

VISAR: A Human-AI Argumentative Writing Assistant with Visual Programming and Rapid Draft Prototyping

Zheng Zhang
University of Notre Dame
Notre Dame, IN, USA
zzhang37@nd.edu

Ranjodh Singh Dhaliwal
University of Notre Dame
Notre Dame, IN, USA
rdhaliwa@nd.edu

Jie Gao*
Singapore University of Technology and Design
Singapore, Singapore
gaojie056@gmail.com

Toby Jia-Jun Li
University of Notre Dame
Notre Dame, IN, USA
toby.j.li@nd.edu

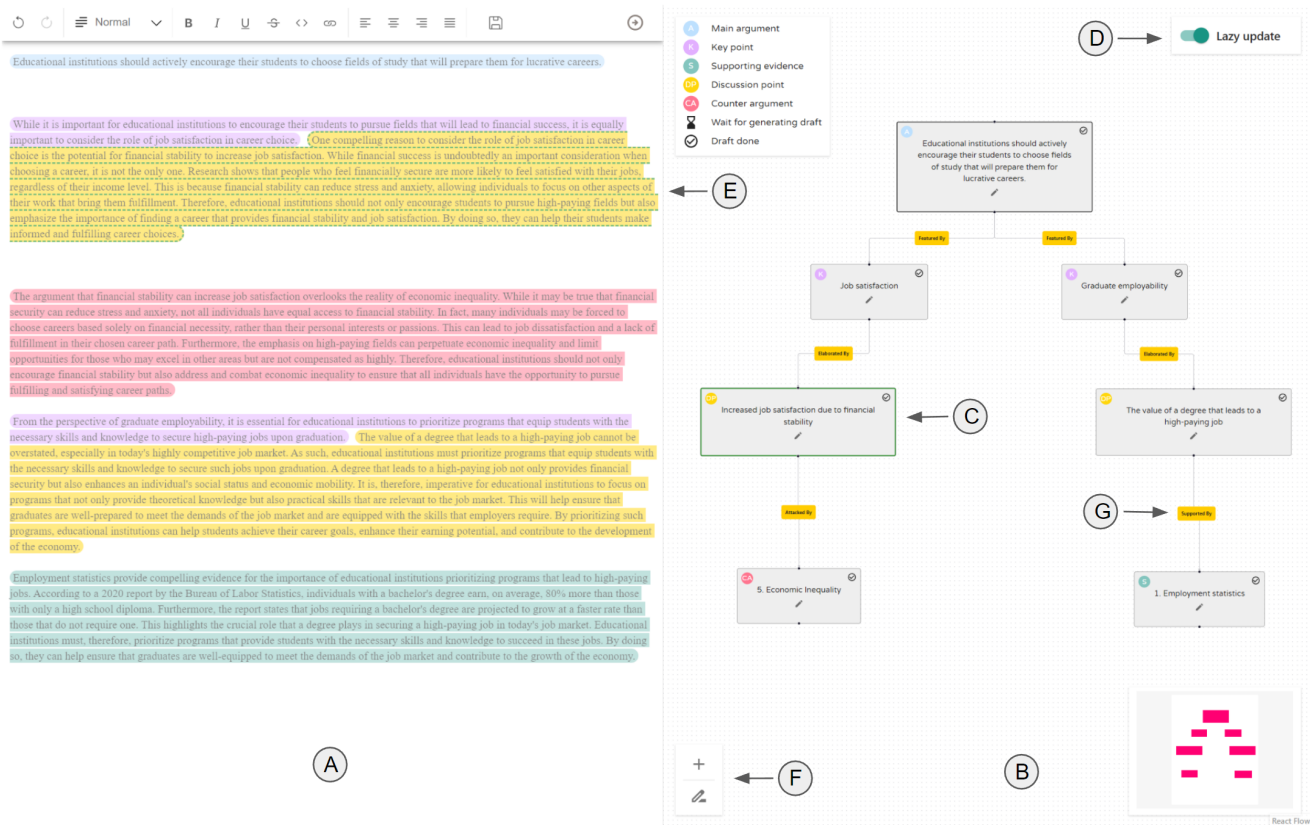


Figure 1: The main interface of VISAR. The user can edit their argumentative writing outline via either the text editor (A) or the canvas of the visual outline (B); VISAR synchronizes these two workspaces and facilitates users to correspond a writing item in editor (E) to the counterpart node in visual outline (C); User can switch between *Lazy* or *non-Lazy* mode to decide when to generate draft prototypes during the planning (D); User can choose between multiple relationships between items in outline (G); User can also add new item to the visual outline and generate a new draft from the outline (F).

*Work done as a visiting student at the University of Notre Dame.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or

republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
UIST '23, October 29–November 1, 2023, San Francisco, CA, USA
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0132-0/23/10...\$15.00
<https://doi.org/10.1145/3586183.3606800>

ABSTRACT

In argumentative writing, writers must brainstorm hierarchical writing goals, ensure the persuasiveness of their arguments, and revise and organize their plans through drafting. Recent advances in large language models (LLMs) have made interactive text generation through a chat interface (e.g., ChatGPT) possible. However, this approach often neglects implicit writing context and user intent, lacks support for user control and autonomy, and provides limited assistance for sensemaking and revising writing plans. To address these challenges, we introduce VISAR, an AI-enabled writing assistant system designed to help writers brainstorm and revise hierarchical goals within their writing context, organize argument structures through synchronized text editing and visual programming, and enhance persuasiveness with argumentation spark recommendations. VISAR allows users to explore, experiment with, and validate their writing plans using automatic draft prototyping. A controlled lab study confirmed the usability and effectiveness of VISAR in facilitating the argumentative writing planning process.

CCS CONCEPTS

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

KEYWORDS

human-AI collaboration, writing support, creativity support

ACM Reference Format:

Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. 2023. VISAR: A Human-AI Argumentative Writing Assistant with Visual Programming and Rapid Draft Prototyping. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*, October 29–November 1, 2023, San Francisco, CA, USA. ACM, New York, NY, USA, 30 pages. <https://doi.org/10.1145/3586183.3606800>

1 INTRODUCTION

Argumentative writing is a meta-genre of writing that is often used by individuals to persuasively convey their thoughts on disputable topics [5, 20, 48, 79]. Unlike other forms and genres of writing, such as narrative and descriptive writing, argumentative writing requires writers to justify their claims with evidence and logical reasoning [5, 18, 53]. Additionally, they must carefully identify and address potential counterarguments in order to construct a compelling argument [5].

However, the argumentative writing process presents significant challenges, requiring considerable effort from writers. From a cognitive perspective, the challenge stems from the inherent conflict between two levels on which the writer must simultaneously operate. First, writers must create and maintain a hierarchical structure of arguments and sub-arguments to ensure global logical coherence and comprehensiveness of their argumentation [33]. This involves brainstorming constructs across various levels of abstraction, including high-level key aspects, specific discussion points and sub-arguments, evidence requirements, and potential counterarguments [5]. Second, at the same time, writers must focus on the lower-level logical relationships between the discussion points to ensure their validity and coherence [17]. This writing process is both *iterative* and *non-linear* as specific content for discussion

points and sub-arguments helps writers solidify their ideas when planning argument structure, while high-level structures support their efforts to maintain global logical coherence and comprehensiveness when working on specific texts [45].

Although several interactive tools have recently been introduced to assist the argumentative writing process, especially with the advancement of large language models (LLMs) such as the GPT family [10], these tools often focus on one of the two levels without effectively connecting them. For example, at the higher level of ideation, Sparks [22] facilitates ideation for scientific writers by using an LLM to suggest potential ideas related to given scientific concepts. Similarly, CoAuthor [39] provides writers with context-based argumentative suggestions, inspiring them with the next piece of content to write. At the lower level of concrete text, some systems prioritize giving feedback on pre-existing drafts to help writers recognize weaknesses and adhere to argumentative structure in order to improve their writing [2, 72, 73]. While these tools have proven useful for argumentative writers, they typically address specific writing constructs within a particular phase of the writing process (e.g., ideation, drafting, and revision) or a specific level of abstraction (e.g., global argumentative structures or local writing suggestions). Consequently, there remains an unmet challenge in developing effective methods to assist argumentative writers in generating and structuring their ideas through hierarchical writing planning and in seamlessly transitioning between global and local levels of abstraction across planning and drafting phases.

