

PAPERBRIDGE: Crafting Research Narratives through Human-AI Co-Exploration

Runhua Zhang

The Hong Kong University of Science and Technology
Hong Kong, China
runhua.zhang@connect.ust.hk

Yang Ouyang

ShanghaiTech University
Shanghai, China
ouyy@shanghaitech.edu.cn

Leixian Shen

The Hong Kong University of Science and Technology
Hong Kong, China
lshenaj@connect.ust.hk

Yuying Tang

The Hong Kong University of Science and Technology
Hong Kong, China
yuying.tang@connect.ust.hk

Xiaojuan Ma

The Hong Kong University of Science and Technology
Hong Kong, China
mxj@cse.ust.hk

Huamin Qu

The Hong Kong University of Science and Technology
Hong Kong, China
huamin@cse.ust.hk

Xian Xu*

Lingnan University
The Hong Kong University of Science and Technology
Hong Kong, China
xianxu0523@gmail.com

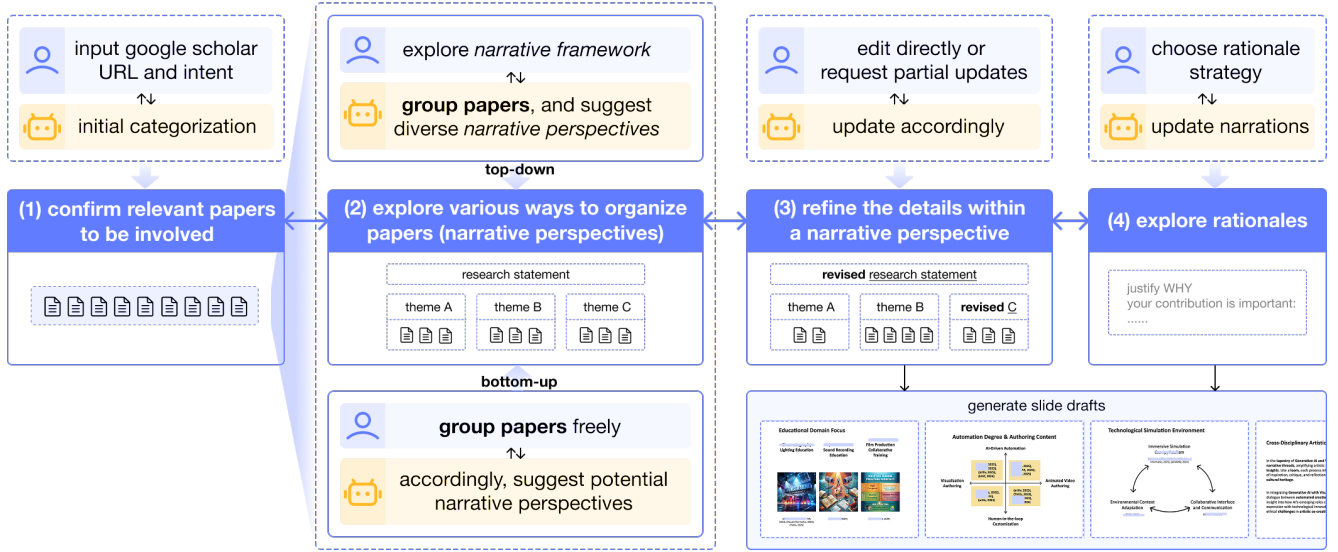


Figure 1: Human-AI co-exploration of alternative research narratives using PAPERBRIDGE. (1) Users begin by selecting relevant papers they wish to organize into a coherent narrative. (2) PAPERBRIDGE supports co-exploration of multiple narrative perspectives through both bottom-up and top-down approaches, allowing dynamic re-grouping of papers. (3) Once a promising narrative perspective is identified, users refine the details with PAPERBRIDGE’s assistance. (4) Finally, users examine rationale strategies to justify the significance of their perspective. In this workflow, users can move forward or backward flexibly, with slide drafts as the outputs of the exploration.

*Corresponding author

ABSTRACT

Researchers frequently need to synthesize their own publications into coherent narratives that demonstrate their scholarly contributions. To suit diverse communication contexts, exploring alternative ways to organize one's work while maintaining coherence is particularly challenging, especially in interdisciplinary fields like HCI where individual researchers' publications may span diverse domains and methodologies. In this paper, we present PAPERBRIDGE, a human-AI co-exploration system informed by a formative study and content analysis. PAPERBRIDGE assists researchers in exploring diverse perspectives for organizing their publications into coherent narratives. At its core is a bi-directional analysis engine powered by large language models, supporting iterative exploration through both top-down user intent (e.g., determining organization structure) and bottom-up refinement on narrative components (e.g., thematic paper groupings). Our user study (N=12) demonstrated PAPERBRIDGE's usability and effectiveness in facilitating the exploration of alternative research narratives. Our findings also provided empirical insights into how interactive systems can scaffold academic communication tasks.

CCS CONCEPTS

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

KEYWORDS

Academic Communication, Storytelling, Human-AI Collaboration

ACM Reference Format:

Runhua Zhang, Yang Ouyang, Leixian Shen, Yuying Tang, Xiaojuan Ma, Huamin Qu, and Xian Xu. 2025. PAPERBRIDGE: Crafting Research Narratives through Human-AI Co-Exploration. In *The 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25), September 28–October 1, 2025, Busan, Republic of Korea*. ACM, New York, NY, USA, 21 pages. <https://doi.org/10.1145/3746059.3747713>

1 INTRODUCTION

Academic communication plays a vital role in enabling researchers to disseminate findings, establish their scholarly identity, and cultivate collaborative relationships within their academic communities [1, 17, 22, 26, 51]. A key facet of this communication is the strategic organization of one's academic work—particularly publications—into coherent **research narratives** that clearly demonstrate a researcher's contributions to the field [25, 47]. Given the wide range of communication scenarios, from keynote speeches to dissertations and public outreach, researchers often need to craft **alternative narratives** tailored to specific audiences (such as academic departments, industry labs, or funding agencies) and communication goals, including job talks, grant proposals, and public engagement [14].

However, crafting alternative research narratives can be challenging. In addition to the need to tailor communication for different contexts, interdisciplinary fields like Human-Computer Interaction (HCI) add complexity, as research often spans diverse epistemologies, methodological traditions, and types of contributions [64]. For instance, even when a researcher's work centers on a single

area such as creativity support, their projects may cut across multiple dimensions, from empirical studies to system development [8], while addressing various stages of creative practice (e.g., ideation, implementation, iteration) across domains like video production, animation, design, and writing.

Moreover, unlike simply summarizing individual papers, constructing a research narrative requires identifying thematic structures across a body of work at an appropriate level of abstraction—one that is neither too broad nor too narrow [13, 45]. This can be particularly challenging for early-career researchers, who are just starting to shape their research trajectory. In addition, as a researcher's publication record grows, this task increasingly demands reframing and reorganizing their work to form a coherent whole [33]. Despite these challenges, little attention has been dedicated to supporting researchers in synthesizing their own publications. This motivates our investigation into how to support narrative exploration, especially in the context of HCI.

To address this gap, we conducted a formative study involving six early-career HCI researchers to examine their current practices, key challenges, and expectations for potential support tools. We identified three essential elements in narrative construction: (1) **Narrative frameworks** that shape the overall flow or arc of the story, determining how subtopics are connected; (2) **Narrative perspectives** that provide specific lenses for highlighting contributions, typically consisting of a central contribution statement and associated thematic clusters that group related papers; and (3) **Narrative rationale** that explains and justifies why a particular perspective was chosen. Among these, the first two elements, which influence or are influenced by how papers are organized, posed the greatest challenges for participants. Based on these findings and their expectations, we derived four design considerations.

Before jumping into the design of PAPERBRIDGE, we first conducted a content analysis of 53 publicly available HCI research talks to identify common practices that would help address our design considerations. This analysis revealed a design space with two key components: (1) Four common narrative frameworks: namely *parallel*, *linear*, *circular*, and *coordinate*. Each offers distinct logic for structuring relationships among publications, and (2) a set of storytelling strategies for the narrative rationale for justifying the significance of one's contribution.

Informed by design considerations from our formative study and the design space derived from content analysis, we present PAPERBRIDGE, a human-AI co-exploration system that supports HCI researchers in exploring diverse perspectives to craft their research narratives. As shown in Figure 1, the system implements a workflow that begins with researchers selecting the papers they wish to organize. Powered by a bi-directional analysis engine, PAPERBRIDGE enables exploration of diverse narrative possibilities through both top-down and bottom-up approaches. In top-down exploration, researchers start by selecting a narrative framework, which guides the generation of multiple narrative perspectives—each composed of a contribution statement and thematic clusters with assigned papers. In bottom-up exploration, users can manually group papers and request matched perspectives based on their groupings. In both approaches, users can iteratively refine the system's suggestions, either by editing directly or requesting partial updates, until they are satisfied with the resulting narratives. Finally, PAPERBRIDGE also

assists in articulating rationale to justify the chosen framing, and generates slide drafts that can be further revised or disseminated.

We evaluated PAPERBRIDGE with 12 HCI researchers from various sub-domains. Participants responded positively to its usability and reported overall engaging experiences with PAPERBRIDGE's co-exploration workflow. Our findings show that PAPERBRIDGE not only improved the efficiency of exploring alternative research narratives, but also revealed how it scaffolded the co-exploration process, from initial trust establishment, to framework-based exploration, and engagement with keyword-driven perspective suggestions.

The contributions of this work are three-fold:

- We conduct a formative study to gain a comprehensive understanding of how research narratives are constructed, leading to a design space encompassing four narrative frameworks and a set of rationale strategies.
- We present PAPERBRIDGE, a human-AI co-exploration system that facilitates diverse research narrative exploration, which features a bi-directional analysis engine that enables both top-down and bottom-up approaches to narrative construction.
- Through a user study with 12 HCI researchers, we demonstrate PAPERBRIDGE's usability and reveal the co-exploration processes. We also discuss the implications for future academic communication support.

2 RELATED WORK

In this section, we review prior work that supports academic profiling, research story construction for communication, and literature sensemaking. We articulate how our work builds on these efforts while identifying the gaps that motivate our research.

2.1 Academic Communication

Effective academic communication is fundamental to advancing research fields through knowledge dissemination and perspectives exchange [17]. Researchers engage in this communication through various channels, including publishing papers, delivering conference talks, participating in academic workshops and seminars, and utilizing social media platforms [14]. Among these scenarios, presenting and profiling one's research trajectory, narratives, or impact, plays a key role in building academic identity and establishing one's position within a research community [1, 25].

To facilitate researchers' academic communication, prior research has explored constructing researcher profiles and stories by organizing personal and academic information, such as affiliations, education, and publication records [5, 72]. Various visualization systems have also been developed to represent researcher-related metrics, including publication volume, citation counts, collaboration networks, and topic evolution [29, 30, 37, 57, 69]. For example, systems such as Citeology [37], PivotPaths [15], PivotSlice [73], and PaperLens [32] enable multi-dimensional exploration of scholarly data, while CiteVis [58], CiteRivers [23], and Wu et al.'s career trajectory visualization [70] provide overviews of research trends.

While these systems support researchers in promoting their academic impact through vivid, quantitative data-driven storytelling, they primarily focus on showcasing research breadth and external influence. On the one hand, this approach may limit researchers with fewer publications from effectively leveraging such tools. On

the other hand, these systems offer limited support for helping researchers retrospectively reflect on their own publications to build research narratives grounded in the content and context of their work, which is a crucial process for demonstrating the depth of contribution within a particular field.

Although prior research has begun to explore content-level analysis of individual papers to support scholarly communication [59], few systems assist researchers in constructing a coherent and holistic reflection across their body of work. Yet, this process is critical in many academic scenarios, such as invited talks, job talks, seminars, and dissertations, where researchers may be expected to reorganize their work, identify meaningful connections, and tailor their narratives for different audiences and purposes [25, 46]. Motivated by this gap, our work aims to explore how to support researchers in constructing their research narratives or profiles through the specific content and context of their own publications.

2.2 Literature Synthesis and Sensemaking

Constructing research narratives based on one's own publications fundamentally relies on literature organization and information synthesis. While most prior research in HCI has focused on supporting literature sensemaking for external publications, *i.e.*, helping researchers explore, interpret, and synthesize knowledge scattered across others' papers, many of the underlying ideas and approaches also inspire our current work.

Broadly, prior studies on literature sensemaking can be categorized into two major threads. The first line of work focuses on fact-based information extraction and synthesis. Early systems often leveraged citation networks to help researchers identify relationships and patterns among a collection of papers [11, 20, 44]. For example, PaperQuest [43] utilized citation structures to assist literature review tasks. CiteSee [9] augmented PDF reading experiences by highlighting citation relationships in context. In addition, tools such as Threddy [27] and Synergi [28] were designed to help users collect, organize, and understand research threads across multiple papers. Beyond citation structures, recent work also extracted factual content within papers, including tables, images, and textual claims, to enable comprehensive literature synthesis [66]. Another line of work has begun to leverage large language models (LLMs) to synthesize paper content and help researchers contextualize the relationships among papers. For example, PaperWeaver [34] generates synthesized descriptions to support paper recommendations, while DiscipLink [74] enables interdisciplinary literature exploration by connecting concepts across fields.

