participants in the system condition reported significantly larger learning gains ($M=2.50, SD=2.468$) compared to the baseline condition ($M=1.08, SD=1.56, p=0.02$), suggesting DiscipLink helps users learn more during the exploration. Also, they self-reported higher efficiency in the system condition ($M=6.18, SD=0.94$) compared to the baseline condition ($M=4.33, SD=1.50, p=0.01$). Participants exhibited a significantly higher task workload in the baseline conditions based on the averaging score of the NASA TLX items ($M=4.01, SD=0.74$) compared to the system condition ($M=2.93, SD=0.97, p<0.01$). For the individual items (illustrated in Figure 7), the system condition is significantly better than the baseline condition in regard to the physical demands ($p=0.01$), performance ($p=0.03$), effort ($p=0.02$), and frustration ($p=0.04$). These results suggest that DiscipLink improves participants' search efficiency, lowers their workload, and helps them effectively learn more in the IIS process.

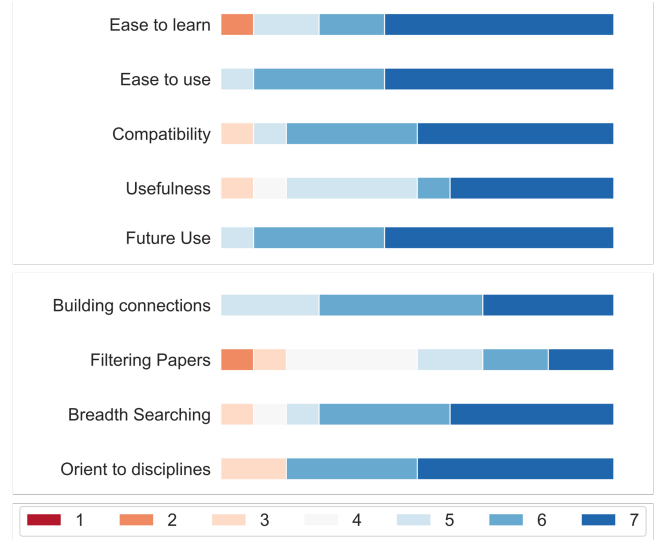


Figure 8: Participants ratings of DiscipLink.

5.4.3 [RQ3] User experience. Participants generally held positive attitudes towards the usefulness, ease of use, ease of learning, and workflow compatibility, as shown in Figure 8. Most participants appreciated the generated EQs. P12 noted that “When I came to this new field, I didn’t know where to go and how to search on Google. There might be some terms in this field, and then it [DiscipLink] could help me ask some questions. It is difficult for me to ask professional questions in this way.”

Nevertheless, some participants complained that DiscipLink provides too many EQs and thus adds load to users: “It has many [EQ] recommendations, and I may still need some efforts to screen them. I need more time to see if this is relevant, and it is unlikely to be completed very well in a relatively short time.” (P06). P07 expects the EQs to be better connected, “There may be a connection or overlap between them, but it didn’t tell me.”

Besides, as illustrated in Figure 8, there were mixed opinions among participants regarding DiscipLink’s effectiveness in filtering papers. Those who viewed DiscipLink positively highlighted its ability to deliver relevant themes and papers aligned with the selected EQs. (“I think the results of the query are very accurate. I can’t get such accurate results using Google Scholar because I feel that it only matches titles. You can search more accurately with this [system].”, P11) and appraise the keyphrase highlighting (“The papers provided by other search engines may not be relevant, and I don’t know exactly what relevant keyword it has. But you can clearly see the fields, the key phrases [in DiscipLink]”, P09). However, some participants felt challenged by the volume of papers DiscipLink recommended, pointing out the difficulty of effectively screening them within the experiment time, “Because it recommends a lot, I still need manual effort to screen them. I can’t complete it very well in a relatively short period of time.” (P6)

We also examined what tools participants used in the baseline condition. We found that all participants used Google Scholar, while P6 and P10 used tools including LitMaps and Connected Papers to check citation networks of papers, and P01, P08, P11, P12 attempted

AI-powered literature search tools including Consensus, Elicit, Paper Digest, and Perplexity. Compared to DiscipLink, participants suggested the tools displaying citation networks offer better visual salience on key papers but provide limited assistance in guiding search directions. P10 noted that “*Some of them [papers] were clustered together [in the citation network], and then I took a quick look at those. [...] But I just picked the standout ones and need to read carefully to proceed, while your tool is much better for getting a comprehensive exploration quickly.*” Regarding AI-powered search tools, P01, P11, and P12 initially tried them but quickly reverted to Google Scholar due to a lack of helpful answers. Conversely, P08 found Paper Digest useful for creating outlines but felt that “*Your system has several exploration levels, allowing me to understand the paper I gave better. In the summary generated by Paper Digest, I am more about reading and accepting papers, and my understanding of these papers may not be as clear.*” This feedback illuminates the unique strengths of DiscipLink in unfolding the IIS process.

6 STUDY 2: CASE STUDY

Encouraged by the results of study 1, we further conducted an open-ended exploratory study with experienced interdisciplinary researchers to get qualitative feedback on the potential benefits and limitations of DiscipLink in IIS.

6.1 Methodology

6.1.1 Participants. We recruited seven researchers (two females and five males) who are focused on interdisciplinary research, including one assistant professor, two postdoctoral researchers, and four senior PhD students. Participants are from various backgrounds, including AI for healthcare, data storytelling, social psychology, AI for science, music technology, and AI for education.