To address these gaps, we introduce VISAR¹², a human-AI collaboration system that supports writers in the hierarchical and iterative planning process of argumentative writing.

VISAR incorporates two novel approaches: *visual programming* and *rapid prototyping* for argumentative writing. The interface of VISAR features side-by-side text editors and an interactive graph of logical relationships among entities for writing planning. During the ideation process, VISAR employs a state-of-the-art LLM to assist writers in interactively exploring potential discussion points using a “chain of thought” approach [76] (Figure 3). Entities and their relationships are visualized in a tree structure that users can freely manipulate and edit (Figure 5). Entities are connected through relationships, and writers can add, edit, or remove entities and connect them using various relationship types in a visual programming interface. VISAR also helps writers identify potential counterarguments, supporting evidence, and logical flaws for entities in the current argumentative plan (Figure 4), synchronizing both the visualization and the text editor. This method allows writers to monitor and evaluate the ideation process across different abstraction levels and facilitates the gradual refinement of the language model’s thoughts in a step-by-step manner, enabling it to evolve its context understanding and provide context-consistent suggestions that respond to writers’ needs.

VISAR’s rapid prototyping approach generates concrete text of what the final write-up of the argumentation *could be* based on the current argumentation structure using state-of-the-art LLMs. It is important to note that VISAR’s objective is *not* to produce the final write-up, as LLMs still have limitations in, e.g., achieving the

¹VISAR stands for Visual Interactive System for Argumentative writing with Rapid draft prototyping

²Live demo website: <https://visar.app>

desired levels of factual correctness and specificity. Instead, VISAR aims to generate *mid-fidelity* (hi-fidelity in form and logical flow; low-fidelity in factual details) prototypes of the write-up to facilitate fast and iterative exploration of ideas. Writers first “diverge” by selecting different key points, argumentation structures, and supporting points, and then “converge” by examining the generated write-up prototypes to understand the strengths, weaknesses, and trade-offs of various approaches, reaping the benefits of *parallel prototyping* in user experience design [49].

The design of VISAR embodies our vision of effective human-AI collaboration on complex and ambiguous tasks with multiple intertwined layers of abstraction, particularly in an era when LLMs like GPT-4 are more powerful than ever. Recently, interfaces that wrap LLMs into multi-turn chatbots, such as ChatGPT, have gained considerable attention. Users find them capable of generating *seemingly* reasonable argumentative essays from simple prompts (e.g., “Can you write an argumentative essay about...”) or auto-completion. Furthermore, the ability to maintain context in multi-turn conversations allows them to support *highly flexible* follow-up requests on both language style (e.g., “Can you revise this section in the style of academic writing?”) and content (e.g., “Can you come up with a specific example for this?”) levels. However, we argue that the chat interface is *not* the ultimate solution for tasks like this. As evidenced by the influential Direct Manipulation vs. Agent debate [64] over two decades ago, a direct manipulation interface offers several key advantages over agents, including better transparency into the system’s state, easier error handling, finer granularity of user control, useful constraints to guide user actions, and more visible system affordances. The design of VISAR incorporates these guidelines by introducing a visual outline of the argumentation structure and enabling object-level direct manipulation (selecting, dragging, actions) on both visual and text representations. In designing VISAR, our goal is to retain the flexibility provided by LLMs while offering users more control, constraints, and transparency to guide users through the task.

The results of a controlled lab evaluation showed that users were able to successfully use VISAR to assist with their argumentative writing planning process. Participants found the various intelligent features and interaction strategies in VISAR useful and expressed interest in incorporating VISAR into their own argumentative writing planning process.

In summary, this paper presents the following contributions:

- a new approach that uses visual programming and rapid prototyping strategies to achieve effective collaboration between human writers and LLMs with adequate user control and autonomy in the ideation and planning stages of argumentative writing.
- VISAR, a writing assistance tool that implements this approach to support writers in argumentative writing tasks.
- a within-subjects user study with 12 participants that validated the usability and effectiveness of VISAR.

2 RELATED WORK

2.1 Writing Support Tools

The development of tools to support users in the writing process has been a long-standing area of interest. On the commercial side, tools

for grammar and spelling check, like the built-in tool of Microsoft Word and third-party tools (e.g., Grammarly³) predominantly address language correctness without providing guidance on content. Beyond grammar and spelling, tools like Ref-N-Write⁴ expand on this by focusing on language style, helping users paraphrase their writing to achieve more concise, professional, or academic expressions. Ref-N-Write also offers practical suggestions for developing research paper outlines and drafts, as well as providing diverse inspirations for academic writing in various fields [46].

The incorporation of natural language generation (NLG) technology into writing support tools has become increasingly prevalent in recent years. Commercial platforms, such as the online AI-powered Smodin⁵, assist users with rewriting and employ text generation technology to produce essay outlines for different genres, including argumentative and descriptive writing.

In the academic domain, extensive research has been conducted on general writing support. For creative writing, Kreminski et al. [35] investigated the need for writers, including maintaining narrative consistency, developing plot structure, crafting reader experiences, and refining expressive intent. To support human-AI collaboration for *creative writing* (as opposed to argumentative writing), Coenen and Ippolito et al. [14, 28] introduced Wordcraft, a story-writing tool that rewrites, elaborates, or provides personalized prompts for subsequent text. CoAuthor [39] is an interactive writing tool that leveraged the GPT-3 model to offer writers ideas for their next writing steps in both creative and argumentative writing, while Chakrabarty et al. developed CoPoet [11], a collaborative poetry tool system based on LLMs that generates poetry in response to user-provided text and customized prompts. Yang et al. [82] explored interaction strategies used by human writers to condense, revise, summarize, and regenerate AI-generated text in fictional writing. The work of Singh et al. [65] explored a multimodal technique that visualizes story plots to support ideation in creative writing. An empirical study [4] by Bhat et al. suggested that when writers encounter next-phase suggestions from an LLM, they take aid from them in multiple nuanced ways even if they do not directly accept the suggestions.

For tools that support specific parts of the argumentative writing process, Dang et al. [15] proposed a text editor that generates continuous text summaries alongside the main text, which was shown useful in helping users revise the content and scope of their drafted paragraphs. For final reflection and quality improvement, Wambganss et al. [73] introduced AL, an *Argumentative writing* support system that provides feedback on argument quality by identifying argument components (e.g., claims, premises, non-argumentative elements) and recognizing argument relations. Xia et al. [81] presented Persua, which employs machine learning models to detect components of argumentative writing and their relationships, and provides example-based guidance on persuasive techniques to improve persuasiveness of arguments. Wong et al. [80] developed a tabletop system to promote argumentation in computer science students. On the machine learning research side of augmenting the argumentation capabilities of LLMs, Yang et al. [83] introduced Re³,

³<https://www.grammarly.com/>

⁴<https://www.ref-n-write.com/>

⁵<https://smodin.io/>

a Recursive Reprompting and Revision framework for automatically generating longer stories by iteratively creating new prompts using intermediate information.

Our work with VISAR complements the existing literature by focusing on interaction strategies and collaborative workflows for human-AI collaboration in two crucial early stages of argumentative writing: prewriting and planning. VISAR proposes a new workflow that employs visual programming and rapid prototyping methodologies to ensure user flexibility, autonomy, and control within the established process of argumentative writing.

2.2 Human-AI Collaboration in Creative Work

Cognitive theory [34] models creative work as a four-stage process: *preparation, incubation, illumination, and evaluation*. In addition, creativity is said to involve two thinking modes: *divergent* and *convergent* thinking [32, 51]. Divergent thinking is most commonly associated with the process of generating original ideas and breaking away from conventional ways of thinking. Once a broad range of ideas has been generated through divergent thinking, convergent thinking can be used to evaluate, refine, and implement these ideas.