Across these systems, a key insight is the combination of top-down and bottom-up workflows in supporting scholarly synthesis. Bottom-up approaches often help users incrementally grow their understanding of relevant literature by exploring specific papers, documents, or keywords [27, 31, 41, 43]. In contrast, top-down strategies aim to provide a high-level structural overview of a research landscape, enabling users to quickly grasp the organization of a domain [49]. A trend emerged and showed that the mixed-initiative approaches were explored [28, 74]. For example, Kang et al.'s Synergi integrated both top-down and bottom-up approaches and further discusses how users' prior knowledge influences their preferences for top-down or bottom-up workflows [28].

Building on the insights from prior work, we adopt both top-down and bottom-up approaches to support researchers in constructing narratives from their own publications. On the one hand, researchers are often highly familiar with their own work, and therefore benefit from open-ended exploration that allows them to revisit and reflect on their ideas in flexible ways. On the other hand, the process of abstracting, reorganizing, and synthesizing familiar content into coherent research stories remains challenging. It may require top-down guidance or structural directions to help researchers identify meaningful patterns. Though existing systems for literature sensemaking have significantly advanced our understanding of information synthesis for text-heavy content [40], they primarily focus on external literature exploration and comprehension. In contrast, constructing research narratives from one's own publications requires a different kind of sensemaking. Our work aimed at addressing this gap.

3 FORMATIVE STUDY

We began by investigating how HCI researchers organize their research narratives, how they find alternatives, the challenges they face in current practices, and their expectations for a potential assistive tool, thereby identifying the design considerations.

3.1 Methods

3.1.1 Participants. Our formative study recruited six HCI researchers. Five participants were final-year PhD students (P1–P5), and one was a newly appointed postdoctoral fellow (P6). Their research backgrounds span diverse HCI subfields, including human–AI collaboration for creativity, data storytelling, visualization, virtual reality, HCI for science, and social computing. All participants had experience both attending and delivering academic talks at the time of the study.

3.1.2 Procedures. Each participant engaged in a semi-structured interview, including open-ended questions such as: “Please describe your typical workflow to prepare your research narratives/stories delivered in academic occasions, such as giving talks”, “What are the considerations when preparing your research narratives?”, and “What challenges have you encountered when preparing such stories, and how did you address them?” Each session lasted around 40 minutes.

3.1.3 Data Analysis. We employed an open-coding approach [24] to inductively explore participants' current practices and challenges. Two authors independently reviewed the transcripts to generate initial codes, with subsequent discussions to resolve discrepancies and ensure coding consistency. Examples of the resulting codes include “research narratives' components”, “paper organization methods”, and “storytelling intent”. During the analysis, two professors with more than 15 and 20 years of experience in the HCI field, respectively, provided expert guidance and contributed to the establishment of analytical criteria that informed the organization and interpretation of the findings.

3.2 Findings

Our findings are organized into two main categories: (1) a hierarchical structure with three key elements for constructing research

stories, and (2) key challenges identified in current practices, particularly associated with paper organization.

3.2.1 Key Elements for Research Narratives and Corresponding Challenges. We identified three essential elements that HCI researchers consider when constructing research narratives: **narrative framework**, **narrative perspective**, and **narrative rationale**. Each of these elements plays a distinct role in how researchers organize and present their work. Our analysis also surfaced corresponding challenges that arise when researchers engage with these elements. Below, we describe each element and its associated challenge.

(1) Narrative Framework refers to the overarching structure or “flow” through which a researcher's body of work is connected and presented. When crafting research narratives, participants often began by identifying central topics and reflecting on questions such as: “What is the main storyline across my work?” or “Can my papers be structured as addressing parallel challenges, or as a staged progression over time?” While some research trajectories lend themselves naturally to linear structures, many researchers actively experimented with alternative framings. For example, work on human–AI collaboration might be organized as a methodological journey (e.g., from identifying needs, to designing techniques, to system evaluation) or as an exploration of diverse contexts (e.g., creativity, decision-making, analytics).

C1: Exploring Alternative Narrative Frameworks. Participants shared that exploring different structural framings often meant rethinking how their papers related to one another. For example, P5 shared: “Previously, I organized my papers into three clusters defined by different challenges. But then I saw others use staged progressions and tried that too. It didn't quite work with my existing ideas.” Regarding this, the current approaches to discovering inspirations for relationships rely heavily on examining existing examples. P6 described searching through recorded talks to find models for research organization but noted that “such a process was quite time-consuming and inefficient.”

(2) Narrative Perspective. While frameworks outline a high-level structure, perspectives provide specific lenses or angles for spotlighting the researcher's contributions. This statement typically synthesizes the fundamental research objective while highlighting the approach that might distinguish their work from others in the field. A typical perspective comprises:

- **A key contribution statement** (e.g., “My research advances human–AI collaboration by developing adaptive interaction techniques that support different collaborative scenarios”), which succinctly articulates the main takeaway or value proposition.
- **Thematic clusters grouping individual papers** (e.g., “Interaction techniques for creative tasks,” “Collaborative systems for decision-making,” “AI-assisted analytical workflows”). These clusters serve as concrete realizations of the contribution statement, each illustrating a particular dimension, approach, or facet of the overall claim.

The components of a narrative perspective and their relationships can be found in [Figure 2](#).

C2: Exploring Alternative Narrative Perspectives. Regarding narrative perspectives, participants consistently emphasized the interconnected challenges of articulating high-level contribution

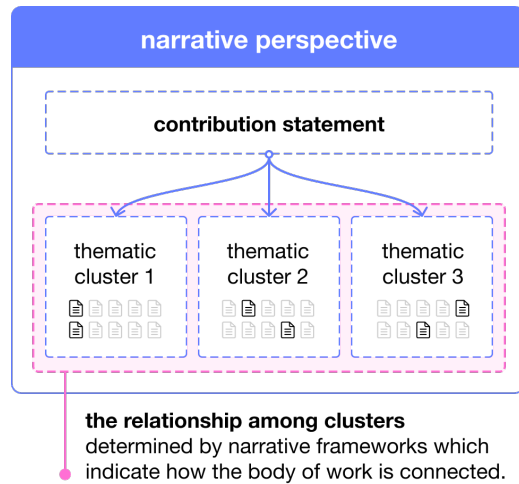


Figure 2: A narrative perspective is composed of three types of components: a contribution statement, thematic clusters, and individual papers, which are organized within the clusters to articulate the overarching themes.

statements and organizing papers into meaningful thematic clusters. These reflect the core of research organization: contribution statements emerge from paper groupings, while effective groupings depend on clear contribution statements. Regarding contribution statements, participants struggled to capture unique intellectual contributions rather than merely describing research topics. P1, a final-year PhD student, explained: “I know my research focus is to support creativity by supporting animation creation, and I am aware that each of my papers contributes to it. But I don’t want to only emphasize my research is about creativity support, but how exactly my research contributes to it.”

Simultaneously, participants struggled to develop thematic clusters that moved beyond surface-level categorizations to reveal deeper insights. P6 shared: “There is a thread of my research regarding VR systems developed to understand different teaching and learning behaviors in such environments. Grouping my research by identifying the teaching content is simple, but I think it is somewhat too superficial to be insightful. However, I have not figured out how to group them from angles with more insightful concepts.” This difficulty becomes particularly acute when synthesizing work for comprehensive presentations, as P1 noted: “To organize my thesis, I really want to abstract some more high-level concepts and form a line from my publications that will be presented in the thesis. Until now, I haven’t figured it out even though I’ve talked a lot to my peers.”

To address these interwoven challenges, participants reported seeking advice from peers or supervisors. However, these approaches were often considered inflexible, since they depended on others’ availability and were not well suited for iterative or self-paced narrative exploration. Some attempted to use tools like large language models (such as ChatGPT) for assistance. Yet these general-purpose tools often required extensive manual refinement, and were commonly found to produce overly generic suggestions that lacked depth and contextual sensitivity, largely because they do not incorporate specific understanding of research narrative practices.

(3) Narrative Rationale. The third element is the rationale, which provides justification for why the chosen perspective matters in the broader research context. It addresses questions like: “Why is developing adaptive interaction techniques for human-AI collaboration important to HCI?” The rationale connects individual contributions to research challenges or social impact, helping researchers articulate not just what they did, but why their approach and findings are significant to advancing the field.

C3: Adapting Narrative Rationale Accordingly. Participants noted that when shifting between different perspectives, they needed to adjust their rationale to align with the new framing. As P4 explained: “Depending on the scenario, I might emphasize the outputs of my visualization tools or the technical pipeline. Each choice requires me to rethink how I justify the significance of my contributions.” This challenge highlights the difficulty of not only selecting a narrative but also articulating its relevance in ways that resonate with different contexts (e.g., job talks, theses, research statements).

3.2.2 Expectations on An Assistive Tool. Our analysis also helped us identify three key expectations for an assistive tool that supports research narrative construction:

Spark Inspirations, Not Complete Stories. Participants expressed interest in tools that could provide inspirations for crafting narratives while preserving their agency in research interpretation. P4 described wanting “automated suggestions for different organizational approaches, so that I can find different perspectives to understand my research.” This sentiment was balanced with a strong desire to maintain control over narrative development. As P2 explained: “When discussing with peers or supervisors, they provide inspirations, and I take the decision on how to interpret their suggestions to form my research story, as I am the one who is most familiar with my research.”

Clear yet Flexible Workflow. All participants characterized narrative crafting and paper organization as non-linear and often chaotic processes, expressing a desire for tools that could help streamline their workflow. P1 shared: “I expected that the tool can help me with a clear workflow, so that I would not get lost in the ideas or thoughts.” However, participants emphasized that exploration should not come at the cost of flexibility. P4 noted: “It should give me some freedom to refine its results and test my own ideas when organizing papers.”

Integration with Existing Workflows. Beyond serving as a thinking aid, participants sought tools that could generate practical outputs compatible with their academic work. “Being able to transform these narrative ideas directly into presentation slides, or at least some images would save significant time,” noted P3. This reflected a desire for tools that could bridge the gap between narrative conceptualization and academic deliverables.

3.3 Design Considerations

Based on the elements, challenges, and expectations identified in our formative study, we propose four design considerations for supporting HCI researchers in exploring paper organization and crafting research narratives, with particular emphasis on facilitating explorations and inspiration to meet their expectations:

DC1: Supporting Exploration of Diverse Narrative Frameworks. To address C1, we aim to provide various narrative framework patterns that inspire researchers to explore different logical approaches to organizing their papers.

DC2: Facilitating Alternative Narrative Perspective Discovery. To address C2, we aim to support researchers in exploring diverse perspectives of their work. Besides, considering the expectations on flexible exploration, we aim to support them in iteratively refining the system's suggestions or testing.

DC3: Supporting Narrative Rationale Development. As it is challenging to adapt the narrative rationale for different contexts (C3), we aim to assist researchers in articulating the significance of their chosen perspectives in different ways.

DC4: Generating Adaptable Visual Outputs. To meet researchers' expectations for practical outputs, we aim to provide editable presentation slides that help researchers transform their narrative ideas into practical academic deliverables.

4 CONTENT ANALYSIS

As our formative study noticed that academic talks serve as a primary source of inspiration for research narratives, we conducted a content analysis of public HCI guest talks to address our design considerations, especially DC1 and DC3. Through this analysis, we examined (1) different approaches to organizing papers into narrative frameworks and (2) rationale strategies for justifying the significance of contribution statements.

4.1 Methods

4.1.1 Dataset. We assembled a dataset of 53 publicly accessible job talks and guest seminars¹ hosted by prominent HCI communities, such as the CMU HCII Seminar Series, the UW CS Seminar Series, Stanford Online, etc. These sources invite researchers from various cultural and research backgrounds, ensuring that our corpus captures a broad spectrum of HCI research domains. Our inclusion criteria were: (1) The speaker had an academic affiliation rather than an industry role; (2) The talk's content drew upon a series of publications, and (3) The talk video included visual presentation materials (e.g., slides).

4.1.2 Preliminary Explorations. We began with a pilot analysis of 10 talk videos to refine our focus. This initial phase helped us identify the part where researchers build the “big picture” of their research stories. This part usually appeared in the first 10–15 minutes of the talk and included presenting the research topics and motivation, stating the contribution perspective (or proposing key research questions or gaps), and showing how their publications supported these ideas. We found this part more relevant to our study than the detailed discussion of individual projects. Therefore, we decided not to analyze the project sections in detail, as our main goal was to understand how researchers construct the “big picture” of their research story.

In addition, we observed diverse approaches to constructing narrative frameworks. According to insights from our formative study, the key of such frameworks lies in how researchers articulate

the connections among their publications to collectively address a shared research problem. Building on this understanding, our observations identified recurring framework patterns, such as illustrating how individual works tackled distinct facets of a broader agenda, or how later publications built upon earlier ones. Based on these patterns, our analysis further examined the underlying relationships among publications as framed by the researchers.

4.1.3 Procedures. Following the pilot, two authors independently coded the talk videos using an inductive thematic analysis approach, focusing on the two focal points outlined earlier and the preliminary insights collected. The disagreements were resolved through rounds of discussion. Finally, two senior HCI researchers were invited to be involved in the discussion to refine the patterns and strategies that had emerged.