6.1.2 Protocol. We recommended that participants choose projects for which they haven’t fully devoted themselves to searching for relevant literature yet, in order to explore those projects in this study. If they do not have such projects or they feel uncomfortable doing that, participants could also explore papers under their other interdisciplinary projects. Participants were informed that none of the information on their research projects would be reported.

After getting the consent of participants, at the beginning of the study, we gave a tutorial on DiscipLink, as the one in Study 1, and suggested participants get familiar with the system through a 5-minute practice with a given research topic. We then asked participants to explore their interdisciplinary projects using DiscipLink. We encouraged participants to think aloud and explore DiscipLink for at least 20 minutes. Participants are allowed to explore DiscipLink for a longer time if they are willing to do so. After the open-ended exploration, we conducted semi-structured interviews with participants. The interview questions cover participants’ general experience with DiscipLink, whether DiscipLink helps in uncovering relevant knowledge in multiple disciplines, whether DiscipLink found helpful papers, whether DiscipLink helps in screening papers, and how well DiscipLink can fit into their daily IIS process. We also asked whether participants felt the level of automation was appropriate.

6.1.3 Data Collection and Analysis. All study sessions were audio-recorded with the informants’ consent. The audio recordings were transcribed using automated tools and subsequently manually verified. Two members of our research team analyzed the transcripts using thematic analysis [8]. The coding results were thoroughly discussed and integrated through iterative discussions.

6.2 Findings

We report our findings based on the participants’ qualitative feedback, focused on those not mentioned by participants in Study 1. Our reports do not include information about participants’ research projects to protect participants’ privacy.

6.2.1 Helpfulness of EQs. All participants found some EQs generated by DiscipLink **relevant and provided insights** to their research projects, enhancing their literature exploration process. P7 considered EQs “*introduces a creative and playful way to pick your sense of curiosity.*” P3 specifically noted that the EQs aligned closely with the fields most pertinent to his research and matched the disciplines associated with the references cited in his paper. He noted how exploring these EQs can be helpful to his paper: “*There are some [EQs] that are relevant and can be helpful for developing the introduction part of the paper. There are also three or four questions that can be explored to develop follow-up future studies or some other discussions.*”

Besides, P4 and P7 both mentioned that the generated EQs would particularly **be beneficial to researchers in the onboarding stage** of interdisciplinary research. P4 found the EQs are related to “*the core knowledge in each field*”; thus “*it is very helpful for understanding knowledge for first-year Ph.D. students.*” Similarly, P7 commented that “*most of the researchers, including myself, have problems with the onboarding process. So, these types of applications or surveys could help us with the onboarding process and also learning the new research area step by step.*”

Moreover, most participants found that DiscipLink **streamlines the IIS workflow that aligns with and augments their original workflow**. P1 noted that:

In the past, I would think of directions in different disciplines and search for them. This process was in my mind, but it would definitely not be so comprehensive. It (DiscipLink) would ask some very related questions that I had not thought of before.

An exception is P6, who usually did not search literature by himself. He collaborated with experts from other disciplines and read the literature they suggested. However, he also found DiscipLink helpful in his workflow:

Sometimes, you don’t want to follow the advice of experts exactly. What experts say is not always correct. [...] This tool is very useful when I want to develop something new [beyond their strengths].

6.2.2 Various expectation of EQs. Although participants generally found EQs helpful, they also found some EQs did not meet their expectations. P4, whose background is chemistry and physics, found that all EQs start with “How” or “What”, but not “Why”. He considered “*asking why*” is more important as “*Understanding the underlying causes is important to conducting my research.*”

Another concern is about **the connection between the generated EQs and their projects**. Interestingly, there are diverging opinions from participants on this. P1, P3, and P6 consider some EQs as not good because the connection between these EQs and their projects is loose and only covers a portion of the key concepts in the projects. P1 noted that *“some questions are simply digging into one field, and it seems that they are less relevant to my interdisciplinary project.”*

However, P5 and P7 desired more EQs to delve deeply into other disciplines and not strongly connected to their own projects. One reason is the lack of existing research on their project; thus, EQs strongly connected to the project *“won’t be able to help me find something very useful.”* Another reason is the need to delve deeply into other disciplines of research, as explained by P5 with her previous experience:

When I was doing [a certain concept] [related interdisciplinary research], I had to read a lot of psychological literature to see how psychologists used [a certain concept]. [...] If I don’t read real psychological papers and do it directly, then I will always read second-hand literature. I’ll never know why people chose that model, and it is problematic.

Moreover, some participants believed **the provided EQs should be different in different stages of the interdisciplinary research**. P2 elaborated on the development of the exploration needs and suggested adding more customization over the EQs generation:

When you first explore this [interdisciplinary] field, you can have more possibilities [...] But I’m actually pretty sure I only want papers with [a concept] and [a concept]. I think before [providing EQs], I can have a choice. For example, am I a person who is exploring this [interdisciplinary] topic for the first time, or am I actually an expert in this topic?

6.2.3 DiscipLink supports human-AI co-exploration. Participants generally considered the level of automation to be appropriate, and they were able to engage and direct the exploration. We observed **how participants actively tailor the EQs for their own goals**. For example, P1 edited EQs by adding concepts more related to his project to strengthen the connection between the found papers and the project; for P7, the original question was about how an intervention affects an element, and P7 edited the question by adding more elements could be affected to enlarge the search scope. Participants also initiated the creation of EQs based on their insights during the exploration process. P1 found “gamification,” and P2 found “music” as elements that they hadn’t considered in their own projects but were relevant through browsing retrieved papers, and they created EQs based on these findings.