Recently, there are some works exploring to use AI models to support people’s workflow of creative work [28, 35, 47]. Examples include collaborative story writing tools such as TaleBrush [13], which allows users to create stories with AI assistance through sketching interactions, indicating different story stages such as positive or negative events. TaleBrush primarily aids the planning stage of writing, enabling writers to generate diverse story lines and refine them iteratively. In a different context, Storybuddy [87] uses question-answer generation (QAG) to create question-and-answer pairs based on the context of a story, facilitating interactive storytelling between parents and children. Similarly, Lee et al. [40] suggested using AI to collaboratively rewrite stories for children and parents, while Akoury et al. [1] presented STORIUM, an evaluation platform that allows users to request language model suggestions during story writing. In addition, CodeToon [67] is an AI-powered story ideation and auto-comic generation tool that maps from code.

Applying AI to creative work presents unique challenges. Creative work requires more than just the spark of an idea—it must exhibit originality and effectiveness [59] of the content creator. While AI models, especially large language models, are good at producing diverse and in-context content on the fly, ensuring the originality and authenticity of their outputs is challenging. For example, generative AI models are known to have tendency to produce outputs that deviate from reality, often referred to as “hallucination” [31]. Users must take responsibility for verifying the credibility of the sources of the generated content. Additionally, the authorship of AI-generated creative work might spark controversy as well [3].

Our work focuses on employing AI to streamline the initial prewriting and planning stages of argumentative writing. The goal is to inspire writers and generate early drafts that serve mainly to enhance understanding and make sense of complex writing topics, rather than incorporating AI-generated content in the final product [21], a practice that could raise ethical concerns.

2.3 Prompting and Prompting Chain in LLMs

VISAR interacts with its underlying large language model (LLM) through *prompting* [44]. Specifically, a new diagram, “pre-train, prompt, and predict” [44], has recently been proposed. To achieve better performance, diverse downstream tasks are reformulated to resemble the type of task the model was originally trained on, rather than adapting the model’s objective function to the downstream task. Hence, LLMs are often considered “general purpose” pre-trained models.

To guide an LLM in generating results, various prompting strategies can be employed [57]. Zero-shot prompting involves a question-answer mechanism, while one-shot prompting provides one example for the AI to base its output on. Few-shot prompting involves providing the AI model with multiple examples, and role prompting requires the AI to complete a task based on a given role rather than examples or templates. Notably, Wei et al. [76] introduced the chain of thought (CoT) to elicit reasoning from the model. This technique enables models to break down complex, multi-step problems into intermediate steps by offering a sequence of prompts that guide them towards a final objective. Implementing CoT significantly enhances the performance of previous language models in reasoning tasks previously considered “off-limits” for language models, even achieving human-parity performance in some tasks [68].

Argumentative writing, as described in Section 3, can be decomposed into several steps, such as proposing issues, claims, and evidence. Importantly, writing is an iterative process that requires writers to continuously refine their thoughts. This process bears a resemblance to the concept of CoT. In fact, Yang et al. [83] demonstrated that the *Re*³ framework can generate longer stories with a CoT approach that implements recursive reprompting and revision. In this paper, VISAR uses a human-in-the-loop prompt chain approach for argumentative writing. By combining the strengths of CoT and human expertise, our approach aims to enhance the effectiveness of human-AI collaboration in argumentative writing.

3 BACKGROUND OF ARGUMENTATIVE WRITING

Argumentative writing is a type of writing in which the writer presents and supports a position on a controversial issue using evidence and reasoning, with the goal of persuading the reader to accept the viewpoint of the writer or take a specific action [5, 20, 48, 79]. This writing style relies on arguments as foundational components, where each argument posits a statement on an issue, substantiated with evidence and reasoning. Multiple arguments can be used to build a larger argument or present a comprehensive perspective on a subject.

Toulmin’s argument model [26] provides a framework for formulating logical arguments. It consists of six components: *claim* (the main point to be supported), *data* (evidence validating the claim), *warrant* (the underlying assumption or rationale linking the data to the claim), *backing* (supplementary evidence or reasoning supporting the warrant and strengthening the argument), *qualifier* (a statement limiting the claim’s scope or force), and *rebuttal* (counterarguments or objections to the claim). This model underscores the necessity of evidence and reasoning to reinforce claims while

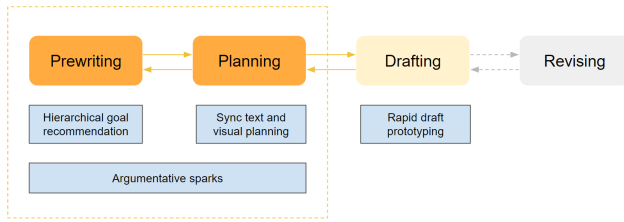


Figure 2: Four cognitive writing stages [19, 24] and how VISAR’s features support them. VISAR primarily focuses on *prewriting* and *planning* stages in argumentative writing. VISAR also assists with rapid draft prototyping as a way to facilitate more effective iterations in the previous two stages.

anticipating and addressing potential objections or counterarguments. Grounded in theoretical argumentation models, argument mining [37, 71] seeks to autonomously detect and extract the structure of inferential logic and reasoning articulated as arguments in natural language. This technique aids in evaluating individuals’ stances and the underlying rationales they employ, helping writers identify the structural and logical weakness in their arguments. Informed by Toulmin’s model and argument mining works, VISAR aims to prompt writers about how to strengthen arguments by alerting them to potential counterarguments (rebuttals), supplying supporting evidence (data, backing), and scrutinizing logical fallacies primarily related to claim, warrant, and qualifier. Compared with traditional argument mining approaches, using LLM to inform argument structure has advantages such as adaptability across various domains, obviating the requirement for dataset collection, and enhanced accuracy in contextual comprehension.

Drawing on Toulmin’s model, VISAR aims to assist writers in strengthening their arguments by alerting them to potential counterarguments (rebuttals), providing supporting evidence (data, backing), and examining logical fallacies, primarily concerning the claim, warrant, and qualifier.

Argumentative writing can be categorized into two types based on the direction of reasoning [6, 53, 54, 62]: *deductive* and *inductive*. *Inductive argumentation*, a *bottom-up* approach, starts with specific observations or instances and progresses toward broader generalizations or theories. Inductive reasoning examples include scientific, statistical, and analogical argumentation. Conversely, *deductive argumentation*, a *top-down* process, begins with general principles or premises and proceeds to specific conclusions through logical reasoning. This approach involves starting with an established theory or hypothesis and using it to derive particular implications, predictions, or conclusions.

VISAR focuses on facilitating deductive, top-down argumentative writing. Compared to inductive arguments, deductive arguments are more common in written communication [30]. In situations such as decision making, problem solving, or expressing opinions, the top-down strategy enables individuals to communicate their ideas coherently and persuasively [30].

Theories of the writing process [19, 24] identify four primary cognitive stages of the writing process: *prewriting*, *planning*, *drafting*, and *revising* (Figure 2). **VISAR primarily concentrates on**

assisting writers during the prewriting and planning stages. However, to support the iterative and non-linear nature of the writing process, we also enable users to rapidly generate draft prototypes, allowing them to review and revise their current outlines within a tangible writing context. In the prewriting stage of deductive writing, writers gather information and brainstorm hierarchical writing objectives, including high-level key aspects, discussion points, and supporting evidence. This stage can generate a high cognitive load, as writers must retrieve information and sequentially elaborate their thought process. In the planning stage, developing a hierarchical and logical structure for the argumentation outline can be challenging in deductive writing. Writers must ensure that their premises logically lead to the intended conclusion with necessary evidence and settlement of counterarguments, while maintaining coherence and consistency throughout the planning.