4.2 Findings

4.2.1 Four Patterns for Narrative Frameworks. Our findings first addressed DC1 by identifying the common narrative frameworks. As shown in Figure 3, we identified four patterns used to organize and connect publications: *parallel*, *linear*, *coordinate*, and *circular*. Each pattern shaped a distinct narrative arc within research trajectories. The visual representations of these patterns were informed by the tree and matrix diagrams that researchers employed. Specifically, the publications were organized into different thematic clusters, and the different relationships among these clusters define each pattern. Below, we define each of them, describe its characteristics, and provide examples.

Parallel Structure: Multiple Facets of Contribution. The parallel structure (Figure 3-i) represents the most common approach (43 out of 53) used by HCI researchers to organize their papers. In this structure, thematic clusters show a non-overlapping relationship and represent distinct technical or methodological approaches or different perspectives for formulating a problem space that collectively addresses a central research challenge. For example, a researcher investigating *accessible design* might organize papers into clusters focused on *vision-based accessibility*, *audio accessibility*, and *haptic feedback*. Each represents a different sensory modality but together serving the main topic.

Linear Structure: Progressive Development. Linear structures (Figure 3-ii) appeared five times in our clusters (5/53). In this structure, the thematic clusters illustrate a sequential flow where papers are organized in a progressive line, highlighting how each stage builds upon insights from previous stages. For instance, a researcher studying *embodied interaction* might organize papers into clusters showing progression from *sensing technologies* to *interaction techniques* to *application implementations*. The linear structure is ideal for showcasing how initial explorations led to sophisticated outcomes or how works span different levels of advancement.

Coordinate Structure: Comparative Dimensions. In coordinate structures (4 out of 53, Figure 3-iii), thematic clusters are positioned within a conceptual space defined by key dimensions that illustrate fundamental tensions in the HCI community, such as “*user control vs. automation*” or “*expressiveness vs. efficiency*.” This structure operationalizes what HCI researchers often refer to as a design space. The coordinate structure helps research statements emphasize the exploration of design spaces or tensions between

¹See <https://docs.google.com/spreadsheets/d/1DfnBhTopjPVJyibi6dTsiTnKkhj8zSZ/edit?usp=sharing&ouid=102936865702623175143&rtfpof=true&sd=true> for 53 public HCI talks used in our analysis

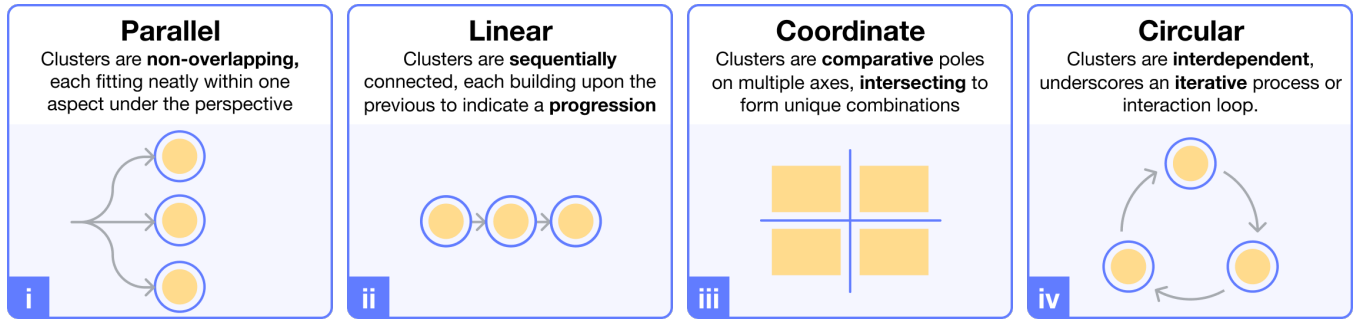


Figure 3: Four common narrative frameworks identified from our content analysis. Each framework describes a distinct way of organizing and connecting a researcher's body of work around a central storyline.

competing factors, providing a systematic way to present how works exist at the intersections between concepts.

Circular Structure: Iterative Refinement. In circular structures (1 out of 53, Figure 3-iv), each thematic cluster represents an interconnected phase that continuously informs others, serving the research statement through iterative processes or feedback loops. The circular structure is particularly valuable for research grounded in iterative processes, design thinking methodologies, or user-centered development cycles. It effectively communicates how different facets of research mutually reinforce each other and how complex problems require recursive refinement.

4.2.2 Rationale Strategies. In addition to the four common narrative frameworks, we also identified a comprehensive set of rationale strategies for justifying why a researcher's contribution statement is important to a domain.

As presented in Table 1, the strategies were categorized according to Aristotle's rhetorical framework [3] of **ethos**, **pathos**, and **logos**, which researchers strategically deploy to justify their work. **Ethos** refers to building credibility and trust — convincing audiences that the authors are knowledgeable and their work is grounded in recognized expertise. For establishing ethos, researchers utilized literature endorsement (8 instances), highlighted industry/academic attention (2), and incorporated quotations from prominent figures in the field (6). **Pathos** appeals to emotion and shared values — helping audiences connect with the problem on a personal or societal level. To engage pathos, researchers shared personal experiences (3), referenced relevant social events (5), aligned with common values and ethical considerations (7), and emphasized public demand or popularity (1). **Logos** relies on logical reasoning and evidence — using facts, data, and analysis to demonstrate the importance of the problem. Logos-based strategies included presenting data with visualizations (7), identifying historical patterns and trends (8), demonstrating tangible impact (9), and articulating the consequences of inaction (1). It is worth noting that researchers frequently employed multiple strategies simultaneously to justify their contribution, resulting in a total count of more than 53. The detailed definitions and examples of each strategy are provided in Table 1. This collection of storytelling strategies informs our approach to addressing DC3.

5 PAPERBRIDGE

Based on the insights from our formative study and content analysis, we present PAPERBRIDGE, a human-AI co-exploration system to inspire HCI researchers to craft their research narratives by supporting them in exploring diverse perspectives to organize their publications. Specifically, we first introduce PAPERBRIDGE's design and the corresponding design considerations through a system overview. Then, we present a detailed usage scenario to walk through how researchers can interact with PAPERBRIDGE. Finally, we describe the backbone of PAPERBRIDGE, a bi-directional analysis engine that enables both top-down and bottom-up explorations.

5.1 Overview

As shown in Figure 4, PAPERBRIDGE consists of three main components that collectively address four key design considerations (DCs): Paper Management (Panels A & B, left view), Narrative Exploration (Panels C, D, and E, center view), and Slide Draft Preview (Panel F, right view).

The workflow begins with paper **selection stage** in Panels A and B (Figure 4). After entering the user's Google Scholar URL and specifying their narrative intent, PAPERBRIDGE automatically retrieves their publications and generates initial, topic-based categorizations. This allows researchers to browse their publication corpus efficiently and iteratively refine both the selected papers and the category labels to align with their intended narrative scope.

Once users have chosen the papers they wish to include, they move to the Narrative Exploration area in the center (Panels C, D, and E, Figure 4).

Narrative Framework Exploration in Panel C (DC1). This panel presents four common narrative frameworks (*parallel*, *linear*, *circular*, and *coordinate*) identified in subsection 4.2.1. By switching among these frameworks, users can explore different high-level structures to see how their selected publications interconnect.

Narrative Perspectives Exploration in Panel D (DC2). Selecting a framework triggers PAPERBRIDGE's analysis engine to produce "sparks": potential narrative perspectives that fit the chosen framework. Each spark includes a candidate contribution statement and one or more thematic clusters, giving users alternate angles for framing their research story. The sparks are designed to be presented in a keyword format, encouraging users to interpret the narratives independently first.

Strategy	Method and Definition
Ethos (establish trustworthiness and expertise)	Literature Endorsement: Cite established research that underscores the importance of this gap. Industry/Academic Attention: Show how academia or industry prioritizes this topic through funding, discussions, or publications. Big Name's Quote: Use expert endorsements to reinforce the significance of the problem.
Pathos (appeal to emotions and values)	Personal Experience: Share a real-life story to make the issue more relatable and engaging. Social Events: Link the research to societal, political, or cultural events to highlight its relevance. Common Values & Ethics: Appeal to widely shared moral or ethical principles to justify the importance of this research. Public Demand & Popularity: Highlight growing public interest or widespread adoption to demonstrate the topic's timeliness.
Logos (appeal to logic and evidence)	Data & Visualization: Use statistics or visual evidence to illustrate the scale or urgency of the problem. Historical Patterns & Trends: Show how this issue fits within broader technological advancements or ongoing trends. Demonstrating Impact: Emphasize the tangible, long-term benefits of addressing this problem. Consequences of Inaction: Highlight risks, negative outcomes, or missed opportunities if the problem is ignored.

Table 1: Rationale strategies used by researchers to justify the importance of their contributions. These are categorized following Aristotle's rhetorical modes: *ethos* (credibility), *pathos* (emotional appeal), and *logos* (logical reasoning).

The interface is divided into four main sections:

- User Input:** Contains fields for Google Scholar URL, Organization Intent, and a Paper List. The Paper List includes a search bar and a list of papers with checkboxes for selection.
- Narrative Frameworks:** Displays four frameworks: Parallel (multidimensional, non-overlapping), Linear (sequential, progressive), Circular (cyclical, interconnected), and Coordinate (comparative, intersectional). Each framework is accompanied by a diagram.
- Sparks:** Lists various visual outputs and workflows. It includes sections for Human-AI Collaboration for Various Visual Outputs, Level of Human-AI Collaboration, Workflow of Tasks, and Automated Visual Content Synthesis.
- Slides Preview:** Shows a preview of the generated slides, including a title slide, a content slide, and a conclusion slide. Each slide is labeled with a letter (A, B, C, D, E, F) corresponding to the panels in the main interface.

Figure 4: PAPERBRIDGE can be navigated through left, middle, and right panels. It supports (1) Paper Management in Panels A & B on the left, (2) Narrative Exploration (main feature) in Panels C, D, and E in the center, and (3) Slide Draft Preview in Panel F on the right.

Perspectives Refinement and Rationale Exploration in Panel E (DC2 & DC3). Panel E is a workspace that provides two modes that can be switched with the button alongside the contribution statement text box.

- One is **Organization Mode** (Panel E, Figure 4), where users can check and refine the components in each spark (*i.e.*, narrative perspective) suggested by PAPERBRIDGE through a top-down analysis. Bottom-up exploration also happens in Panel E, where users assign papers to clusters manually, then have PAPERBRIDGE synthesize contribution statements and cluster themes. In either way, users can refine the components, such as editing statements and clusters' themes, and dragging and dropping papers into different clusters. They can also confirm elements they like and request the regeneration of anything unsatisfactory. To facilitate the exploration, papers are represented as numbered circles with a hover-activated title display, and contextual explanations on the elements are accessible through hover interactions.
- Another is **Rationale Mode** (see Figure 7 where this mode is activated), where users can explore various rationale strategies (identified in subsection 4.2.2 and displayed in Panel C) by dragging and dropping them into Panel E, helping them articulate the significance of their chosen perspective.

Throughout the workflow, users can switch freely among frameworks (Panel C), perspectives (Panel D), workspace (Panel E) to explore diverse narratives that suit their objectives. Whenever satisfied with the content in work space, they can confirm their workspace content, and PAPERBRIDGE will transform the content into editable slides (DC4) and provide a preview in Panel F (Figure 4). Users can export them for further refinement.

5.2 Usage Scenarios Walkthrough

In this section, we present a usage scenario that walks through PAPERBRIDGE, illustrating how it facilitates co-exploration on research narrative based on user intent and interactions. Below, we use *italics* to indicate Flora's thoughts, and teletype to indicate the button on the interface. We use "Initial Capitalization" for the content suggested by PAPERBRIDGE to the user.

Scenario Background. Imagine Flora², a postdoctoral fellow in HCI who focuses on visual storytelling, with an emphasis on motion effects. Her recent research portfolio has expanded significantly with new publications related to human-AI collaboration (*e.g.*, the underlying collaboration paradigms to achieve the outputs). She wants to reorganize these works under this broader topic but struggles to pinpoint suitable perspectives. Flora turns to PAPERBRIDGE for inspiration.

Paper Selection (Panels A & B). As shown in Figure 4-A, after entering her Google Scholar URL into PAPERBRIDGE, Flora types her current intent: "I want to figure out how humans and AI can collaborate to ease visual materials authoring, such as visualization, data videos, slides, etc." She clicks the `let's go` button. In response, PAPERBRIDGE automatically extracts her publications from Google

Scholar and categorizes them into three clusters (Figure 4-B): "Human-AI Collaboration"; "Visual Storytelling"; "Systems for Data Analysis". "These categories look pretty close to my mental map of my work," Flora notes as she navigates the tags in Panel B. She checks the relevant papers potentially relevant to Human-AI collaboration, and refines "Human-AI Collaboration" to "Human-AI Collaboration for Visual Storytelling" to better align with her intended focus. Satisfied with these selected papers, she clicks `Confirm` in Panel B.