Additionally, P7 highlighted how the language framing of the EQs encourages him to tailor the EQs and join the co-exploration:

If your system was giving me very sophisticated questions. Then, it might restrict my own creativity. But because the questions are broad, easy-to-understand. I think it helps me also think by myself. And these [EQs] can act like hints.

We also found **most of the participants not only rely on the provided themes and the corresponding papers in DiscipLink to seek information, but also deliberately attempt more ways to discover papers**, including digging into the “possibly relevant paper” list, the citation and references of papers, and even browse the recommended paper list in the Semantic Scholar paper page after they clicked the link of papers in DiscipLink. This behavior aligns with the nonlinear model that researchers would deliberately increase the breadth of the search in IIS [20].

When asked about their sense of control during the exploration process with DiscipLink, **all participants affirmed they were in control**. They attributed this to having a clear objective for their exploration to advance their own projects. Thus, their engagement with the system’s recommendations was driven not by the system’s guidance alone but by the perceived value of these suggestions to their projects. P6 encapsulated this sentiment by stating, *“I have been deeply thinking about this research topic for a while. It’s always on my mind. AI cannot diminish my goal or change my focus.”*

7 DISCUSSION

This paper explores the potential of LLMs in supporting interdisciplinary information seeking and contributes to the growing body of research on human-LLM interaction across various domains [36, 37, 40, 61, 84]. It is worth noting that there are differing opinions in academia regarding the role of LLMs in enhancing search systems. Metzler et al. [45] envision that future search systems can build on LLMs to offer answers to user queries instead of references that require users to further explore. Conversely, Shah and Bender [58] argue against “dropping a LLM-based agent as a one-size-fits-all solution” and emphasize the importance of user interaction with information in search processes, which enables activities such as learning and serendipity. Our work aligns with Shah and Bender [58]’s perspective by offering a human-AI co-exploration solution that integrates LLM support while enabling rich user interaction during the search process. In this section, we discuss the lessons learned in designing and evaluating DiscipLink, as well as its limitations and potential directions for future work.

7.1 Exploratory Questions as Shared Representation in AI-powered IIS

In DiscipLink, we conceptualize exploratory questions as a shared representation between users and AI to facilitate co-exploration. The concept of shared representation, as introduced by Heer [25], refers to the use of a representation, be it textual or visual, in human-AI interaction that is mutually intelligible to both humans and AI. This shared foundation allows for collaborative contributions to the task at hand, thereby striking a balance between automation and user agency. Participants in Study 1 infrequently modified the EQs, possibly due to time constraints. However, during open-ended exploratory studies, we noticed participants customizing the generated EQs and creating their own, showing the potential of EQs to act as a shared representation that encourages engagement from both humans and AI.

Feedback from participants in our first two studies was generally positive regarding the utility of EQs. Yet, some noted the lack of visible connections between EQs, particularly when they span

across disciplines and overlap in themes after querying. DiscipLink currently does not highlight these links by presenting EQs in a list. To address this, future efforts could consider better supporting the curation and exploration of EQs. One possible direction is to consider the existing research on curating examples for creativity support, and visually contextualizing the EQs when they are generated [14, 41, 77].

7.2 Support Learning with Human-AI Co-Exploration

As noted in the literature [20, 52] and commented by our participants in study 2 (subsubsection 6.2.2), learning in IIS helps researchers conduct interdisciplinary research over a long period of time. This insight underpins our decision to create a human-AI co-exploration workflow rather than merely employing AI agents to autonomously compile research papers for scholars [80]. Our approach is in line with the idea of AI aiding human cognitive processes [21, 46]. Feedback from study 2 indicates that DiscipLink not only piques user curiosity but also facilitates their initiation into interdisciplinary research, thereby supporting learning within IIS.

However, much of the existing research on human-AI collaboration has prioritized enhancing collaboration to boost immediate performance outcomes [5, 12, 42, 43], often focusing on short-term metrics such as the accuracy of decision-making with AI support. While these immediate results are valuable, we advocate for a broader focus within the field, emphasizing support for enduring outcomes of human-AI collaborations. This includes promoting long-term project success (as in interdisciplinary research), enhancing worker well-being, and contributing to societal benefits.

7.3 Contextualization in IIS

Interdisciplinary research often involves sifting through a vast array of potentially relevant papers. However, the specific contexts of researchers guide them in concentrating on particular studies. DiscipLink acknowledges this by using the user-provided research topic and selected EQs as a framework for offering contextualized support: recommending EQs tailored to the research topic and synthesizing themes from the papers based on these EQs. Feedback from user studies confirms the relevance of EQs to their research topics and the extracted themes to the EQs.

Yet, findings from our Study 2 also indicate that interdisciplinary researchers expect DiscipLink to grasp their research contexts in a more nuanced way, beyond just the topics at hand. This includes understanding their preferences regarding the framing of questions (e.g., P4 shows a preference for “why” questions over “what” and “how”), their desired exploration scope (e.g., some participants favor an in-depth exploration into other disciplines while others prefer a stronger linkage with their own research project), and their current phase in the research process (e.g., P2 suggested questions could initially be broader and become more targeted as the research progresses). Although DiscipLink allows for customization of EQs, offering more nuanced EQ recommendations could significantly enhance the user experience by saving time and boosting efficiency. Therefore, future endeavors should delve deeper into the specific

factors of the needs of interdisciplinary researchers to offer more finely tuned exploration and research support.