Here, we define the key terms we use throughout this paper:

- **Argument:** A statement or group of statements called premises intended to determine the degree of truth or acceptability of another statement called a conclusion [29].
- **Argumentation:** A process of constructing, presenting, and defending a position or claim through the use of logical reasoning, evidence, and persuasive techniques [5].
- **Writing goal:** An intended outcome that a writer seeks to achieve while composing an argumentative piece in order to persuade audiences to accept the central claim [19].
- **Key aspect:** A primary theme or central concept that guides the development and organization of argumentation to support the central claim from a particular perspective [19].
- **Discussion point:** A specific point or idea that supports the central claim related to a particular theme [19].
- **Counter argument:** An opposing viewpoint or objection that challenges the central claim, or position presented by the writer [26].
- **Evidence:** Information, facts, data, or examples that support the writer’s central claim or argument [26].
- **Logical fallacy:** The use of flawed or unsound reasoning in the development of an argument, which may appear to be a well-reasoned argument if unnoticed [5].
- **Textual outline:** A structured text-based framework that organizes the writing components such as main ideas, arguments, counterarguments, and supporting evidence. An example is nested bullet points [43].
- **Visual outline:** A graphical representation of the structure and organization of argumentation that highlight the logical connections among components. Examples include mind maps, flowcharts, and hierarchical diagrams [43].

4 VISAR SYSTEM

4.1 Design Goals

Our goal is to develop a human-AI collaborative system that facilitates writers in the hierarchical and iterative planning process of argumentative writing. To inform our system design, we conducted a semi-systematic literature review using method proposed by Synder [66] to gain insights into (1) the cognitive processes of argumentative writing, (2) the challenges faced by argumentative writers, and (3) the capabilities and limitations of existing writing

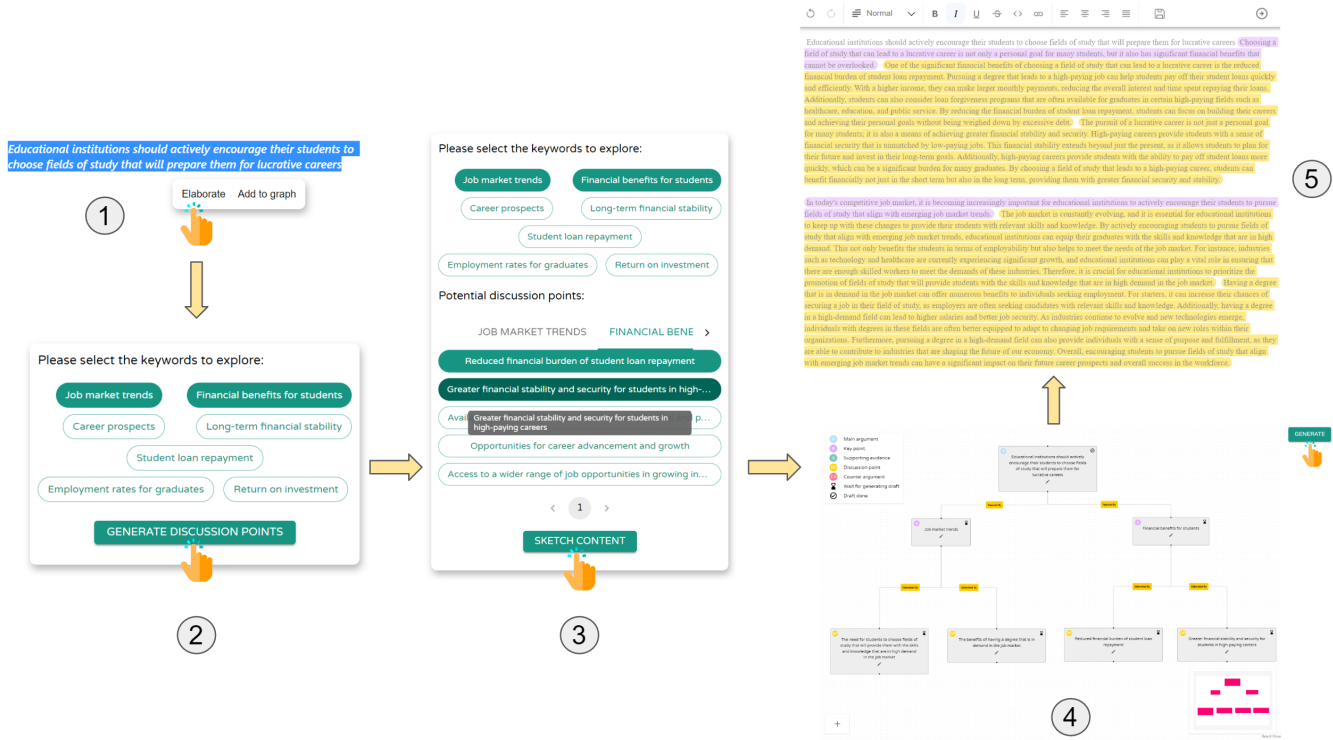


Figure 3: The interaction flow of hierarchical goal recommendation: (1) Users can select a range of text and click *Elaborate* to invoke hierarchical goal recommendation; (2) VISAR proposes a set of possible key aspects, users can select key aspects that they want to further explore; (3) VISAR proposes specific discussion points for each selected key aspect; (4) Users can review and modify the the logical structure of the writing goals. By click *Generate*, users can use VISAR to implement the draft for the current writing outline; (5) VISAR generates a prototype of the draft in the text editor.

assistance tools. Specifically, we first identified seminal and extensively referenced works in the fields of cognitive writing theory, argumentation theory, writing pedagogy theory and writing assistance tools. This step helped us gain a fundamental understanding of the aforementioned questions. Afterwards, we expanded our inquiry to more recent works in those fields via keyword searches and by tracing the citation networks from the key works recognized in the preceding stage. This allowed us to assimilate recent advancements in these domains as well as understand the persisting challenges. Through the process of this literature review, we identified four design goals that our system ought to achieve:

DG1: Accommodating the non-linear and hierarchical nature of writing planning process. Rhetorical theories often conceptualize writing as a goal-directed process, in which writers must establish a hierarchical network of goals to guide their efforts [19, 53]. In argumentative writing, these objectives involve introducing the topic and main argument, elaborating on key points and sub-arguments supporting the central argument, providing evidence for the same, and addressing potential counterarguments. However, users typically determine local goals only after engaging in a hierarchical planning process that provides a concrete context [45]. Furthermore, instead of following a linear process, writers frequently revise goals or adjust their relationships as they gain deeper understanding during the planning phase [23]. Due to

the hierarchical structure of planning, these revisions may cascade and affect subordinate subgoals, making it difficult to maintain coherence [17]. To address these characteristics, the system should help writers construct a hierarchical sequence of writing goals while concurrently aiding them in maintaining overall coherence during goal revision.

DG2: Facilitating the ideation of writing plans. Writing sometimes begins as a serendipitous experience. Writers may start planning without a clear endpoint or a comprehensive understanding of required components [17, 19]. Additionally, the writing planning process requires writers to quickly retrieve relevant information from long-term memory in accordance with the current local goal [19]. However, this information may be underdeveloped or disorganized, which can lead to omissions [77]. In response, our proposed system should provide writers with diverse options at each ideation stage, inspiring possible subsequent writing goals based on the existing context.

DG3: Scaffolding to enhance persuasiveness of users' argumentation. Argumentative writing requires writers to justify their arguments' credibility through persuasive strategies (e.g., *logos*, *ethos*, *pathos*) and addressing counterarguments and logical fallacies [5, 58]. However, maintaining persuasiveness can be challenging. Writers may be unfamiliar with persuasive strategies and

common types of logical fallacy or struggle to select and implement suitable strategies for a specific context. Therefore, our system should prompt writers with relevant persuasive strategies, highlight potential logical fallacies, and present counterarguments tailored to the specific local context in which the writers operate.

DG4: Facilitating users' reflection and assessment of plans.