Exploring Potential Narrative Perspectives (Panel C, D, & E). Flora then moved to the middle area. She starts in Panel C (Figure 4-C), which displays four different narrative frameworks: parallel, linear, circular, and coordinate. "My prior approach seems like a linear approach," Flora reflects. She clicks the `Linear` first to see how PAPERBRIDGE maps her publications in a linear arc. PAPERBRIDGE responds by generating "sparks" in Panel D (Figure 4-D). The first spark closely matches her previous storyline, though phrased differently: "Human-AI Collaboration for Various Visual Outputs," with thematic clusters labeled "2D Visualization," "Video Effects," and "Narrative Storytelling". She sees that hovering over the numbered circles in Panel E reveals each paper's title (Figure 4-E). "Good. These align with the developmental path of my work—starting with 2D data viz and moving into animations," Flora thinks. Encouraged by this familiarity, she decides to explore more potential perspectives on human-AI collaboration with the help of PAPERBRIDGE.

Flora clicks the `Coordinate` framework next, and PAPERBRIDGE analyzes her publications again. The first candidate perspective in Panel D is labeled "Automation Degree & Authoring Content". As shown in Figure 5-i, this perspective is mapped onto a coordinate plane with two dimensions: "AI-driven Automation" vs. "Human-in-the-loop Interaction" on one axis, and "Visualization Authoring" vs. "Animated Video Authoring" on the other. "This is interesting," Flora thinks, "It distinguishes how AI's role varies with different output formats." She hovers the perspective text box to reveal PAPERBRIDGE's explanation (Figure 5-i): "This coordinate system categorizes research in terms of the level of AI involvement and the complexity of the content created as two directions, which can be used to discuss the human's role vs. AI's within visual storytelling domain." "That's spot on!" Flora thinks. She checks the assigned papers by hovering over their numbered circles and relocates two papers by dragging and dropping them into another cluster (Figure 5-ii). Besides, she revises "AI-driven Automation" to "AI-powered Automation" (Figure 5-iii). With that small tweak done, she clicks `Confirm`, and PAPERBRIDGE generates a corresponding slide reflecting this coordinate-based perspective (first slide shown in Panel F, Figure 4).

"Maybe there are other angles worth trying," Flora thinks. She decides to explore a bottom-up approach under the `Parallel` framework. This time, she first clicks `reset` button in Panel E, and manually groups her papers based on her understanding of the common and distinct traits in her research (Figure 6-i). She then requests PAPERBRIDGE to generate suitable contribution statements and cluster themes by clicking the `update` button in Panel E. Among the various attempts, one especially catches her eye (Figure 6-ii): "Human-AI Collaboration for Various Interaction Paradigms," yielding three thematic clusters: "HAI for Natural Language Interaction", "HAI for Template-based Interaction", and "HAI for Direct Manipulation Interaction". "I never thought of it this way, but it nicely highlights

²To comply with anonymity requirements during review, we constructed this scenario using a synthetic profile. We combined publications from multiple researchers, re-framed paper titles, altered venues and publication years, and removed identifying system names. The context here and the content shown in Figure 4 do not reflect any single researcher's actual publication history.

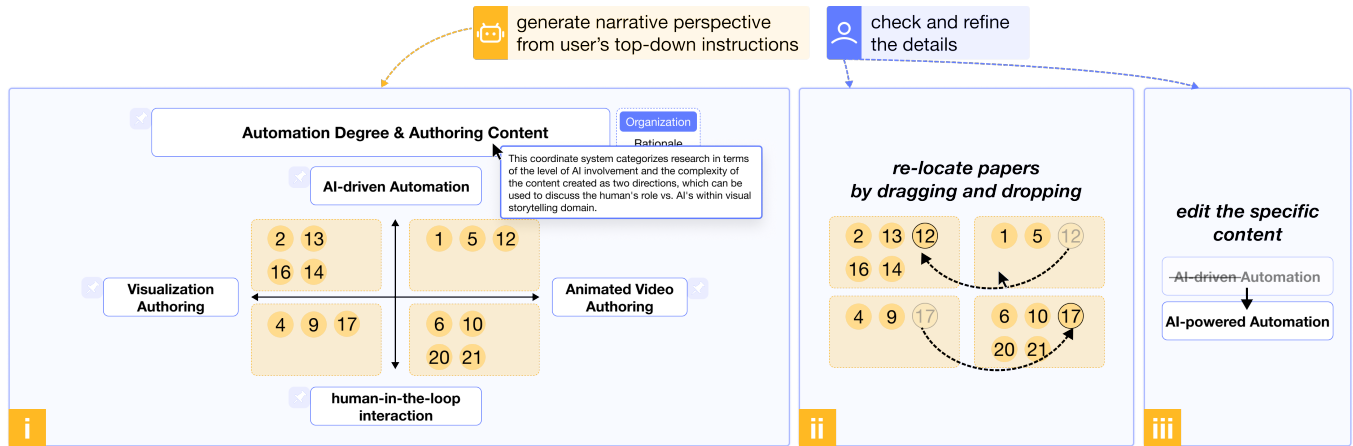


Figure 5: PAPERBRIDGE supports user in checking, adjusting, and revising the top-down exploration results. (i) One narrative perspective suggested by PAPERBRIDGE, including contribution statement, thematic themes, and assigned papers. The user can hover on the components to see the explanations from PAPERBRIDGE. (ii) User can re-locate papers to other clusters by dragging and dropping interaction. (iii) The user can edit the specific content to satisfy their intent.

how I've been supporting diverse interaction methods," she realizes. She re-checks the paper assignments in each cluster, adjusts a few accordingly, and decides "Template-based Interaction" feels a bit vague. She locks the other two themes as final and clicks update in Panel E again. PAPERBRIDGE regenerates that theme as "HAI for Example-based Interaction" (Figure 6-iii), which Flora appreciates. She confirms it, prompting PAPERBRIDGE to generate another slide (second slide shown in Panel F, Figure 4).

Crafting the Rationale (Panel E). Though Flora is satisfied with the interaction paradigm in organizing her research, she is not sure how to justify its importance in a talk. Therefore, Flora clicks the Rationale button in Panel E, enabling Rationale Mode. As shown in Figure 7, simultaneously, PAPERBRIDGE displays various rationale strategies in Panel C. "Let's see what with big name's quote," she thinks. She drags the big name's quote strategy to the text boxes in Panel E. PAPERBRIDGE automatically generates the narrations based on this strategy and shows the content in the text box (see Figure 7). Clicking Confirm, she sees the current content in the workspace is organized into a new rationale slide.

Exporting the Slides (Panel F). Finally, Flora compares the different frameworks and saves several promising perspectives. "I've got a nice variety of narratives," she thinks. When she feels she has gathered enough directions, she clicks the Export button in Panel F (Figure 4-F). PAPERBRIDGE generates a slide file with scripts (in .pptx format). "Perfect. I can show these different ways of framing my work to my advisor," Flora thinks, "This might spark new ideas and feedback."

In this way, Flora capitalizes on both top-down and bottom-up exploration, stepping freely between different frameworks, customizing thematic clusters, and refining her rationale. PAPERBRIDGE's iterative workflow helps her craft diverse narrative perspectives that reflect her research contributions in new and insightful ways.

5.3 Backend Engine: Bi-directional Analysis

The backend of PAPERBRIDGE is designed to support users in constructing research narratives from their collection of publications. It consists of four modules: (1) paper collection and sifting, (2) bi-directional analysis engine as the core, (3) evaluation and ranking, and (4) slide generation. We describe each module in detail below.

5.3.1 Paper Collection and Sifting. Given a Google Scholar URL provided by the user, we use the scholarly package³ to retrieve the author's publication metadata, including each paper's title, abstract, citation string, and unique ID. To support efficient navigation and selection, the system leverages LLMs (OpenAI's o3-mini) to categorize the retrieved papers by topic based on user intent. The user-selected filtered papers, together with the specified research focus and their intent, serve as the contextual input for the subsequent bi-directional analysis engine.

5.3.2 Bi-directional Analysis. As shown in Figure 8, the key to generate each narrative perspective is to identify its core components (see Figure 2): a contribution statement, thematic clusters, and the individual papers within each cluster. Our solution leverages LLMs (OpenAI's o3-mini) for their strong natural language understanding and reasoning capabilities to generate diverse narrative perspectives and their constituent components.

Narrative Schema: A Unified Schema for Data Flow and Mixed-Initiative Interaction. To support both top-down and bottom-up exploration, we designed a structured JSON schema (see Figure 8-iii) to represent the key components of a candidate narrative perspective. Specifically, the schema includes:

- a particular clustering of papers, `paper_assign`
- the extracted themes of these clusters, `cluster_theme`
- a synthesized contribution statement that captures the unique framing of this perspective, `contribution_statement`

³See <https://pypi.org/project/scholarly/> for documentation.

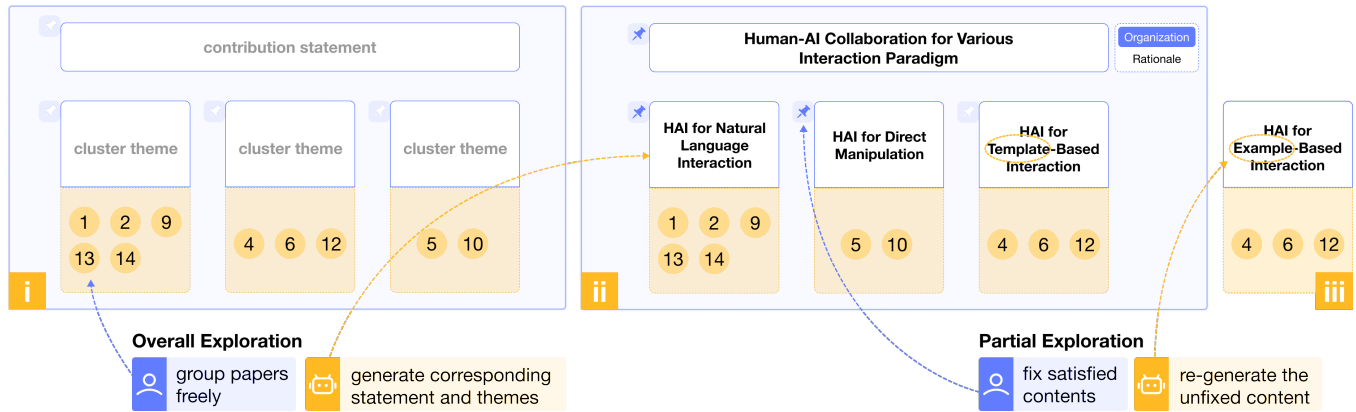


Figure 6: PAPERBRIDGE supports bottom-up explorations in both overall or partial manner. (i) User can group their papers freely, and request PAPERBRIDGE to generate overall suggestions. (ii) PAPERBRIDGE suggests the corresponding narrative perspectives, including contribution statement and clusters' themes. (iii) User can fix the satisfactory content and request PAPERBRIDGE to re-generate the unfixed content.

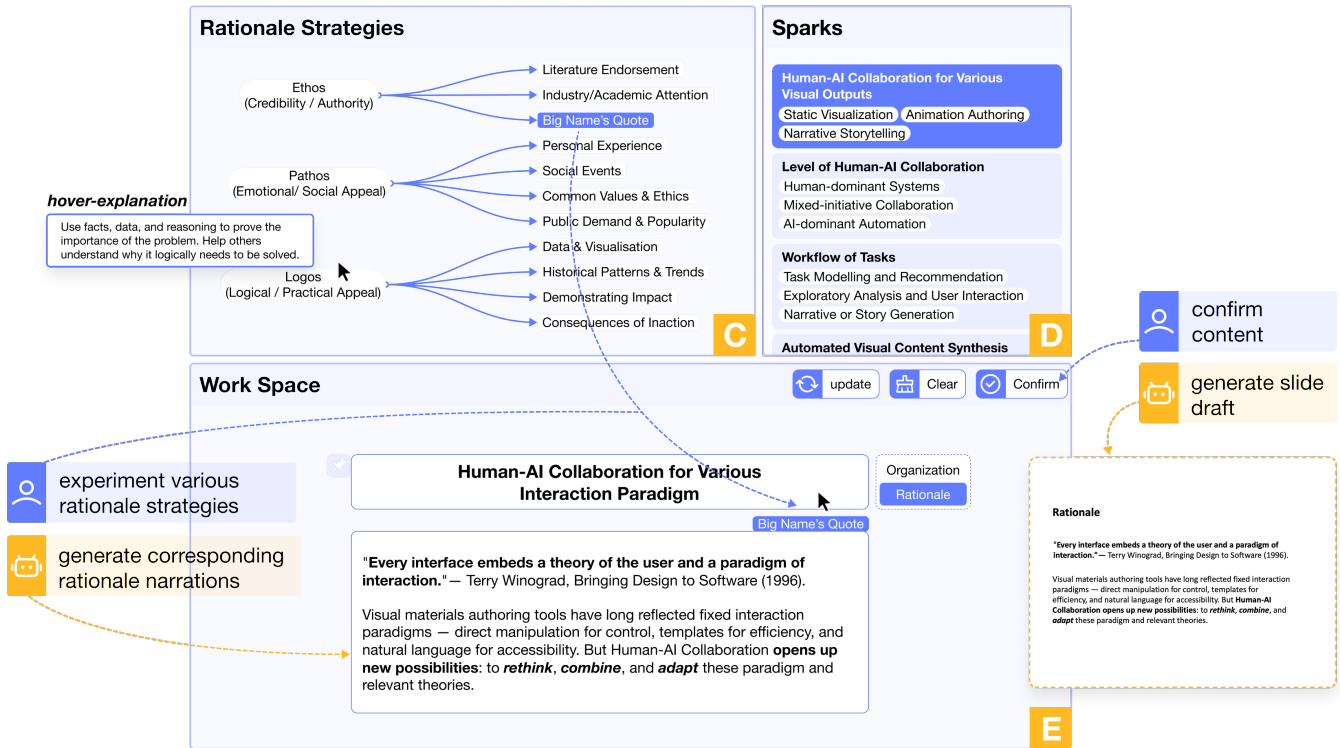


Figure 7: PAPERBRIDGE supports users in exploring various rationale strategies to justify the significance of the specific contribution statement by generating corresponding narration drafts for users.