7.4 Limitation and Future Work

7.4.1 Deepening the Support to the Nonlinear Exploration Process. The design of DiscipLink is motivated by the nonlinear model of IIS behavior [20]. DiscipLink supports the nonlinear exploration process primarily through generating EQs for orientation, query expansion for opening, and theme extraction for consolidation. However, support for researchers could be further extended, especially in the consolidation stage and the transitions between different stages. For example, DiscipLink does not provide direct supports for activities in consolidation such as incorporation and verifying, which is challenging due to the diverging tastes and paradigms across disciplines [9]. Future work can explore methods to help researchers understand how to use the found papers effectively. One inspiration could come from research on analogical search [31], which supports researchers in transferring solutions from other disciplines. Also, improving the interdisciplinary reading experience through features such as guiding questions or in-situ summarization as suggested by August et al. [4] can be helpful.

Moreover, participants indicated that DiscipLink is limited in assisting researchers in exploring the citation networks of papers of interest, which could facilitate transitioning from consolidation to opening or orientation. Existing work, such as Synergi [33] and Relatedly [50], can enhance this support. For example, Synergi can aid in theme extraction from citation networks based on text snippets of seed papers [33].

7.4.2 Scale of the user study. In our usability experiment, the task scenarios are mainly related to HCI research. We acknowledge that this design may limit the generalizability of our findings to other research contexts. Testing DiscipLink in a broader range of scenarios quantitatively would enable us to better evaluate its efficacy in supporting IIS. Additionally, both the usability experiment and the open-ended exploratory study are relatively short in comparison to the substantial amount of time typically required for researchers in IIS. Long-term field studies are needed to fully assess DiscipLink’s support in IIS, and benefits as well as limitations of human-AI co-exploration. For example, some participants in study 2 exhaustively explore all presented papers by the system, which can be explained by their open-minded IIS search habits [20]. It would be interesting to see whether long-term usage of DiscipLink would change such habits.

8 CONCLUSION

In this paper, we present DiscipLink, a novel interactive system designed to enhance the process of interdisciplinary information seeking through human-AI co-exploration. DiscipLink leverages the capabilities of large language models (LLMs) to assist users in their exploration efforts by generating exploratory questions (EQs) and facilitating automatic query expansion. It enables users to actively participate in the exploration by customizing EQs. Additionally, DiscipLink employs contextualized clustering to extract themes from the gathered papers and annotates key phrases within those papers, aiding users in quickly identifying how these papers relate to their research objectives. Our evaluation of DiscipLink includes

two user studies. The first study assesses usability, gathering both quantitative and qualitative feedback from 12 researchers, which indicates that DiscipLink significantly enhances researchers' ability to conduct thorough paper searches. The second study examines the potential integration of DiscipLink into the daily routines of interdisciplinary researchers, with qualitative findings highlighting the value researchers find in co-exploring with DiscipLink. However, these studies also highlight certain limitations of DiscipLink, pointing towards areas for future investigation and improvement.

ACKNOWLEDGMENTS

This work is supported by the seed fund of the Big Data for Bio-Intelligence Laboratory (Z0428) from The Hong Kong University of Science and Technology. We thank the anonymous reviewers for their constructive comments and our user study participants for their efforts and insights. We also thank Yujie Zheng for helping with the literature review and Chengzhong Liu for providing suggestions during the early system design stage.