The planning and drafting stages are often intertwined in the writing process. As writers begin to draft, they may realize that some aspects of their outline are incomplete or lack sufficient detail, possibly due to missing evidence, weak arguments, or unclear transitions between sections [19, 24]. Drafting allows writers to identify gaps, return to the planning stage, and revise their outline accordingly [17, 19]. It also enables writers to reassess ideas and restructure their plan. However, transforming ideas from planning into a coherent, well-structured piece of writing can be time-consuming and laborious. To alleviate these efforts, our system should support users in reflecting on and evaluating their plans through automated rapid draft prototyping.

4.2 Example scenario

In this section, we provide an illustrative example of how VISAR operates in a practical context. Imagine that Alice needs to write an argumentative essay supporting the statement “Universities should require every student to take a variety of courses outside the student’s field of study”. Unsure of the best structure and needing guidance to create a persuasive argument, Alice decides to use VISAR to plan and draft her essay.

Alice begins by entering the statement as her initial sentence in VISAR (users may also choose to write more themselves before seeking AI assistance). After highlighting the text and clicking the *Elaborate* button (Figure 3-1), VISAR provides Alice with step-by-step writing goal recommendations. First, it suggests key aspects of the argument that Alice may consider, such as *breadth of knowledge, well-rounded education, career preparation, and development of critical thinking skills*. Alice selects *well-rounded education* and *career preparation* for further exploration.

By clicking the *Generate discussion points* button (Figure 3-2), VISAR generates potential points related to the two selected aspects. For *well-rounded education*, it prompts Alice to consider the *benefits of exposure to diverse thinking styles* and the *opportunities for lifelong learning and personal growth*. For *career preparation*, VISAR recommends addressing *preparation for interdisciplinary jobs* and *increased adaptability in a fast-changing job market*. Alice selects the points of lifelong learning and adaptability to the changing market and clicks the *Sketch Content* button (Figure 3-3) to review the current logical structure of her writing goals.

Alice then clicks the *Generate* button (Figure 3-4) to request prototype drafts for each goal in her outline, which she can directly edit. She also asks VISAR to generate alternative drafts for the point on *lifelong learning* and *personal growth* and refines the draft through conversation, providing further instructions.

To identify counterarguments, logical fallacies, and necessary supporting evidence in her drafts, Alice uses VISAR’s *Argumentative Sparks* feature. She selects the draft block and clicks the *Argumentative sparks* button (Figure 4-1), which provides her with likely counterarguments, potential logical fallacies, and suggestions

for supporting evidence. VISAR also enables Alice to view draft implementations of counterarguments and supporting evidence, along with revised drafts of original discussion points that address logical weaknesses.

To generate a visual outline, Alice takes advantage of VISAR’s visual programming feature. A visual outline of the current argumentation outline is automatically generated, synchronized, and displayed next to the text editor. Using the intuitive drag-and-drop interface (Figure 1-B), she can add, edit, and rearrange goals within her outline (Figure 5). For instance, Alice may want to include a new key aspect, *development of critical thinking skills*, by creating a new node and connecting it as a child of the main argument node. As she modifies the visual outline, Alice can immediately view the updated draft in a pop-up window and update dependent nodes and goals in-situ. By activating the *Lazy update* mode (Figure 1-D), Alice can focus on editing the visual outline and update the entire draft simultaneously by clicking the *Generate text* (✎) button (Figure 1-F). Additionally, Alice can edit an argument or goal in the text editor and incorporate it into the visual outline by clicking the *Add to graph* button (Figure 3-1).

4.3 Key Features

This section introduces the key features of VISAR, as illustrated in the example scenario. These features include hierarchical writing goal recommendation, synchronized text and visual planning, argumentative sparks, and automatic draft prototyping.

4.3.1 Hierarchical writing goal recommendation. As noted in DG1 and DG2, writers often face challenges when generating a comprehensive set of potential goals at each planning level for their writing. VISAR addresses this need by providing users with a step-by-step goal recommendation process, allowing them to direct goal generation at each level. The design of VISAR facilitates the integration of user-defined and AI-generated goals, promoting user autonomy.

Inspiring writer with step-by-step writing goal recommendation. Theories of the writing process [19, 24] suggest that during the planning phase of writing an article, writers typically begin by brainstorming a wide range of relevant topics. They then identify specific points to support the core argument related to the chosen topics. To accommodate this process, VISAR prompts writers in a “chain of thought” approach, allowing them to explore various potential topics connected to their core argument. VISAR then provides recommendations for specific discussion points that can be further developed based on the selected topics.

To perform this step-by-step planning with an LLM, writers can select the text containing the main argument. By clicking the *Elaborate* button in the displayed menu (Figure 3-1), writers are presented with a set of suggested potential aspects or topics that can help substantiate the selected argument. Writers can review and select topics for further exploration. When writers click the *Generate discussion point* button (Figure 3-2), the system prepares specific discussion points organized by the selected topics. Writers can navigate these by clicking the corresponding tabs and select discussion points to include in their outline (Figure 3-3). After writers selected the points to discuss, VISAR will show the current visual outline (see details in Section 4.3.2) for writers to review

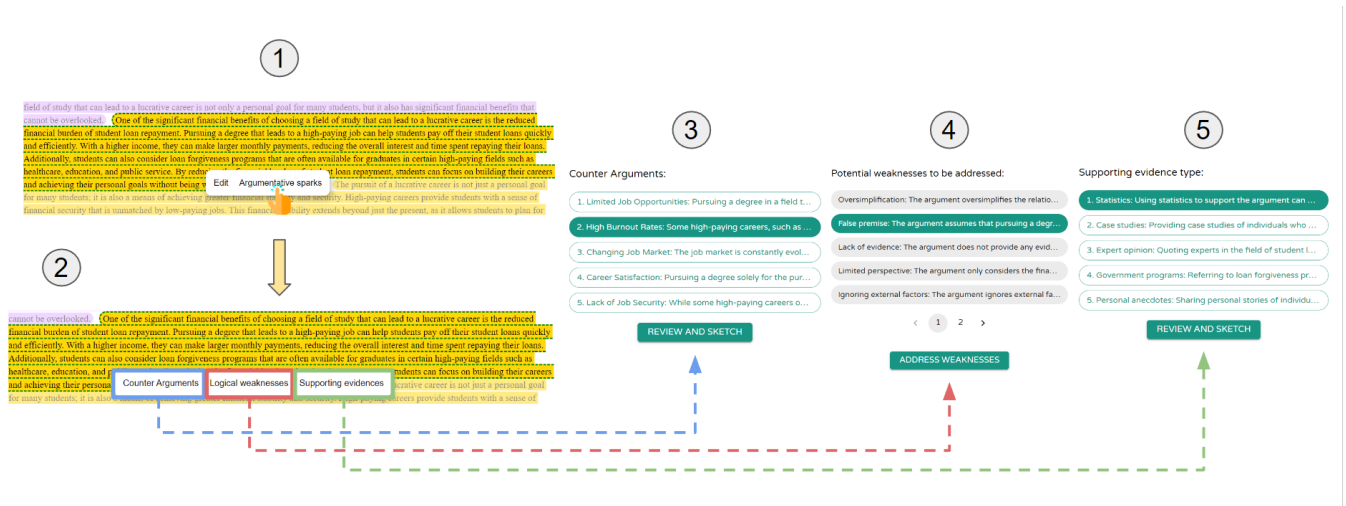


Figure 4: The interaction flow of argumentative sparks: (1) users can invoke argumentative sparks by clicking on a text block; (2) users can select the specific spark type; (3) VISAR proposes counterarguments for the selected discussion point; (4) VISAR identifies logical fallacies in the selected content; (5) VISAR recommends supporting evidences for the selected content.