As the central data layer of our system, this schema can be populated via LLM-generated outputs (top-down) or directly edited by users (bottom-up). It thus serves as a consistent and manipulable representation of narrative components, enabling seamless transitions between automated generation and human-driven refinement.

This schema also underpins our prompt chain design, which we detail below in both top-down and bottom-up workflows.

Top-Down Reasoning: Framework-guided Clustering and Theme Synthesis. In the top-down workflow, the prompt chain includes two modules of context and instructions before the JSON output restrictions (see [Appendix A](#) for the full prompt). The first

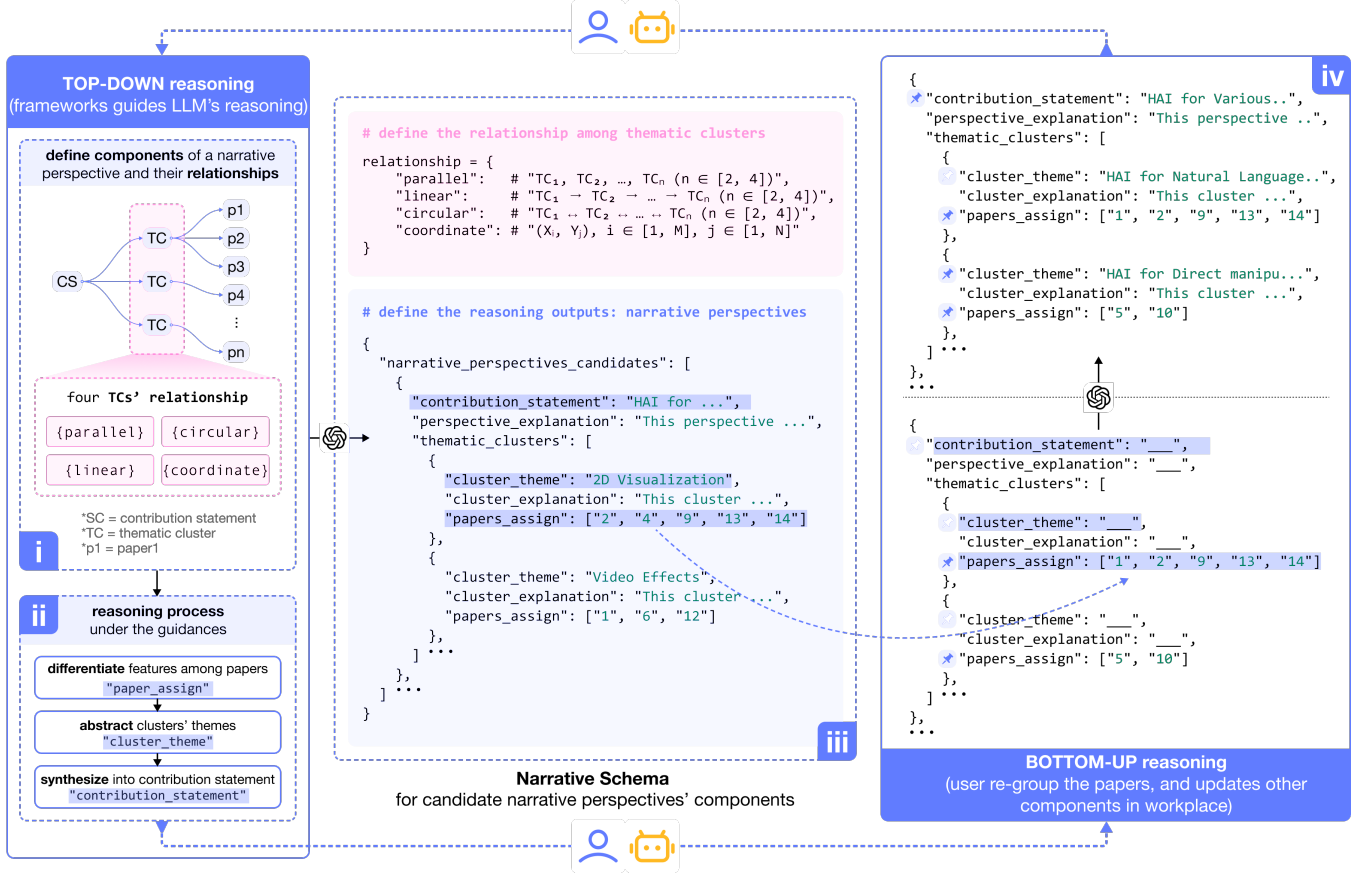


Figure 8: Backend implementation for PAPERBRIDGE, a bi-directional analysis engine that supports top-down and bottom-up reasoning for organizing and synthesizing publications (abstracts and titles) into narrative components. *Top-down*, we define structured narrative components, especially inter-cluster relationships, to guide LLMs in grouping thematically coherent paper clusters. *Bottom-up*, users can re-group papers themselves, refine clusters, and reshape contribution perspectives. The *Narrative Schema* serves as a bridge between both reasoning processes by storing and updating narrative components.

module is a narrative structure-specific instruction module (see Figure 8-i). In which, we translated each narrative framework into natural language, defining a) the concepts of research statement, paper clusters, and themes; b) the hierarchy from statement, to clusters and individual papers; and c) semantic relations among clusters (e.g., sequential in linear, orthogonal in coordinate). The second module is a reasoning chain guiding the LLMs through three steps (see Figure 8-ii): identifying distinctive features among papers, abstracting cluster themes, synthesizing potential contribution statements. The output of this process is the structured JSON schema storing a set of candidate narrative perspectives and their components (see Figure 8-iii).

Bottom-Up Reasoning: User-driven Re-Clustering and Narrative Refinement. In the bottom-up workflow, users interact with the Narrative Schema through the frontend interface to re-group papers, refine cluster themes, or revise contribution statements. Specifically, users can selectively lock components to preserve their edits, such as `paper_assign` (by manually re-grouping papers), `cluster_theme` (by locking thematic cluster boxes), or

`contribution_statement` (by locking the statement box). Once an update is requested, the LLM updates only the unlocked components (see Appendix B for the full prompt).

For example, in Figure 8-iv, a user groups the papers first, locks the `paper_assign`, and requests updates from PAPERBRIDGE. Upon request, PAPERBRIDGE updates the remaining components (i.e., `contribution_statement` and `cluster_theme` in this case) based on the locked parts. The analysis yields corresponding `cluster_theme` and `contribution_statement`.

5.3.3 Evaluation and Ranking. Before presenting the candidate perspectives to users, we evaluate each LLM-generated narrative perspective using SentenceTransformer-all-MiniLM-L6-v2⁴. The system computes a *Final Score* by aggregating five equally weighted metrics that assess both semantic coherence and structural alignment. The first four metrics focus on semantic relationships: Statement-Cluster Alignment (SCA) evaluates semantic alignment between clusters and their parent statements; Intra-Cluster Cohesion (ICC)

⁴See <https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2>

quantifies internal cluster consistency; Adjusted Rand Index (ARI) measures cluster coherence against unsupervised clustering results; and Paper-Cluster Similarity (PCS) verifies paper alignment with assigned cluster themes.

The fifth metric, Structural Consistency (SC), ensures that generated structures align with their intended narrative frameworks through pattern-specific evaluations. For Parallel patterns, we compute Cluster Separability to ensure distinct thematic boundaries between clusters. In Linear patterns, we assess Cluster Ordered-ness to verify logical progression between consecutive clusters. For Circular patterns, we measure Cluster Interdependence to examine mutual relationships among all clusters. In Coordinate patterns, we evaluate Axis Separability to confirm the proper distribution of clusters along different axes.

The final score aggregates these metrics as: $\text{FinalScore} = 0.2 \cdot \text{SCA} + 0.2 \cdot \text{SC} + 0.2 \cdot \text{ARI} + 0.2 \cdot \text{PCS} + 0.2 \cdot \text{ICC}$. Each metric is normalized to $[0, 1]$, and the four highest-scoring perspectives (the number is based on the findings of our pilot study, see [subsection 6.1](#)) are presented to users through the interface.

5.3.4 Slide Generation. The final backend module generates presentation slides from user-confirmed structures using the Python PPTX package⁵. We developed four templates aligned with four narrative frameworks⁶, enabling export to PowerPoint for further refinement. In addition, we utilized Recraft⁷ to automatically generate images corresponding to each narrative perspective, aligning the visuals with the narration and enhancing overall engagement.

The system's frontend is implemented in Vue⁸, with a Flask⁹ backend. Asynchronous data exchange facilitates seamless communication, while D3.js¹⁰ powers interactive tree chart visualization corresponding to the four frameworks with real-time updates.

6 EVALUATION

To evaluate the PAPERBRIDGE, we conducted a user study with 12 HCI researchers. Specifically, our evaluation aims to investigate: 1) the usability and overall experience of PAPERBRIDGE, *i.e.*, how useful and easy it is to use PAPERBRIDGE, 2) if and how PAPERBRIDGE supported researchers' exploration of diverse research narratives.

6.1 Methods

Before our main user study, we performed a pilot study with 4 participants to refine the study procedures. A key insight from this pilot study concerned the number of sparks (narrative perspectives) to provide under each framework. Participants indicated that 3-5 sparks struck an appropriate balance in terms of cognitive workload. Based on this feedback, we standardized the number of sparks to 4 per framework in our main user study to ensure a manageable yet sufficient workload for participants.

6.1.1 Participants. We recruited 12 researchers (P1-P12) through an open call on social media (4 female, 8 male, other gender options were provided). They span various career stages from senior PhD

students to postdoctoral researchers and a research assistant professor. The participants had substantial publication records, ranging from 8 to 26 publications. Their research expertise covered a diverse range of HCI subfields, including virtual and augmented reality, visual analytics, accessibility, computer graphics, affective computing, cultural heritage, information visualization, human-AI collaboration, and online communities. This diversity in research backgrounds ensured that our evaluation captured perspectives from different areas within the HCI community.

6.1.2 Procedures and Data Collection. Our user study comprised four stages: (1) a brief introduction; (2) a pre-interview before interaction with PAPERBRIDGE; (3) free exploration with PAPERBRIDGE; and (4) a post-interview including questionnaires. The whole procedure lasts around 1 to 1.5 hours. Details are as follows:

Brief Introduction. The study began with obtaining participant consent and introducing the project. Participants were presented with their primary task: organizing publications for a hypothetical job talk. Specifically, the task required them to re-organize their publications from perspectives different from their previous approach. Using PAPERBRIDGE to assist in this process, they needed to draft at least two pages of slides. Time was allocated for participants to consider their potential audience and their intentions for highlighting their research.

Pre-interview and Brainstorming. Participants then engaged in a pre-interview where they described their preliminary thoughts on organizing and presenting their body of work, along with their justifications. Pen and paper were provided for participants to illustrate their ideas. The facilitators asked follow-up questions when clarification was needed. This phase was crucial for establishing a baseline understanding of participants' organizational thinking before system interaction. The pre-interview was audio recorded.

Free Exploration with PAPERBRIDGE. After the pre-interview, we provided a 2-minute tutorial on PAPERBRIDGE and allowed participants 5 minutes for initial exploration. Participants then interacted with PAPERBRIDGE freely without time constraints. During the interaction, the laptop's screen was recorded.

Post-interview and Questionnaires. Upon completing their interaction with PAPERBRIDGE, participants completed questionnaires about their experience and then engaged in a post-interview. Specifically, the data collected included:

- **Quantitative:**

- *Overall usability:* System Usability Scale (SUS) [6] using a 5-point Likert scale.
- *Workload:* NASA Task Load Index (NASA-TLX) [18], using a 7-point Likert scale.
- *User experience questionnaire:* A 10-question survey using 5-point Likert scales, evaluating participants' perceptions of (1) the exploration process (inspired by [67, 68]), (2) inspiration (inspired by [12, 61]), and (3) overall satisfaction. The specific questions can be found in [Figure 9](#).
- *Ratings on generated narrative perspectives:* Participants were asked to rate each spark (16 in total) on a 5-point scale based on its reasonableness and helpfulness. Additionally, they were asked to indicate which sparks they had previously considered, allowing us to distinguish between novel ideas generated by PAPERBRIDGE and those already conceived by participants.

⁵See <https://pypi.org/project/python-pptx/> for documentation

⁶See supplementary for four slide templates

⁷See <https://www.recraft.ai/docs> for documentation

⁸See <https://vuejs.org/>

⁹See <https://flask.palletsprojects.com/en/stable/>

¹⁰See <https://d3js.org/> for documentation



Figure 9: Assessment of participants' perception of PAPERBRIDGE in terms of exploration (Q1-Q4), inspiration (Q5-Q6), and satisfaction (Q7-Q10).