REFERENCES

- Andrew Abbott. 2010. *Chaos of disciplines*. University of Chicago Press.
- Badr AlKhamissi, Millicent Li, Asli Celikyilmaz, Mona Diab, and Marjan Ghazvininejad. 2022. A review on language models as knowledge bases. *arXiv preprint arXiv:2204.06031* (2022).
- Kumaripaba Athukorala, Alan Medlar, Antti Oulasvirta, Giulio Jacucci, and Dorota Glowacka. 2016. Beyond relevance: Adapting exploration/exploitation in information retrieval. In *Proceedings of the 21st international conference on intelligent user interfaces*. 359–369.
- Tal August, Lucy Lu Wang, Jonathan Bragg, Marti A Hearst, Andrew Head, and Kyle Lo. 2023. Paper plain: Making medical research papers approachable to healthcare consumers with natural language processing. *ACM Transactions on Computer-Human Interaction* 30, 5 (2023), 1–38.
- Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the whole exceed its parts? the effect of ai explanations on complementary team performance. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–16.
- Marcia J Bates. 1996. Learning about the information seeking of interdisciplinary scholars and students. (1996).
- Michel Beaudouin-Lafon and Wendy E Mackay. 2000. Reification, polymorphism and reuse: three principles for designing visual interfaces. In *Proceedings of the working conference on Advanced visual interfaces*. 102–109.
- Virginia Braun and Victoria Clarke. 2012. *Thematic analysis*. American Psychological Association.
- Rebekah R Brown, Ana Deletic, and Tony HF Wong. 2015. Interdisciplinarity: How to catalyze collaboration. *Nature* 525, 7569 (2015), 315–317.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- Christine Susan Bruce. 1994. Research students' early experiences of the dissertation literature review. *Studies in Higher Education* 19, 2 (1994), 217–229.
- Carrie J Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S Corrado, Martin C Stumpe, et al. 2019. Human-centered tools for coping with imperfect algorithms during medical decision-making. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–14.
- Joel Chan, Joseph Chee Chang, Tom Hope, Dafna Shahaf, and Aniket Kittur. 2018. Solvent: A mixed initiative system for finding analogies between research papers. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–21.
- Joel Chan, Zijian Ding, Eesh Kamrah, and Mark Fuge. 2024. Formulating or Fixating: Effects of Examples on Problem Solving Vary as a Function of Example Presentation Interface Design. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.
- Joseph Chee Chang, Amy X Zhang, Jonathan Bragg, Andrew Head, Kyle Lo, Doug Downey, and Daniel S Weld. 2023. CiteSee: Augmenting Citations in Scientific Papers with Persistent and Personalized Historical Context. *arXiv preprint arXiv:2302.07302* (2023).
- Duen Horng Chau, Aniket Kittur, Jason I Hong, and Christos Faloutsos. 2011. Apolo: making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 167–176.
- Kiroong Choe, Seokweon Jung, Seokhyeon Park, Hwajung Hong, and Jinwook Seo. 2021. Papers101: Supporting the discovery process in the literature review workflow for novice researchers. In *2021 IEEE 14th Pacific Visualization Symposium (PacificVis)*. IEEE, 176–180.
- Bernard CK Choi and Anita WP Pak. 2006. Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. *Clinical and investigative medicine* 29, 6 (2006), 351.
- Max Coltheart. 1981. The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A* 33, 4 (1981), 497–505.
- Allen Foster. 2004. A nonlinear model of information-seeking behavior. *Journal of the American society for information science and technology* 55, 3 (2004), 228–237.
- Krzysztof Z Gajos and Lena Mamykina. 2022. Do people engage cognitively with AI? Impact of AI assistance on incidental learning. In *27th international conference on intelligent user interfaces*. 794–806.
- Luyu Gao, Xueguang Ma, Jimmy Lin, and Jamie Callan. 2022. Precise zero-shot dense retrieval without relevance labels. *arXiv preprint arXiv:2212.10496* (2022).
- Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
- Andrew Head, Kyle Lo, Dongyeop Kang, Raymond Fok, Sam Skjonsberg, Daniel S Weld, and Marti A Hearst. 2021. Augmenting scientific papers with just-in-time, position-sensitive definitions of terms and symbols. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.
- Jeffrey Heer. 2019. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences* 116, 6 (2019), 1844–1850.
- Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 159–166.
- Jill Jesson, Lydia Matheson, and Fiona M Lacey. 2011. Doing your literature review: Traditional and systematic techniques. (2011).
- Peiling Jiang, Jude Rayan, Steven P Dow, and Haijun Xia. 2023. Graphologue: Exploring large language model responses with interactive diagrams. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–20.
- 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.
- Hyeonsu B Kang, Rafal Kocielnik, Andrew Head, Jiangjiang Yang, Matt Latzke, Aniket Kittur, Daniel S Weld, Doug Downey, and Jonathan Bragg. 2022. From who you know to what you read: augmenting scientific recommendations with implicit social networks. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–23.
- Hyeonsu B Kang, Xin Qian, Tom Hope, Dafna Shahaf, Joel Chan, and Aniket Kittur. 2022. Augmenting scientific creativity with an analogical search engine. *ACM Transactions on Computer-Human Interaction* 29, 6 (2022), 1–36.
- Hyeonsu B Kang, Nouran Soliman, Matt Latzke, Joseph Chee Chang, and Jonathan Bragg. 2023. ComLittee: Literature Discovery with Personal Elected Author Committees. *arXiv preprint arXiv:2302.06780* (2023).
- Hyeonsu B Kang, Tongshuang Wu, Joseph Chee Chang, and Aniket Kittur. 2023. Synergi: A Mixed-Initiative System for Scholarly Synthesis and Sensemaking. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–19.
- Jeffrey A Knapp. 2012. Plugging the "whole": Librarians as interdisciplinary facilitators. *Library Review* 61, 3 (2012), 199–214.
- Dan Lahav, Jon Saad Falcon, Bailey Kuehl, Sophie Johnson, Sravanthi Parasa, Noam Shomron, Duen Horng Chau, Diyi Yang, Eric Horvitz, Daniel S Weld, et al. 2022. A search engine for discovery of scientific challenges and directions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 11982–11990.
- Michelle S Lam, Janice Teoh, James A Landay, Jeffrey Heer, and Michael S Bernstein. 2024. Concept Induction: Analyzing Unstructured Text with High-Level Concepts Using LLoM. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–28.
- Yoonjoo Lee, Hyeonsu B Kang, Matt Latzke, Juho Kim, Jonathan Bragg, Joseph Chee Chang, and Pao Siangliulue. 2024. PaperWeaver: Enriching Topical Paper Alerts by Contextualizing Recommended Papers with User-collected Papers. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.
- Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* 33 (2020), 9459–9474.
- Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, et al. 2022. Holistic evaluation of language models. *arXiv preprint arXiv:2211.09110* (2022).