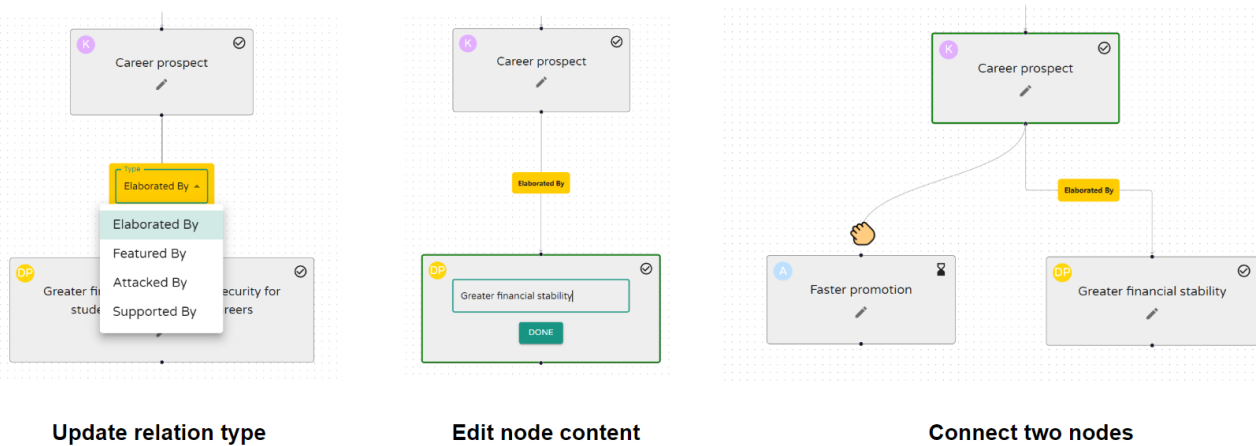


Figure 5: Illustration of three types of manipulations on the visual outline

the outlining structure. Writers can also click the *Generate* button (Figure 3-4) to produce a prototype write-up that implements this outline (see details in Section 4.3.4).

Integration of user-defined goals with AI-generated goals. VISAR allows writers to interweave their own defined goals with AI-generated goals during the ideation process. Writers can enter a goal description in the text editor, select the range, and click the *Add to graph* button to insert a new node in the visual outline that corresponds to the user-defined goal (see details in Section 4.3.2). Writers can also create the node directly by clicking *Add (+)* button (Figure 1-F) and editing the node description. After adding the node, writers can position it appropriately in the visual outline by dragging (Figure 5). Writers can collaborate with the LLM to further

recursively elaborate their defined goals or any AI-generated discussion point by clicking the *Elaborate* button on the corresponding text description.

4.3.2 Synchronized text and visual planning. As highlighted in DS4, VISAR supports both text-based and visual-based writing planning, catering to diverse planning strategies. This design combines the benefits of “WYSIWYG (What You See Is What You Get)” directness with text editing and the flexibility and clarity of organizing argumentation structure through visual programming.

Enabling interactive visual planning of writing. VISAR allows users to plan their writing by constructing a hierarchical tree-based outline on an interactive canvas that represents their writing outline. This visual representation is similar to existing visual programming

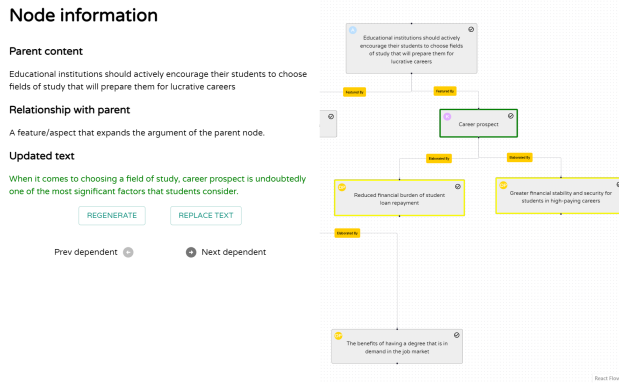


Figure 6: When the content of a node has changed, VISAR prompts users to update all of its dependent nodes recursively.

and online whiteboard tools such as Miro⁶. The nodes represent individual components or discussion points in the outline. The directed edges between nodes represent their relationships. Writers can easily create, remove, or edit nodes and edges through direct manipulation. In specific, VISAR supports five types of nodes: *main argument* (default), *key aspect*, *discussion point*, *counterargument and supporting evidence* and four types of edges: *featured by* (child is a high-level topic/aspect regarding the parent), *elaborated by* (child is a concrete discussion point regarding the parent), *attacked by* (child is a counterargument that attacks the parent) and *supported by* (child is an evidence/argument supporting the parent). The node type will change accordingly as writers modify the edge type between it and its parent (Figure 5). Writers can edit the contents of the nodes by clicking the *Edit* (✎) button (Figure 1-C).

Integrating text- and visual-based planning. VISAR enables writers to plan their writing through the correspondence between text-based and visual-based planning. Writers can initiate the writing process with text-based approaches such as listing (with bullet points) or freewriting, and integrate their arguments into a visual graph by selecting text and clicking the *Add to graph* button (Figure 3-1). To facilitate the correspondence of components between two representations, when an entity is selected in one representation, VISAR highlights its counterpart in the other presentation. VISAR also renders text blocks in the same color as their corresponding nodes to reinforce the correspondence.

4.3.3 Argumentative sparks. As highlighted in DG3, VISAR aims to provide scaffolding support for users to enhance the persuasiveness of their outlined argumentation. According to argumentation theory [5], effective persuasive writing should comprehensively address counterarguments that could potentially weaken the main argument. Additionally, the writing should avoid logical fallacies and incorporate a diverse range of supporting evidence to boost credibility. Building on these points, we have provided VISAR’s underlying LLM with a set of few-shot examples that enable VISAR

to generate contextually relevant counterarguments, logical fallacies, and supporting evidence related to the current selected draft sentence or bullet point (see details in Section 4.4.3).

Prompts for counterarguments. Writers can explore potential counterarguments for a paragraph or discussion point by clicking on the corresponding text block, selecting *Argumentative sparks*, and then *Counter arguments* on the float menu. VISAR will display a list of counterargument points generated by the LLM. Writers can choose multiple counterarguments to include in their outline. By clicking the *Review and sketch* button (Figure 4-3), they can add the selected counterarguments to the visual outline and review the current outline (Figure 3-4). Finally, when writers clicking the *Generate* button, VISAR will produce a draft that describes the counterargument points in detail and insert it into the draft after the corresponding text block.

Prompts for logical fallacies. Similarly, writers can use VISAR to identify potential logical fallacies of a discussion point in their outline. To do that, they click on the text block, select *Argumentative sparks*, and then *Logical weaknesses* on the float menu (Figure 4-4). VISAR will suggest a list of logical fallacies generated by the LLM that may apply to the selected discussion point. For each suggestion, VISAR also displays an explanation of how this fallacy may apply to the paragraph. Writers can select the fallacies they want to address and, by clicking the *Address weaknesses* button, VISAR will present possible revision options for fixing the selected fallacies. Writers can click the replace (↻) button to update the draft accordingly.

Prompts for supporting evidence. VISAR can suggest supporting evidence for a discussion point by adopting Aristotle’s four persuasive strategies [55]: *ethos* (persuade by sharing professional experience), *pathos* (persuade by arousing audience’s emotions), *logos* (persuade by indicating facts or logical reasoning), and *example* (persuade by presenting concrete example). Details of suggestion generation are discussed in Section 4.4.3. Writers can explore the available types of supporting evidence by clicking on the corresponding text block and then selecting *Argumentative sparks* and *Supporting evidence*. VISAR will present users with potential types of evidence, accompanied by descriptions in the context of the selected discussion point. Writers can add the selected evidence types and review the updated visual outline by clicking the *Review and sketch* button (Figure 4-5). Upon confirming the outline, they can click *Generate* to draft prototype paragraphs that implement the selected evidence types.

4.3.4 Rapid draft prototyping. As emphasized in DG4, VISAR should support users in quickly generating drafts, allowing them to reflect upon, evaluate, and iteratively improve their current plan. VISAR offers flexible draft generation timing, helps maintain consistency among interdependent nodes, and assists in updating the draft.