• Qualitative:

- We conducted semi-structured interviews to collect participants' thoughts and feedback during the exploration process. Sample questions included: "Please describe your overall experiences with PAPERBRIDGE," "What are your opinions on the frameworks and sparks provided by the system?," "Did you find new perspectives to frame your research, and how did our system support this?," and "What suggestions do you have for improving the system?"

6.2 Findings 1: Overall Usability and User Experiences

6.2.1 Usability and Workload. The SUS and NASA-TLX measurements indicate that PAPERBRIDGE is easy to use and imposes a low cognitive burden, while supporting high levels of task success during exploratory activities. Specifically, the System Usability Scale yielded an average score of 84.38 (SD = 19.54), which falls within the range of excellent usability [6]. In terms of workload (measured on a 7-point scale), participants reported low mental demand (M = 3.08, SD = 2.07), very low physical demand (M = 1.42, SD = 0.67), and low time pressure (M = 2.33, SD = 1.67) while using PAPERBRIDGE. Notably, participants reported a high sense of success in completing the task (M = 5.50, SD = 1.31), suggesting that the system effectively supported their goals. Reported effort levels were moderate (M = 3.42, SD = 1.51), and frustration remained low (M = 1.75, SD = 0.97).

6.2.2 Overall User Experience. As shown in Figure 9, results from the user experience questionnaire (N = 12, 5-point Likert scale) indicate that participants were generally satisfied with PAPERBRIDGE (M = 4.39, SD = 0.59). Specifically, participants gave positive evaluations across three key dimensions. The *exploration* process dimension (4 questions) received high ratings (M = 4.48, SD = 0.52), indicating that participants found the system effective for organizing research and brainstorming alternative perspectives. The *inspiration* dimension (2 questions) received the highest ratings (M = 4.54, SD = 0.69), suggesting that PAPERBRIDGE helped users discover new ideas and view their work from different angles. The *satisfaction* dimension

(4 questions) was also rated positively (M = 4.23, SD = 0.73), with slide output satisfaction showing the most room for improvement.

6.3 Findings 2: Co-Exploration Processes with PAPERBRIDGE

The analysis of post-interviews and participants' ratings on the generated narrative perspectives, complemented by in-situ observations during the study, revealed how PAPERBRIDGE: (1) enhanced the efficiency of narrative ideation; (2) supported the co-exploration through both top-down and bottom-up approaches; and (3) inspired reflections on participants' research corpus beyond the immediate task of narrative construction. These findings demonstrated that PAPERBRIDGE was helpful in exploring alternative research narratives, and also surfaced insights about participants' further expectations.

6.3.1 Facilitating Efficient Narrative Ideation. All participants reported that PAPERBRIDGE improved the efficiency of narrative exploration by offering diverse options or perspectives for them. This efficiency emerged from multiple system affordances:

First, most participants noted that both the frameworks and narrative perspectives provided by PAPERBRIDGE **offered a wide range of starting points** for exploration. Frameworks gave access to high-level, abstract structures, while perspectives under each framework served as more concrete, interpretable sparks. As P7 described: "It's really helpful to have these frameworks available. It makes it much easier to try different structures and think about how to organize my publications." When it came to narrative perspectives, participants highlighted how they helped reduce the barrier to starting narrative ideation. P8 remarked: "I probably couldn't have come up with so many different approaches on my own. Having these perspectives presented to me made the task of creating my research narrative feel less daunting."

In addition to the rich and diverse suggestions provided, participants appreciated the PAPERBRIDGE's **clear workflow that streamlined their exploration**. We found that the top-down exploration helped participants structure their thoughts more systematically. P9 shared: "I like its ability to guide me through a step-by-step exploration of different narrative aspects; when I think on my own, my

thoughts tend to be scattered.” At the same time, participants mentioned PAPERBRIDGE facilitated their bottom-up exploration with real-time updates. P9 emphasized: “Grouping my papers and testing different ideas with PAPERBRIDGE is more efficient than sketching on paper. I can quickly get feedback on how these papers might work together when organized in different ways.”

Finally, the PAPERBRIDGE’s **ability to output slide drafts** was also appreciated by participants, as these drafts helped them transition more directly into discussion and collaboration. As P6 noted: “With these slides, I can more quickly take the materials for discussion with others; previously, I had to make separate slides from the beginning after organizing my thoughts.” Beyond immediate use, some participants imagined integrating these outputs into longer-term academic assets. For instance, P12 suggested: “With some optimization, I think I will put the visuals on my personal website.”

6.3.2 Co-Exploration Process. Beyond improving efficiency, our findings revealed that PAPERBRIDGE supported an exploratory and reflective process for crafting research narratives. This process began with trust building, human-AI co-exploration of narrative perspectives, and participants’ reflection on their research beyond the immediate task of narrative construction.

Initial Trust Establishment. We noticed the participants commonly began by quickly scanning among the frameworks and sparks. A sense of initial trust was often formed when they recognized familiar or sensible ideas within the system’s suggestions. As P2 explained: “When I saw ideas I had thought of before, I felt the system worked well; it made me feel that the sparks it provided were logical and reasonable.”

Co-Exploration of Narrative Frameworks. Participants often initiated deeper exploration by examining high-level narrative frameworks. The presence of multiple frameworks sparked curiosity and motivated participants to explore the perspectives under each of them. As P7 reflected: “These frameworks seemed reasonable, making me want to explore each of them.”

Co-Exploration of Diverse Narrative Perspectives. When exploring the sparks (i.e., narrative perspectives) within a specific framework, the **discovery of novel perspectives** often facilitated deeper engagement with the exploration process. For example, P1 noted: “I designed and observed different user behavior in VR. However, I never thought to categorize my work through the lens of ‘perception.’ It surprised me and made me want to explore more.”

Besides, we found that **keywords** associated with contribution statements and thematic clusters played a critical role. The keywords generated by PAPERBRIDGE offered diverse entry points and remained open to interpretation, allowing participants to project their own understanding. As P6 noted: “Seeing these keywords, I will interpret them first before I check the details in the workspace.” We also observed that participants tended to prefer abstract or conceptual keywords over specific ones, as the former provided more room for interpretation. As P4 commented: “There were two cluster names—one was ‘tutorial-guided interaction’ and the other was ‘hand gesture interaction.’ The former left room for imagination, while the latter felt too specific and less suitable.”

Additionally, the interpretive space created by these keywords often inspired participants to **shift into bottom-up exploration** to test and refine their own ideas. For example, P12 mentioned:

“When I saw the keywords ‘decision-making process,’ I got an aha moment. But I didn’t like the way it grouped my papers. So, I tried to fix the contribution statement, assign the papers myself, and request the system to return new cluster themes.”

Reflections Emerged During the Exploration. We noticed that the exploration process was a process to help researchers rethink their research through different perspectives. We also noticed that, in addition to the diverse candidates that served as the reflection trigger, PAPERBRIDGE’s visual and conceptual organization prompted participants to think more broadly about their research positioning and future directions. Specifically, the visual layout of the workspace enabled participants to identify latent patterns and gaps across their body of work. P4 shared: “When it showed three clusters describing how my work innovates tangible interface input methods—gesture input, object input, and spatial context input—I noticed I had fewer papers in the last cluster, suggesting a potential direction for future research.”

6.4 Findings 3: Challenges and Tensions Observed

Since participants are most familiar with the content and context of their own publications, we consider their feedback particularly valuable. We summarize their expectations as part of our findings to surface the tensions and challenges involved in supporting the reframing of one’s own intellectual work.

6.4.1 Complexities in How Users Perceive and Rate Sparks. We collected a total of 192 spark ratings (4 sparks × 4 narrative structures × 12 participants), yielding an overall mean score of 3.18 (SD = 1.29). To further examine how familiarity shaped evaluations, we compared ratings of sparks that captured ideas participants had already considered (“thought-of,” N = 66) versus those that were entirely new (“not-thought-of,” N = 126). A paired-samples *t*-test showed that “thought-of” sparks were rated slightly higher (M = 3.41, SD = 1.29) than “not-thought-of” sparks (M = 3.06, SD = 1.27), though the effect was only marginal, $t(11) = 2.14$, $p = .056$.

We noticed an inconsistency between the average spark scores and the otherwise positive quantitative and qualitative data. Despite a moderate mean rating, the SUS scores, user experience measures, and participants’ interview responses indicated that PAPERBRIDGE was effective in helping users explore diverse research narratives. This discrepancy calls for a more nuanced interpretation of the spark ratings.

Specifically, we observed that spark scores were sensitive to the characteristics of users’ inputs, particularly the number of papers and the degree of thematic coherence among them. For example, P11 selected four papers, two of which were topically disjointed. P11 gave lower ratings to most sparks, with only one spark receiving a score of 4. Reflecting on this, the participant acknowledged, “I know my publications were not that coherent, which made me headache. But at least, your system yielded one helpful perspective for me called ‘technological mediation in the interaction.’” This was the one spark that received a rating of 4. Drawing from this case, we interpret that the spark scores do not fully capture the extent to which participants were inspired by PAPERBRIDGE, as a smaller number of papers may limit the system’s ability to form meaningful clusters from the input, while greater topical diversity can reduce the semantic coherence

required to generate compelling sparks. In other words, both factors may constrain the quality of the generated sparks.

6.4.2 Expectations for More Flexible Framework Organization. The exploration of different narrative frameworks made participants express interest in more open-ended and customizable structural frameworks. They wished for the ability to go beyond predefined templates and experiment with representations that better reflect the nuances of their work. For instance, P3 remarked: “*I wish I could arrange and combine these frameworks arbitrarily. For instance, I want to add two additional nodes to one edge of a circular structure and see what happens.*” P6 emphasized the value of aligning frameworks with the structure of specific theoretical models: “*Different theories have different levels, and these levels and nodes could potentially resonate with one’s papers.*” Meanwhile, both P1 and P5 pointed out the importance of relative positioning in coordinate-based structures. As P5 explained: “*I know that this paper (pointing at the screen), the degree of AI automation is higher, so it should be placed closer to the extreme of this axis.*” These insights indicate a strong desire for more granular, theory-aligned, and manipulable narrative structures.

7 DISCUSSION

In this paper, we targeted HCI researchers’ own challenges in exploring alternative research narratives, and presented PAPERBRIDGE to support this process. In this section, we reflect on the key design decisions and associated trade-offs, and discuss implications and directions for future research.

7.1 Supporting Academic Storytelling and Communication

By targeting the challenge of helping HCI researchers construct alternative narratives around their own body of work, our study contributes to a growing line of research on how interactive systems can support academic practice [22, 26, 65]. While many existing tools focus on literature discovery [28, 63, 71], writing mechanics [38, 50], or citation management [63], our work centers on a distinct yet critical scenario: **academic storytelling and communication**. We approach this direction by examining and supporting research narratives constructions based on one’s own publications. In the following, we reflect on our work and offer insights that may inform future research in this emerging space.

7.1.1 Narrative Frameworks as Design Patterns and Conceptual Scaffolds for Academic Storytelling. Our investigation started from identifying narrative frameworks (*i.e.*, *parallel*, *linear*, *coordinate*, and *circular*). While the four structures are not exhaustive, we argue that they may serve as design patterns [4, 60] as well as conceptual scaffolds [7, 48] in guiding future research and system development.

As design patterns, these frameworks offer reusable structural templates for organizing publications into narrative perspectives. They provide a foundation for interface designs that support narrative construction, and can be extended to accommodate greater flexibility. For example, future systems might allow users to adjust the number of clusters, combine multiple structural logics, or interactively manipulate narrative layouts to better reflect individual goals and preferences. However, it is worth noting that increased

flexibility may also introduce new challenges. For instance, over-fragmentation could compromise narrative coherence, highlighting the need for additional design considerations to support manageable and meaningful structure selection. **As conceptual scaffolds**, these frameworks offer cognitive support for complex narrative construction. They help users interpret and decompose ideas, enabling them to make sense of the bodies of their own work. Beyond academic storytelling, such scaffolds may generalize to other interpretive tasks that require organizing multiple components into a coherent whole, such as structuring course modules planning [16].

7.1.2 Academic Storytelling as an Exploratory and Reflective Process. The evaluation of PAPERBRIDGE showed that researchers appreciated the support in academic storytelling, and revealed that engaging with PAPERBRIDGE not only improved efficiency but also fostered a co-exploration process that encouraged reinterpretation and reflection on their research. These observations suggest that academic storytelling can be understood as **an exploratory and reflective activity**, rather than merely a task of communication. In light of this, we call for greater attention to the design of systems that support researchers in exploring and reflecting on their own intellectual trajectories. In our study, we observed that using keywords as outputs created space for reflection. This **lightweight and suggestive design** encourages users to reconsider and reframe their own work, rather than passively accept system-generated content. Such an approach offers a promising direction for future systems that aim to balance guidance and interpretive flexibility. Although our study focused on HCI researchers, the need to make sense of one’s academic trajectory exists across disciplines and contexts, such as grant writing, portfolio curation, or job applications. We see this as an emerging space for academic and professional storytelling that is worth further exploration.