- [40] Michael Xieyang Liu, Tongshuang Wu, Tianying Chen, Franklin Mingzhe Li, Aniket Kittur, and Brad A Myers. 2023. Selenite: Scaffolding Online Sensemaking with Comprehensive Overviews Elicited from Large Language Models. *arXiv preprint arXiv:2310.02161* (2023).
- [41] Nic Lupfer, Andruid Kerne, Andrew M Webb, and Rhema Linder. 2016. Patterns of free-form curation: Visual thinking with web content. In *Proceedings of the 24th ACM international conference on Multimedia*. 12–21.
- [42] Shuai Ma, Qiaoyi Chen, Xinru Wang, Chengbo Zheng, Zhenhui Peng, Ming Yin, and Xiaojuan Ma. 2024. Towards human-ai deliberation: Design and evaluation of llm-empowered deliberative ai for ai-assisted decision-making. *arXiv preprint arXiv:2403.16812* (2024).
- [43] Shuai Ma, Ying Lei, Xinru Wang, Chengbo Zheng, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2023. Who should i trust: Ai or myself? leveraging human and ai correctness likelihood to promote appropriate trust in ai-assisted decision-making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [44] Pablo N Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. 2011. DBpedia spotlight: shedding light on the web of documents. In *Proceedings of the 7th international conference on semantic systems*. 1–8.
- [45] Donald Metzler, Yi Tay, Dara Bahri, and Marc Najork. 2021. Rethinking search: making domain experts out of dilettantes. In *Acm sigir forum*, Vol. 55. ACM New York, NY, USA, 1–27.
- [46] Tim Miller. 2023. Explainable ai is dead, long live explainable ai! hypothesis-driven decision support using evaluative ai. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 333–342.
- [47] Lionel JB Mote. 1962. Reasons for the variations in the information needs of scientists. *Journal of Documentation* 18, 4 (1962), 169–175.
- [48] Jill Newby. 2011. Entering unfamiliar territory: Building an information literacy course for graduate students in interdisciplinary areas. *Reference & User Services Quarterly* 50, 3 (2011), 224–229.
- [49] OpenAI. 2024. New Embedding Models and API Updates. <https://openai.com/index/new-embedding-models-and-api-updates/>. Accessed on 7.22.2024.
- [50] 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.
- [51] Srishti Palani, Yingyi Zhou, Sheldon Zhu, and Steven P Dow. 2022. InterWeave: Presenting Search Suggestions in Context Scaffolds Information Search and Synthesis. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–16.
- [52] Carole L Palmer and Katrina Fenlon. 2010. Information research on interdisciplinarity. *The Oxford handbook of interdisciplinarity* (2010), 174–188.
- [53] Peter Pirolli and Stuart Card. 1999. Information foraging. *Psychological review* 106, 4 (1999), 643.
- [54] Peter Pirolli, Patricia Schank, Marti Hearst, and Christine Diehl. 1996. Scatter/gather browsing communicates the topic structure of a very large text collection. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 213–220.
- [55] Rui Qiu, Yamei Tu, Yu-Shuen Wang, Po-Yin Yen, and Han-Wei Shen. 2022. DocFlow: A Visual Analytics System for Question-based Document Retrieval and Categorization. *IEEE Transactions on Visualization and Computer Graphics* (2022).
- [56] Stephen Rowland. 2002. Overcoming fragmentation in professional life: The challenge for academic development. *Higher education quarterly* 56, 1 (2002), 52–64.
- [57] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. 2017. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)* 42, 3 (2017), 1–21.
- [58] Chirag Shah and Emily M Bender. 2022. Situating search. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*. 221–232.
- [59] Dafna Shahaf, Carlos Guestrin, and Eric Horvitz. 2012. Metro maps of science. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1122–1130.
- [60] Nikhil Sharma, Q Vera Liao, and Ziang Xiao. 2024. Generative Echo Chamber? Effects of LLM-Powered Search Systems on Diverse Information Seeking. *arXiv preprint arXiv:2402.05880* (2024).
- [61] Chuhan Shi, Weiwei Cui, Chengzhong Liu, Chengbo Zheng, Haidong Zhang, Qiong Luo, and Xiaojuan Ma. 2023. NL2Color: Refining color palettes for charts with natural language. *IEEE Transactions on Visualization and Computer Graphics* (2023).
- [62] Chuhan Shi, Fei Nie, Yicheng Hu, Yige Xu, Lei Chen, Xiaojuan Ma, and Qiong Luo. 2022. MedChemLens: An Interactive Visual Tool to Support Direction Selection in Interdisciplinary Experimental Research of Medicinal Chemistry. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 63–73.
- [63] Chuhan Shi, Yicheng Hu, Shenan Wang, Shuai Ma, Chengbo Zheng, Xiaojuan Ma, and Qiong Luo. 2023. RetroLens: A Human-AI Collaborative System for Multi-step Retrosynthetic Route Planning. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.
- [64] Neil R Smalheiser. 2017. Rediscovering Don Swanson: the past, present and future of literature-based discovery. *Journal of Data and Information Science* 2, 4 (2017), 43–64.
- [65] Kihoon Son, DaEun Choi, Tae Soo Kim, Young-Ho Kim, and Juho Kim. 2024. Genquery: Supporting expressive visual search with generative models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.
- [66] Don Spanner. 2001. Border crossings: Understanding the cultural and informational dilemmas of interdisciplinary scholars. *The Journal of Academic Librarianship* 27, 5 (2001), 352–360.
- [67] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. *arXiv preprint arXiv:2305.11483* (2023).
- [68] Don R Swanson. 1986. Undiscovered public knowledge. *The Library Quarterly* 56, 2 (1986), 103–118.
- [69] Don R Swanson and Neil R Smalheiser. 1996. Undiscovered Public Knowledge: A Ten-Year Update. In *Proceedings of the KDD*. 96, 295–298.
- [70] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. 990–998.
- [71] Jaime Teevan, Christine Alvarado, Mark S Ackerman, and David R Karger. 2004. The perfect search engine is not enough: a study of orienteering behavior in directed search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 415–422.
- [72] Richard Van Noorden et al. 2015. Interdisciplinary research by the numbers. *Nature* 525, 7569 (2015), 306–307.
- [73] Viswanath Venkatesh and Hillol Bala. 2008. Technology acceptance model 3 and a research agenda on interventions. *Decision sciences* 39, 2 (2008), 273–315.
- [74] Vijay Viswanathan, Kiril Gashevtovski, Carolin Lawrence, Tongshuang Wu, and Graham Neubig. 2023. Large language models enable few-shot clustering. *arXiv preprint arXiv:2307.00524* (2023).
- [75] Liang Wang, Nan Yang, and Furu Wei. 2023. Query2doc: Query expansion with large language models. *arXiv preprint arXiv:2303.07678* (2023).
- [76] Shuai Wang, Harrison Scells, Bevan Koopman, and Guido Zuccon. 2023. Can ChatGPT write a good boolean query for systematic review literature search?. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1426–1436.
- [77] Andrew M Webb, Andruid Kerne, Rhema Linder, Nic Lupfer, Yin Qu, Kade Keith, Matthew Carrasco, and Yvonne Chen. 2016. A free-form medium for curating the digital. *Curating the Digital: Space for Art and Interaction* 73–87.
- [78] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C Schmidt. 2023. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382* (2023).
- [79] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–22.
- [80] Chang Xiao. 2023. AutoSurveyGPT: GPT-Enhanced Automated Literature Discovery. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–3.
- [81] Zhengyuan Yang, Zhe Gan, Jianfeng Wang, Xiaowei Hu, Yumao Lu, Zicheng Liu, and Lijuan Wang. 2022. An empirical study of gpt-3 for few-shot knowledge-based vqa. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 3081–3089.
- [82] Yuwei Zhang, Zihan Wang, and Jingbo Shang. 2023. Clusterllm: Large language models as a guide for text clustering. *arXiv preprint arXiv:2305.14871* (2023).
- [83] Chengbo Zheng, Dakuo Wang, April Yi Wang, and Xiaojuan Ma. 2022. Telling stories from computational notebooks: AI-assisted presentation slides creation for presenting data science work. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–20.
- [84] Chengbo Zheng, Kangyu Yuan, Bingcan Guo, Reza Hadi Mogavi, Zhenhui Peng, Shuai Ma, and Xiaojuan Ma. 2024. Charting the Future of AI in Project-Based Learning: A Co-Design Exploration with Students. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.