Supporting different timing of prototyping. Writers may have varying preferences on the timing of draft generation during the planning process. Some writers might prefer to view the draft of an individual component immediately upon its addition to the outline, thereby enabling a direct assessment of its suitability. Conversely, others may prefer to concentrate on constructing the outline and initiate the draft generation process only after completion of the entire outline. To accommodate this, VISAR supports flexible timing

⁶<https://miro.com/>

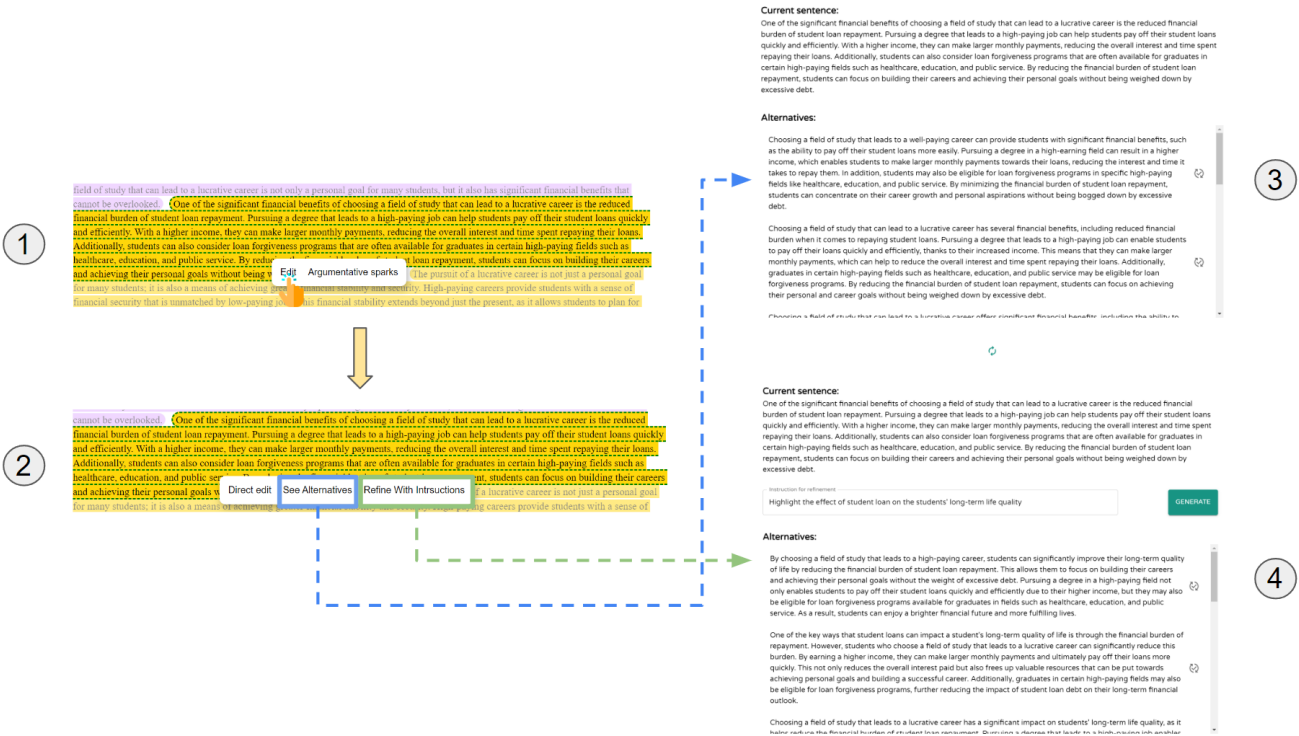


Figure 7: The interaction flow of text editing: (1) Users can edit the draft by clicking the corresponding text block; (2) Users can select different editing strategies; (3) Users can ask VISAR to generate alternative drafts; (4) Users can provide VISAR with flexible natural language instructions revising a specific part of the draft content.

of draft generation. With the *Lazy update* mode off (Figure 1-D), users see the updated draft immediately after modifying the visual outline. When the mode is on, users can edit the outline and click the *Generate text* (✍) button (Figure 1-F) to update drafts for all components at once.

Maintaining consistency among interdependent components. The writing components could be highly interdependent, which complicates the maintenance of consistency. For example, if a writing component has child dependents (e.g., counter argument, supporting evidence), modifying the component could cause inconsistency between its new content and the original children. To address this, VISAR allows writers to *recursively* update dependent children components. In the update window (Figure 6), writers can navigate between affected dependent components using ⬅ and ➡ buttons. Upon moving to a dependent component to be updated, users can select a new topic from a recommended list and generate the updated draft.

Flexible editing strategies on the draft. VISAR allows writers to edit and update drafts using three strategies (Figure 7). They can manually edit a block of text by clicking the *Direct Edit* option. In addition, users can request VISAR to propose alternatives using the *See Alternative* button. Writers can replace the old component with an alternative (Figure 7-3) by clicking the *Replace* (🔄) option. Lastly, the *Refine With Instruction* option allows users to interact

with VISAR in a ChatGPT-like conversational interface (Figure 7-4), where writers can provide flexible instructions to refine the selected component. Writers can engage in multiple iterations of edits to refine the draft to their satisfaction.

4.4 Implementation

4.4.1 Web application. The interactive Web app of VISAR is implemented in React. Specifically, the text editor is constructed with Lexical⁷, which models and organizes the editor’s content into a hierarchical structure comprising nodes. We designed custom node classes for each type of writing component supported by VISAR (i.e., key aspects, discussion points, counterarguments, and evidence). The back-end server uses the Flask framework⁸ for communication with the LLM and accessing a MongoDB⁹ database.

4.4.2 Writing goal recommendation. VISAR used OpenAI’s API (GPT-3.5-turbo¹⁰) to recommend writing goals. The detailed prompt templates are shown in the Appendix. Specifically, the model is first prompted to play the role of “a helpful writing assistant that aims to help writers come up with high level aspects or topics that they can think of to support their argument”. As users select arguments and click the *Elaborate* button to obtain inspiration for key aspects,

⁷<https://lexical.dev/>

⁸<https://flask.palletsprojects.com/en/2.2.x/>

⁹<https://www.mongodb.com/>

¹⁰<https://platform.openai.com/docs/models/gpt-3-5>

we prompt the model with *"Please list key aspects that are worth discussing to support the argument: [selected argument]."* We also provide the model with several examples in the prompt, which inform it of the desired output format (i.e., a numbered list).

Similarly, to elicit *discussion points* regarding a key aspect of an argument, we direct the model to adopt the role of *"a helpful writing assistant that aims to come up with pertinent discussion points based on a specified aspect to reinforce the given argument"*. We supply the model with several examples and prompt: *"List key discussion points worth including in the discussion to support argument [selected argument] from the aspect of [selected aspect]"*. This prompt is executed for each selected aspect. In practice, we discovered that employing a "chain of thought" strategy (initially considering aspects and subsequently generating discussion points based on the selected aspects) allows the model to produce more detailed and relevant discussion points. This observation confirms the findings of previous studies [61, 76] that prompting LLMs to undertake explicit intermediate steps enhances their problem-solving and reasoning capabilities.

4.4.3 Argumentative spark generation. VISAR leverages the GPT 3.5 model to create argumentative sparks¹¹. Instead of relying solely on the model's zero-shot capabilities, we supply it with several examples of each type of argumentative spark (counterargument, logical fallacy, and supporting evidence) in the prompts. These examples are based on proven methods for identifying counterarguments or taxonomies of logical fallacies and supporting evidence from argumentation theory, ensuring that the generated sparks are consistent with established principles of argumentation. We found that, when we provide the model with a few examples, it can proficiently consider a wide range of categories, even if the model is not explicitly shown all of these categories. The detailed example and prompts can be found in the Appendix.

Counterargument generation. Prior research in argumentation theory [5, 70] suggests various strategies for opposing an argument, such as attacking the premise, warrant and backing component of the argument, or raising an alternative explanation for a phenomenon. We familiarize the model with the strategies summarized by Blair [5] by providing it with several examples mentioned in the Blair's work. Specifically, we inform the model about the applicable counterargument types and their application to particular arguments in the examples. When writers click the *Counterargument* button on a selected argument, we prompt the model to act as a *"helpful writing assistant specializing in argumentative essay tutoring and generating counterarguments for the given statement"*, then enter the following prompt: *"Here are several examples of how to raise counterarguments for an argument: [examples]. Please list potential counterarguments that can challenge the argument [selected argument]"*.