7.1.3 Challenges in Supporting Narrative Construction with Familiar Content. Throughout our study, we identified several design challenges specific to the academic storytelling and communication scenario. Unlike tasks involving unfamiliar content [53, 56], participants were deeply familiar with the material, as they were working with their own publications. This familiarity introduced a unique tension: participants wanted the system to inspire them with new ideas, but also expected its suggestions to align with their self-understanding. As a result, they tended to be more critical of the system’s outputs, and their assessments were heavily influenced by subjective expectations. For example, spark ratings were slightly higher for ideas they had already considered, suggesting a preference for outputs that resonated with their existing thinking. To support such scenarios, future systems could consider providing users with more agency in the interaction [21, 55]. This may include exposing parts of the system’s reasoning process or enabling interaction with scaffolded reasoning steps [10], in order to provide transparency that fosters trust and supports more reflective evaluation during research narrative construction.

7.2 Designing Human-AI Collaboration for Mixed-Initiative Interaction

To support the exploratory nature of research narrative construction, PAPERBRIDGE enables human-AI co-exploration through both

top-down and bottom-up approaches. To this end, we followed established practices in guiding LLMs by translating narrative construction practices into structured prompt instructions [42]. To further support mixed-initiative interaction around the generated narrative components, we *designed the LLM outputs as shared, editable representations* that bridge model reasoning and user refinement. In the following, we reflect on the design of these editable representations and discuss their broader generalizability.

7.2.1 Editable Representations for Mixed-Initiative Interaction. In our current work, the shared, editable representations are implemented using a structured JSON schema, which defines the key elements of a narrative perspective, including the contribution statement, thematic clusters, and paper assignments. By making these fields editable and accessible across both the frontend and backend, the system allows users to iteratively adjust, reframe, and reorganize the generated content [39, 52]. In addition to *enabling mixed-initiative interaction*, this schema design also *enhances the extensibility* of PAPERBRIDGE. For example, new narrative frameworks can be easily incorporated by defining structure-specific prompts and mapping their outputs to the same schema fields, without altering the overall system architecture.

Beyond narrative construction, the concept of editable representations offers an alternative solution for other interpretive tasks that involve LLMs reasoning and human refinement. This design *reframes a generative task as an interactive process of interpretation and reconstruction* between human and AI. Based on our findings, decomposing narrative construction into modular components helps users avoid the cognitive burden of starting from scratch and provides concrete entry points for reflection. This design principle may be generalized to other domains, including literature surveys [71], idea brainstorming [36], Socratic-style learning [35], and creativity support tools [2, 54]. While our current interface supports text-based refinement, the underlying concept aligns with recent visions of dynamic abstractions [19, 62], which advocate for interpretable and manipulable intermediate representations. Future systems may build on this approach by designing multi-modal interactions, such as sketch- or diagram-based interfaces.

7.2.2 Trade-offs in Input Scope for Narrative Generation. However, our use of LLMs also revealed challenges, particularly around how the scope of input influences the perceived quality of output. Our current approach, which relies on paper titles and abstracts, introduces a trade-off. On one hand, it reduces the input burden (e.g., users do not need to upload full papers), and abstracts often capture the core ideas of a publication, supporting our goal of surfacing narrative inspirations. On the other hand, this limited input may constrain the specificity and depth of the generated clusters and contribution statements. We observed that when users selected only a small number of papers, the system had fewer materials from which to extract meaningful patterns or construct abstract framings. At the same time, users often had highly specific interpretations of these papers, making them more sensitive to mismatches or inaccuracies in the system's suggestions. Although this tension was amplified by the limited input scope, we believe the underlying challenge would persist even with full-text access. Researchers with more publications naturally provide a richer foundation for the system to identify various themes. Addressing this challenge

may require models or systems with more adaptive reasoning capabilities. Such systems would need to operate effectively across varying input sizes and tailor the level of abstraction and suggestion to the quantity and specificity of the available content.

7.3 Limitations and Future Work

We acknowledge several limitations of our current work and would like to outline potential directions for future research. First, our current study focused on HCI researchers, and both the narrative frameworks and user study findings primarily reflect experiences from within the HCI community. The participants in our formative study were mainly early-career researchers preparing for the job market. Although the insights from the formative study and content analysis were discussed with senior HCI researchers, the findings have yet to be examined more broadly across the HCI community. A natural next step is to extend PAPERBRIDGE to a more diverse group of HCI researchers, as well as to other computer science subfields, in order to explore both shared narrative patterns and domain-specific differences in research narrative construction.

Second, the four narrative frameworks are not intended to be exhaustive. Future work may explore additional frameworks or investigate more flexible, customizable structures to support personalized and diverse forms of narrative construction. Third, although PAPERBRIDGE goes beyond standard chat-based interfaces by embedding domain knowledge and providing a structured workflow, our current study did not include a baseline comparison. A valuable direction for future research is to evaluate how PAPERBRIDGE performs in comparison with general-purpose tools like ChatGPT, particularly when both are equipped with narrative knowledge.

Finally, in terms of input for LLMs, we currently rely on titles and abstracts retrieved from Google Scholar to reduce user effort. Future work could incorporate multi-modal inputs uploaded by users, such as PDFs, slides, or project documents, to support deeper narrative analysis and more diverse output formats. Additionally, beyond the published papers, the ongoing projects are valuable resources to be considered.

8 CONCLUSION

In this paper, we targeted the challenges HCI researchers face in constructing alternative research narratives. Building on empirical findings from a formative study and content analysis, we developed PAPERBRIDGE, a human-AI co-exploration system that supports narrative construction through both top-down and bottom-up approaches. Powered by a bi-directional analysis engine, PAPERBRIDGE enables users to generate, revise, and reinterpret their research narratives from diverse perspectives. Our user study with 12 HCI researchers demonstrated that PAPERBRIDGE not only improved the efficiency of narrative construction but also supported an exploratory and reflective process. Based on these findings, we offer implications for future research on 1) academic storytelling and communication and 2) designing human-AI collaborative systems for mixed-initiative interaction.

ACKNOWLEDGMENTS

This work was partially supported by RGC GRF grant 16210722, and was supported in part by the EdUHK-HKUST Joint Centre for

Artificial Intelligence (JC_AI) research scheme: Grant No. FB454. We sincerely thank all participants for their time and efforts, and the anonymous reviewers for their constructive feedback, which significantly improved the quality of this paper.

We would like to thank Dr. Chengbo Zheng and Liwenhan Xie for their valuable discussions during the revision of this paper. We also thank Junze Li, Dr. Yao Wang, Hanlu Ma, Dr. Linping Yuan, Dingdong Liu, and Zheng Wei for their insightful comments on earlier drafts, and Jiawen Liu and Meiyong Li for their help with proofreading. Finally, the first author wishes to express special thanks to Dr. Yulin Tian, Ziqi Pan, Qiaoyi Chen, and to those who showed patience and support during a personally challenging time. Their encouragement and willingness to simply be there made it possible to carry this work through to completion.

REFERENCES

- [1] Avril Aitken. 2010. Becoming an academic: Professional identity on the road to tenure. *The Journal of Educational Thought (JET)/Revue de la Pensée Éducative* (2010), 55–68.
- [2] Tyler Angert, Miroslav Suzara, Jenny Han, Christopher Pondoc, and Hariharan Subramonyam. 2023. Spellburst: A Node-based Interface for Exploratory Creative Coding with Natural Language Prompts. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–22.
- [3] Aristotle Aristotle et al. 1954. *Rhetoric*. Vol. 11. Modern Library New York.
- [4] Benjamin Bach, Euan Freeman, Alfie Abdul-Rahman, Catatay Turkay, Saiful Khan, Yulei Fan, and Min Chen. 2023. Dashboard Design Patterns. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 342–352.
- [5] Maricela Bravo, José A Reyes-Ortiz, and Isabel Cruz. 2020. Researcher profile ontology for academic environment. In *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1*. Springer, 799–817.
- [6] John Brooke. 2013. SUS: a retrospective. *Journal of usability studies* 8, 2 (2013).
- [7] Arthur Cámara, Nirmal Roy, David Maxwell, and Claudia Hauff. 2021. Searching to Learn with Instructional Scaffolding. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval (Canberra ACT, Australia) (CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 209–218.
- [8] Tara Capel and Margot Brereton. 2023. What is Human-Centered about Human-Centered AI? A Map of the Research Landscape. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI'23*. ACM, 1–23.
- [9] 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. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [10] Wei-Hao Chen, Weixi Tong, Ph.D. Case, Amanda, and Tianyi Zhang. 2025. Dango: A Mixed-Initiative Data Wrangling System using Large Language Model. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 389, 28 pages.
- [11] Kiroong Choe, Eunhye Kim, Sangwon Park, and Jinwook Seo. 2024. Fields, Bridges, and Foundations: How Researchers Browse Citation Network Visualizations. In *2024 IEEE Visualization and Visual Analytics (VIS)*. IEEE, 146–150.
- [12] Jean-Peic Chou, Alexa Fay Siu, Nedim Lipka, Ryan Rossi, Franck Dérnoncourt, and Maneeesh Agrawala. 2023. Talestream: Supporting story ideation with trope knowledge. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–12.
- [13] William Cronon. 1992. A place for stories: Nature, history, and narrative. *Journal of American history* 78, 4 (1992), 1347–1376.
- [14] Andres De Los Reyes. 2024. Delivering Academic Job Talks as an Early Career Researcher. In *The Early Career Researcher's Toolbox: Insights into Mentors, Peer Review, and Landing a Faculty Job*. Springer, 117–122.
- [15] Marian Dörk, Nathalie Henry Riche, Gonzalo Ramos, and Susan Dumais. 2012. PivotPaths: Strolling through Faceted Information Spaces. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2709–2718.
- [16] Haoxiang Fan, Guanzheng Chen, Xingbo Wang, and Zhenhui Peng. 2024. Lesson-Planner: Assisting Novice Teachers to Prepare Pedagogy-Driven Lesson Plans with Large Language Models. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (Pittsburgh, PA, USA) (UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 146, 20 pages.
- [17] John Flowerdew and Simon Ho Wang. 2015. Identity in academic discourse. *Annual Review of Applied Linguistics* 35 (2015), 81–99.
- [18] 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.
- [19] Devamardeep Hayatpur, Brian Hempel, Richard Lin, and Haijun Xia. 2025. The Shapes of Abstraction in Data Structure Diagrams. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 883, 12 pages.
- [20] Jiangen He, Qing Ping, Wen Lou, and Chaomei Chen. 2019. PaperPoles: Facilitating adaptive visual exploration of scientific publications by citation links. *Journal of the Association for Information Science and Technology* 70, 8 (2019), 843–857.
- [21] Jeffrey Heer. 2019. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences of the United States of America* 116, 6 (2019), 1844–1850.
- [22] Jeffrey Heer, Matthew Conlen, Vishal Devireddy, Tu Nguyen, and Joshua Horowitz. 2023. Living Papers: A Language Toolkit for Augmented Scholarly Communication. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–13.
- [23] Florian Heimerl, Qi Han, Steffen Koch, and Thomas Ertl. 2016. CiteRivers: Visual Analytics of Citation Patterns. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 190–199.
- [24] Judith A Holton. 2007. The coding process and its challenges. *The Sage handbook of grounded theory* 3 (2007), 265–289.
- [25] Ken Hyland. 2011. The presentation of self in scholarly life: Identity and marginalization in academic homepages. *English for Specific Purposes* 30, 4 (2011), 286–297.
- [26] Ken Hyland. 2018. Narrative, identity and academic storytelling. *ILCEA. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie* 31 (2018).
- [27] 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.
- [28] 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.
- [29] Takeshi Kurosawa and Yasufumi Takama. 2011. Predicting researchers' future activities using visualization system for co-authorship networks. In *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, Vol. 1. IEEE, 332–339.
- [30] Shahid Latif and Fabian Beck. 2018. VIS Author Profiles: Interactive descriptions of publication records combining text and visualization. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 152–161.
- [31] Machine Learning. 2011. Apolo: Making Sense of Large Network Data by Combining. (2011).
- [32] Bongshin Lee, Mary Czerwinski, George Robertson, and Benjamin B. Bederson. 2005. Understanding research trends in conferences using paperLens. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, 1969–1972.
- [33] Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. 2015. More Than Telling a Story: Transforming Data into Visually Shared Stories. *IEEE Computer Graphics and Applications* 35, 5 (2015), 84–90.
- [34] 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 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [35] Jiayu Liu, Zhenya Huang, Tong Xiao, Jing Sha, and Jinze Wu. 2024. SocraticLM: Exploring Socratic Personalized Teaching with Large Language Models. In *NeurIPS*. 1–29.
- [36] Yiren Liu, Si Chen, Haocong Cheng, Mengxia Yu, Xiao Ran, Andrew Mo, Yiliu Tang, and Yun Huang. 2024. How AI Processing Delays Foster Creativity: Exploring Research Question Co-Creation with an LLM-based Agent. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–25.
- [37] Justin Matejka, Tovi Grossman, and George Fitzmaurice. 2012. Citeology: visualizing paper genealogy. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*. ACM, 181–190.
- [38] Andy Nguyen, Yvonne Hong, Belle Dang, and Xiaoshan Huang. 2024. Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education* 49, 5 (2024), 847–864.
- [39] Yang Ouyang, Leixian Shen, Yun Wang, and Quan Li. 2024. NotePlayer: Engaging Computational Notebooks for Dynamic Presentation of Analytical Processes. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–20.
- [40] Cassandra Overney, Belén Saldías, Dimitra Dimitrakopoulou, and Deb Roy. 2024. Sensemate: An accessible and beginner-friendly human-ai platform for qualitative data analysis. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 922–939.
- [41] 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.