A EXAMPLE OUTPUTS OF PROMPTS FOR EXPLORATORY QUESTION (EQ) GENERATION AND SEARCH QUERY EXPANSION

We design DiscipLink to address the *terminology gap* by generating simple EQs to ease sensemaking and search queries with specific

terminology to ensure search proficiency. In this appendix, we provide example outputs of the generated EQs and queries.

A.1 EQ Generation

DiscipLink uses EQs to help users explore unfamiliar fields by asking LLM to take expert personas and adding a condition in the prompt that the generated EQs need to be easy to understand. We tested prompts for EQ generation, including the one used in DiscipLink, and two ablation prompts: 1) a prompt without persona; 2) a prompt without the simplification condition. The prompts are presented in Table 1.

We used these prompts to generate three EQs under three different relevant disciplines for eight research topics (in total 72 EQs). Part of the results are presented in Table 2. We include all the results in the supplementary material. These results present hints that using persona makes EQs more concrete to domains and adding a simplification requirement enhances the comprehensibility.

A.2 Search Query Expansion

DiscipLink first generates pseudo-answers (PA) and then uses PA to prompt for queries (subsubsection 4.2.2). We tested query expansion with PA generation first and without PA. For each EQ generated in subsection A.1, nine search queries were generated. Part of the results are presented in Table 2 and all are included in the supplementary material. We also compared the concreteness of the generated queries of the two query expansion approaches (with and without PA) using the MRC Psycholinguistic Database [19]. The average concreteness ratings of the queries generated with PA is 319.56 (SD=52.37), and for the queries generated without PA is 292.42 (SD=39.69). A comparison between the two conditions regarding the generated queries of each input EQ using paired t-test shows the queries generated with PA is significantly more concrete than the ones generated without PA ($p < 0.01$). These results provide evidence on the addition of PA helps inject domain-specific terms into the queries.

Table 1: Tested prompts for EQ generation.

Condition	Prompt
w/o persona	<p>You are assisting a researcher who is looking for critical research questions studied in a certain field.</p> <p>The research idea they are interested in is: {research_idea}.</p> <p>Please provide a list of questions that:</p> <ul style="list-style-type: none"> - Are simple, and easy to understand. Are concise and do not exceed 15 words each. - Relate directly to the research idea. - Focus on concepts within the field, {field}. <p>Provide {num_rq} questions.</p>
w/o simplification	<p>You are an expert in {field}. A researcher is looking for critical research questions studied in your field.</p> <p>The research idea they are interested in is: {research_idea}.</p> <p>Please provide a list of questions that:</p> <ul style="list-style-type: none"> - Relate directly to the research idea. - Focus on concepts within the field, {field}. <p>The research idea involves multidisciplinary concepts, but you should provide questions that researchers in {field} have studied. Do not give questions that you, as a {field} expert, cannot answer confidently.</p> <p>Provide {num_rq} questions.</p>
DiscipLink	<p>You are an expert in {field}. A researcher is looking for critical research questions studied in your field.</p> <p>The research idea they are interested in is: {research_idea}.</p> <p>Please provide a list of questions that:</p> <ul style="list-style-type: none"> - Are simple, and easy to understand. Are concise and do not exceed 15 words each. - Relate directly to the research idea. - Focus on concepts within the field, {field}. <p>The research idea involves multidisciplinary concepts, but you should provide questions that researchers in {field} have studied. Do not give questions that you, as a {field} expert, cannot answer confidently.</p> <p>Provide {num_rq} questions.</p>

Research Field	Topic & Prompt	Generated EQs
The relationship between mobile phone use and mental well-being [Social Psychology]	w/o persona	<ul style="list-style-type: none"> - How does mobile phone use impact social interactions and mental well-being? - Does excessive phone use contribute to feelings of loneliness or depression? - Can mindful usage of mobile phones improve mental health outcomes?
	w/o simpl.	<ul style="list-style-type: none"> - How does the frequency of mobile phone use impact perceived social support and interpersonal relationships among different age groups? - What role does social comparison via social media on mobile phones play in affecting self-esteem and mental well-being? - How do patterns of mobile phone use influence feelings of loneliness and social connectedness in various social settings?
	DiscipLink	<ul style="list-style-type: none"> - How does mobile phone use impact social connectedness and isolation? - Do smartphone interactions influence users' self-esteem and self-worth? - Can excessive mobile phone use contribute to social anxiety or fear of missing out (FOMO)?
Promote sustainable travel choices among urban commuters [Transport Economics]	w/o persona	<ul style="list-style-type: none"> - How does pricing affect urban commuters' choice for sustainable travel methods? - What role does convenience play in urban commuters' decision to use public transport? - How do travel time comparisons influence commuters' preference for car vs. public transport?
	w/o simpl.	<ul style="list-style-type: none"> - What are the elasticity effects of pricing policies (e.g., congestion pricing or parking fees) on urban commuters' travel behavior towards more sustainable modes of transportation? - How do investments in public transit infrastructure economically impact commuters' propensity to shift from private car usage to public transportation in urban areas? - What are the cost-benefit implications of implementing incentives such as subsidies for electric vehicle purchases or biking to work programs on overall urban travel sustainability?
	DiscipLink	<ul style="list-style-type: none"> - How do fare subsidies impact urban commuters' adoption of public transport? - What are the economic benefits of reducing car dependency in cities? - How do congestion pricing policies influence commuter behavior and mode choice?
Facilitate community engagement in local environmental conservation efforts [Public Policy]	w/o persona	<ul style="list-style-type: none"> - How can public policy increase community participation in environmental conservation? - What policies effectively engage youth in local environmental efforts? - How do incentives in policies impact community-led conservation activities?
	w/o simpl.	<ul style="list-style-type: none"> - What policy mechanisms are most effective in incentivizing local communities to participate in environmental conservation efforts? - How do local governance structures impact the level of community engagement in environmental conservation programs? - What role does public policy play in building partnerships between local governments, non-profit organizations, and community groups to support environmental conservation initiatives?
	DiscipLink	<ul style="list-style-type: none"> - How can policy frameworks enhance public participation in local conservation programs? - What incentives can encourage local communities to engage in conservation efforts? - Which governance structures best support community-led environmental conservation initiatives?

Table 2: EQs generated by GPT-4 with different prompts. The GPT-4 version we used is GPT-4-0125.

EQ	Condition	Generated Queries
Does heavy social media use affect code-switching behaviors in multilingual speakers?	w/o PA	<ol style="list-style-type: none"> "heavy social media use code-switching multilingual speakers" "social media impact multilingual code-switching" "code-switching frequency social media multilingualism" "effects of social media on bilingual code-switching" "digital communication code-switching multilingual" "online interaction code-switching behavior multilingual speakers" "social media language switching bilingualism" "social media multilingual communication behavior" "impact of social networks on code-switching"
	w/ PA	<ol style="list-style-type: none"> "social media influence code-switching multilingual" "code-switching normalization social media environments" "multilingualism social media language practices" "identity negotiation code-switching social media" "digital communication multilingual audience design" "social contexts code-switching social media" "social capital code-switching online communities" "strategic code-switching social media multilingual" "building online communities multilingual code-switching"
How effective are price incentives in shifting commuter preferences towards sustainable travel methods?	w/o PA	<ol style="list-style-type: none"> "price incentives commuter preferences sustainable travel" "financial incentives sustainable transportation" "commuter behavior price incentives sustainable travel" "effectiveness of price incentives on sustainable commuting" "incentives for sustainable travel modal shift" "price reductions sustainable transportation methods" "commuter choices financial incentives sustainable travel" "economic incentives public transportation uptake" "reward systems for sustainable commuting"
	w/ PA	<ol style="list-style-type: none"> "Price incentives commuter preferences sustainable travel" "Discounted fares impact sustainable commuting" "Subsidies cost-benefit analysis commuting" "Behavioral economics nudges sustainable commuting" "Price incentives behavioral change transportation" "Environmental nudges commuting behavior" "Longitudinal effectiveness price incentives sustainable transport" "Infrastructure public awareness sustainable commuting" "Complementary measures price incentives transportation"
What ethical guidelines should govern the use of robots in elderly care?	w/o PA	<ol style="list-style-type: none"> "ethical guidelines robots elderly care" "robotics ethics senior care" "AI ethics in elderly care" "robot caregivers ethical considerations" "elderly care robotics ethical standards" "robot use in elder care ethical issues" "ethics of humanoid robots in eldercare" "moral guidelines for caregiving robots" "ethical framework for AI caregivers"
	w/ PA	<ol style="list-style-type: none"> "patient autonomy elderly robot care" "dignity human-robot interaction elder care" "ethical design robots elder care" "privacy elderly care robots" "data security ethical concerns elder care" "surveillance data ethics elderly" "social well-being companion robots elderly" "emotional support robots mental health elderly" "human-robot interaction elderly acceptance"

Table 3: Queries generated by GPT-4 with or without PA. The GPT-4 version is GPT-4-0125.