- [42] Rock Yuren Pang, Hope Schroeder, Kynneddy Simone Smith, Solon Barocas, Ziang Xiao, Emily Tseng, and Danielle Bragg. 2025. Understanding the LLM-ification of CHI: Unpacking the Impact of LLMs at CHI through a Systematic Literature Review. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 456, 20 pages.
- [43] Antoine Ponsard, Francisco Escalona, and Tamara Munzner. 2016. PaperQuest: A visualization tool to support literature review. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2264–2271.
- [44] Napol Rachatasumrit, Jonathan Bragg, Amy X Zhang, and Daniel S Weld. 2022. CiteRead: Integrating localized citation contexts into scientific paper reading. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*. 707–719.
- [45] Gabriele Rosenthal. 2006. The narrated life story: On the interrelation between experience, memory and narration. University of Huddersfield.
- [46] Joshua Schimel. 2012. *Writing science: how to write papers that get cited and proposals that get funded*. OUP USA.
- [47] Salomé Schulze. 2014. Finding the academic self: Identity development of academics as doctoral students. *Koers: Bulletin for Christian Scholarship= Koers: Bulletin vir Christelike Wetenskap* 79, 1 (2014), 1–8.
- [48] Vidya Setlur, Enamul Hoque, Dae Hyun Kim, and Angel X. Chang. 2020. Sneak pique: Exploring autocompletion as a data discovery scaffold for supporting visual analysis. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, UIST'20*. ACM, Minneapolis, Minnesota, USA, 966–978.
- [49] 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.
- [50] Yijia Shao, Yucheng Jiang, Theodore A. Kanell, Peter Xu, Omar Khattab, and Monica S. Lam. 2024. Assisting in Writing Wikipedia-like Articles From Scratch with Large Language Models. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2024* 1, 2018 (2024), 6252–6278.
- [51] Zekai Shao, Leixian Shen, Haotian Li, Yi Shan, Huamin Qu, Yun Wang, and Siming Chen. 2025. Narrative Player: Reviving Data Narratives with Visuals. *IEEE Transactions on Visualization and Computer Graphics* (2025), 1–15.
- [52] Leixian Shen, Haotian Li, Yun Wang, Tianqi Luo, Yuyu Luo, and Huamin Qu. 2024. Data Playwright: Authoring Data Videos With Annotated Narration. *IEEE Transactions on Visualization and Computer Graphics* (2024), 1–14.
- [53] Leixian Shen, Haotian Li, Yun Wang, and Huamin Qu. 2024. From Data to Story: Towards Automatic Animated Data Video Creation with LLM-Based Multi-Agent Systems. In *IEEE VIS 2024 Workshop on Data Storytelling in an Era of Generative AI, GEN4DS'24*. IEEE, 20–27.
- [54] Leixian Shen, Haotian Li, Yun Wang, and Huamin Qu. 2025. Reflecting on Design Paradigms of Animated Data Video Tools. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, 1–21.
- [55] Leixian Shen, Haotian Li, Yifang Wang, Xing Xie, and Huamin Qu. 2025. Prompting Generative AI with Interaction-Augmented Instructions. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '25*. ACM, 1–9.
- [56] Leixian Shen, Yizhi Zhang, Haidong Zhang, and Yun Wang. 2024. Data Player: Automatic Generation of Data Videos with Narration-Animation Interplay. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 109–119.
- [57] Lei Shi, Hanghang Tong, Jie Tang, and Chuang Lin. 2015. VEGAS: Visual influence GrAph Summarization on Citation Networks. *IEEE Transactions on Knowledge and Data Engineering* 27, 12 (2015), 3417–3431.
- [58] John Stasko, Jaegul Choo, Yi Han, Mengdie Hu, Hannah Pileggi, Ramik Sadana, and Charles D Stolper. 2013. Citevis: Exploring conference paper citation data visually. *Posters of IEEE InfoVis* 2 (2013), 2–3.
- [59] Carola Strobl, Emilie Ailhaud, Kalliopi Benetos, Ann Devitt, Otto Kruse, Antje Prose, and Christian Rapp. 2019. Digital support for academic writing: A review of technologies and pedagogies. *Computers & education* 131 (2019), 33–48.
- [60] Hari Subramonyam, Roy Pea, Christopher Pondoc, Maneesh Agrawala, and Colleen Seifert. 2024. Bridging the Gulf of Envisioning: Cognitive Challenges in Prompt Based Interactions with LLMs. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 1039, 19 pages.
- [61] Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2024. Luminate: Structured generation and exploration of design space with large language models for human-ai co-creation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–26.
- [62] Sangho Suh, Hai Dang, Ryan Yen, Josh M. Pollock, Ian Arawjo, Rubaiat Habib Kazi, Hariharan Subramonyam, Jingyi Li, Nazmus Saquib, and Arvind Satyanarayan. 2024. Dynamic Abstractions: Building the Next Generation of Cognitive Tools and Interfaces. In *Adjunct Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) (UIST Adjunct '24). Association for Computing Machinery, New York, NY, USA, Article 91, 3 pages.
- [63] Nicole Sultanum, Christine Murad, and Daniel Wigdor. 2020. Understanding and supporting academic literature review workflows with litsense. In *Proceedings of the 2020 International Conference on Advanced Visual Interfaces*. 1–5.
- [64] Harold Thimbleby. 2004. Supporting diverse HCI research. In *Proceedings BCS HCI Conference*, Vol. 2. 125–128.
- [65] Nees Jan Van Eck and Ludo Waltman. 2014. CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *Journal of informetrics* 8, 4 (2014), 802–823.
- [66] Xingbo Wang, Samantha L Huey, Rui Sheng, Saurabh Mehta, and Fei Wang. 2024. SciDaSynth: Interactive Structured Knowledge Extraction and Synthesis from Scientific Literature with Large Language Model. *arXiv preprint arXiv:2404.13765* (2024).
- [67] Yun Wang, Zhitao Hou, Leixian Shen, Tongshuang Wu, Jiaqi Wang, He Huang, Haidong Zhang, and Dongmei Zhang. 2023. Towards Natural Language-Based Visualization Authoring. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 1222 – 1232.
- [68] Yun Wang, Leixian Shen, Zhengxin You, Xinhuan Shu, Bongshin Lee, John Thompson, Haidong Zhang, and Dongmei Zhang. 2024. WonderFlow: Narration-Centric Design of Animated Data Videos. *IEEE Transactions on Visualization and Computer Graphics* (2024), 1–17.
- [69] Yong Wang, Conglei Shi, Liangyue Li, Hanghang Tong, and Huamin Qu. 2018. Visualizing Research Impact through Citation Data. *ACM Transactions on Interactive Intelligent Systems* 8, 1 (2018), 1–24.
- [70] Meng Qi Yelena Wu, Robert Faris, and Kwan-Liu Ma. 2013. Visual exploration of academic career paths. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '13)*. ACM, 779–786.
- [71] Chang Xiao. 2023. AutoSurveyGPT: GPT-Enhanced Automated Literature Discovery. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–3.
- [72] Limin Yao, Jie Tang, and Juanzi Li. 2007. A unified approach to researcher profiling. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI'07)*. IEEE, 359–366.
- [73] Jian Zhao, Christopher Collins, Fanny Chevalier, and Ravin Balakrishnan. 2013. Interactive Exploration of Implicit and Explicit Relations in Faceted Datasets. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2080–2089.
- [74] Chengbo Zheng, Yuanhao Zhang, Zeyu Huang, Chuhan Shi, Minrui Xu, and Xiaojuan Ma. 2024. DiscipLink: Unfolding Interdisciplinary Information Seeking Process via Human-AI Co-Exploration. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–20.

A PROMPT FOR TOP-DOWN EXPLORATION (LINEAR EXAMPLE)

You are an AI assistant and an expert in Human Computer Interaction research. Your task is to help researchers structure a set of papers and craft coherent research narratives.

DATA & CONTEXT

- Paper set : {paper_set}
- Overall research focus : {overall_focus}
- Researchers intent : {researcher_intent}

KEY CONCEPTS

- Contribution Statement: a high-level criterion that partitions the entire paper set into meaningful, non-overlapping dimensions.
- Cluster: a subgroup of papers that share a distinctive feature under the same Contribution Statement.
- Cluster Theme: the abstract feature (e.g., method, technology, user group) that defines a Cluster.
- Paper: an individual publication, referenced only by its ID.

Hierarchy

1. Contribution Statement > Clusters > Papers. Each Paper appears in every Contribution Statement but in exactly one Cluster per Contribution Statement (no overlaps). Together, the Clusters must cover the entire paper set.

Rules for Linear Framework

Within one Contribution Statement, Clusters illustrate a sequential flow where papers are organized in a progressive line, highlighting how each stage builds upon insights from previous stages. Therefore, the Linear Framework is ideal for showcasing how initial explorations led to sophisticated outcomes or how works span different levels of advancement.

INSTRUCTIONS

Step 1: Differentiate Features Among Papers:

Examine each Paper to identify salient features (topics, technologies, research methodologies, target groups, interaction techniques, etc.).

Step 2: Form Preliminary Clusters and Abstract Cluster Themes

Group similar Papers into 3 to 6 Clusters. For each Cluster provide: 1) a concise, descriptive name (cluster_theme), and 2) a list of Paper IDs

Step 3: Synthesize Contribution Statements

Review the preliminary Clusters and abstract 4 to 6 higher-level, fundamental criteria that distinguish them. These criteria become the contribution_statements.

Important Constraints

- Focus on distinctions that genuinely help researchers articulate unique contributions.
- Avoid excessively broad, vague, or redundant categories.

OUTPUT FORMAT (strict JSON)

Return only valid JSON. No extra keys, comments, or text. Follow exactly the schema below. Use only Paper IDs inside "papers_assign".

```
{
  "contribution_statements": [
    {
      "contribution_statement": "Contribution Statement 1",
      "contribution_statement_description": "One-sentence explanation of this Contribution Statement and how its Clusters form a progression.",
      "clusters": [
        {
          "cluster_theme": "Cluster 1 Theme",
          "cluster_description": "Brief explanation of the Cluster Theme.",
          "papers_assign": ["id", "id"]
        },
        {
          "cluster_theme": "Cluster 2 Theme",
          "cluster_description": "Brief explanation of the Cluster Theme.",
          "papers_assign": ["id", "id"]
        }
      ]
    },
    {
      "contribution_statement": "Contribution Statement 2",
      "contribution_statement_description": "...",
      "clusters": [
        {
          "cluster_theme": "Cluster 1 Theme",
          "cluster_description": "...",
          "papers_assign": ["id"]
        }
      ]
    }
  ]
}
```


B PROMPT FOR BOTTOM-UP EXPLORATION (LINEAR EXAMPLE)

You are an AI assistant and an expert in Human-Computer Interaction research. You are given a structured research narrative that organizes papers into contribution statements and clusters. Your task is to refine the specific field indicated by `key_to_modify`, using the surrounding context for reasoning.

DATA & CONTEXT

The input is a JSON snippet that belongs to the following hierarchy: Contribution Statement > Clusters > Papers

KEYS USED

- `contribution_statement`: A high-level category that groups multiple research themes
- `cluster_themeX`: A specific research theme that falls under the `contribution_statement`
- `papers_assignX`: A list of paper IDs assigned to the corresponding cluster
- `key_to_modify`: The key whose value should be updated

INPUT FORMAT

```
{
  "contribution_statement": "{contribution_statement}",
  "cluster_theme0": "{cluster_theme0}",
  "papers_assign0": ["id", "id"],
  "cluster_theme1": "{cluster_theme1}",
  "papers_assign1": ["id"],
  "...": "...",
  "key_to_modify": "{key_to_modify}"
}
```

INSTRUCTIONS

Step 1: Examine the Context of `key_to_modify`:

- If it is a `cluster_themeX`: Examine the papers in the associated `papers_assignX`. Identify their shared characteristics (such as research method, user group, interaction type, or technology), and summarize them into a concise, meaningful cluster theme that reflects the essence of this group.
- If it is a `papers_assignX`: Based on the corresponding `cluster_themeX` and the overall `contribution_statement`, assign the most relevant papers from the dataset to this cluster. Ensure that the assignment is coherent with the theme and mutually exclusive with other clusters under the same `contribution_statement`.
- If it is a `contribution_statement`: Consider all available cluster themes and papers to propose a more coherent and meaningful statement.

Step 2: Check the Relationship Among Clusters:

- Check that the Clusters illustrate a sequential flow where papers are organized in a progressive line, highlighting how each stage builds upon insights from previous stages. Make sure the Clusters are ideal for showcasing how initial explorations led to sophisticated outcomes or how works span different levels of advancement.
- Confirm that each paper appears in exactly one cluster within this `contribution_statement`.

Step 3: Update the New Value

Locate the field indicated by `key_to_modify`. Replace its value with an improved one-line description. Do not modify any other fields.

OUTPUT

Return only the new value you assign to the key. No extra keys, comments, or text.

```
{new_value}
```