Logical fallacy generation. To identify logical fallacies in a discussion point, we use the taxonomy of logical fallacies summarized by argumentation theories [5, 16] and guide the model to identify which fallacy types may apply. During inference, we first prompt the model to play a role as *"a helpful writing assistant focusing on argumentative essay tutoring and trying to suggest logical weaknesses*

in the given statement", and then input the prompt: *"Here are several examples of demonstrating how to find logical weaknesses for an argumentation: [examples]. Please list potential logical weaknesses in the argument [selected argument]"*.

Supporting evidence generation. We introduce the model to Aristotle's four persuasive strategies [55] (*ethos, pathos logos, and example*) for generating supporting evidence for a selected argument. Specifically, we first ask the model to play a role as *"a helpful writing assistant focusing on argumentative essay tutoring and trying to suggest supporting evidence types in the given statement"*. During inference, we prompt the model: *"Please list potential supporting evidences that can back the argument: [selected argument]. You can think from the following aspects: sharing professional experience (ethos), arousing audience's emotion (pathos), providing facts and strict logical reasoning (logos) and presenting concrete practical examples (example)"*.

4.4.4 Draft generation. We utilize the model to generate prototype drafts that implement writers' argumentation plans. The prompt templates for creating each specific writing component, as well as for revision and refinement, are detailed in the Appendix (Table 2). When a writing component has a parent node in the plan, we include the parent's content as context to improve coherence between the two elements.

5 USER STUDY

We conducted a lab user study¹² with 12 participants to evaluate the usability, effectiveness, and usefulness of VISAR. The study aims to answer the following questions:

- **RQ1:** Can users successfully plan argumentative writing using VISAR?
- **RQ2:** How useful is VISAR in facilitating the ideation, organization, and revision of users' writing planning?
- **RQ3:** What challenges do users encounter when using VISAR to plan their writing?

5.1 Participants

We recruited 12 participants from a private R1 university in the United States. The participant group comprised 1 sophomore, 3 juniors, 4 seniors, and 4 graduate students. 4 participants identified themselves as intermediate writers (having some experience but still honing their skills), 7 as advanced writers (with significant experience and confidence in their abilities), and 1 as an expert writer (having extensive experience and a high level of skill). We anticipated that most, if not all, of our participants would be proficient writers. It is important to note that the average incoming freshman at this institution scores in the 96th percentile in the Evidence-Based Reading and Writing (EBRW) section of the SAT test, and all students complete two semesters of Writing & Rhetoric in their first year. All participants were proficient in English, and 8 were native English speakers. Each participant received a \$30 USD compensation for their time.

Regarding their experience with AI-supported writing tools, 8 participants had used AI tools for grammar and spelling assistance, 7 for writing prompt or idea suggestions, 5 for refining writing

¹¹See prompt templates in Appendix

¹²The study protocol was reviewed and approved by the IRB at our institution.

styles or tones, 6 for polishing content or language, and 1 for plagiarism detection. In terms of their familiarity with writing planning methods, 2 were familiar with mind maps and 11 were familiar with outline sketches. All 12 participants had previously used generative AI tools (such as ChatGPT) to support their work in various scenarios, including brainstorming outline and ideas, refining the language, summarizing text, and generating code snippets.

When asked about the challenges they commonly face in planning argumentative writing, 7 participants reported having previously experienced difficulties in generating ideas, 9 in organizing ideas and information logically, 3 in developing counterarguments, and 4 in making their arguments more persuasive.

5.2 Study Design

Each study session, lasting approximately 90 minutes, was conducted in-person at a usability lab. The study used a within-subjects design, where the participants experienced three conditions during three 25-minute argumentative writing planning sessions. The study procedure is detailed in Section 5.2.3.

5.2.1 Tasks. To design a representative domain-general task for argumentative writing, we adopted the Issue Essay task from the Graduate Record Examinations (GRE)¹³ for our study. This task necessitates that writers create an argumentative essay exhibiting critical thinking while clearly expressing their thoughts on a given topic that can be approached from various perspectives and applied to multiple situations and conditions. In our study, we selected three topics from the GRE’s issue writing sample pool¹⁴. Our selection criteria stipulated that the topic should not require expert knowledge and must have an appropriate scope, allowing writers to generate specific points from different aspects within a limited time frame. The selected topics and the corresponding task descriptions are provided in Appendix. Users encountered a different topic in each writing session, during which they were asked to use the system corresponding to the current condition to create an outline for the argumentative essay and draft as much of the essay as possible. The order in which the participants addressed the three writing topics was randomized.

5.2.2 Conditions. The study incorporated the following three conditions:

- **Baseline:** A plain text editor without AI support.
- **GPT Playground:** A text editor resembling OpenAI Playground¹⁵ that allowed users to input prompts and select text for the LLM (GPT-3.5-turbo was used, the same as VISAR) to generate a response. The model’s output was appended to the participants’ prompt in place.
- **VISAR:** A full version of VISAR featuring visual programming, step-by-step ideation support, argumentative sparks, and rapid draft prototyping capabilities.

We used a Latin square experimental design [60] to ensure a balanced sequence of task topics and conditions for each study session.

¹³<https://www.ets.org/gre/test-takers/general-test/prepare/content/analytical-writing/issue.html>

¹⁴<http://words.greemat.com/greessay.html>

¹⁵<https://platform.openai.com/playground>

5.2.3 Study procedure. At the beginning of each study session, the experimenters collected informed consent and demographic information from the participants. The study coordinators then provided a high-level introduction of the writing tasks (as described in Section 5.2.1). Each participant completed three writing sessions, each session comprising a 5-7 minute tutorial and a 20-minute writing planning task. The tutorial explained the key features of the tool associated with the current condition. The task sessions were independent of each other, which means that the output created in each condition was not carried over to subsequent conditions. After the participants completed the three writing sessions, they completed a post-study questionnaire, which mainly assessed usability, usefulness, and enjoyment of VISAR system. The study ended with a 10-minute semi-structured interview, during which the experimenter asked questions to understand the rationale behind the participants’ observed behaviors, tool usage, and reflections on their experiences with the tools. The purpose of the semi-structured interview was to explore participants’ overall experiences, and to elicit their specific feedback on VISAR features, final draft quality, improvement suggestions, and potential benefits from the system.

5.3 Result

5.3.1 Post-study Questionnaire. Figure 8 summarizes the results of the post-study questionnaire of 12 participants. In general, the participants found that VISAR was more effective in helping them generate argumentative elements, organize these elements into outlines, and validate the outlines compared to the GPT Playground and baseline conditions. The participants also expressed confidence in the quality of outlines produced using VISAR and considered the system easy to learn and enjoyable to use.

Although VISAR did not score as high as the other two conditions on “easy to use” and “learn to use quickly” due to the many new features added, our participants still considered it generally easy to use.

5.3.2 Post-study interview. Following established open-coding protocols [8, 38], two study team members first performed a round of independent qualitative coding of interview transcripts. Afterward, they had a discussion to achieve agreement and created a consolidated codebook. Using this codebook, a thematic analysis was performed to uncover emerging topics from the interview. The entire research team collectively reviewed the coding outcomes to identify high-level themes. We report the following key findings:

KF1: Assistance for hierarchical ideation aids users in developing structured outlines. The participants found the hierarchical goal recommendations provided by VISAR useful to help them explore and develop structured outlines. The participants found that VISAR could “give us (them) more points than we (they) would have thought about” (P3), allowing them to “realize the missing thinking aspects and efficiently explore extensive possible directions by a few clicking” (P9). In addition, some participants found that the ideation assistance process of VISAR aligned well with their natural thinking process. For example, P7 said “it was kind like what I did in my own writing, creating opening sentence and then thinking of concrete aspects and examples to support my argument step by step. It was like compile whatever the thing was and gave me some options at each

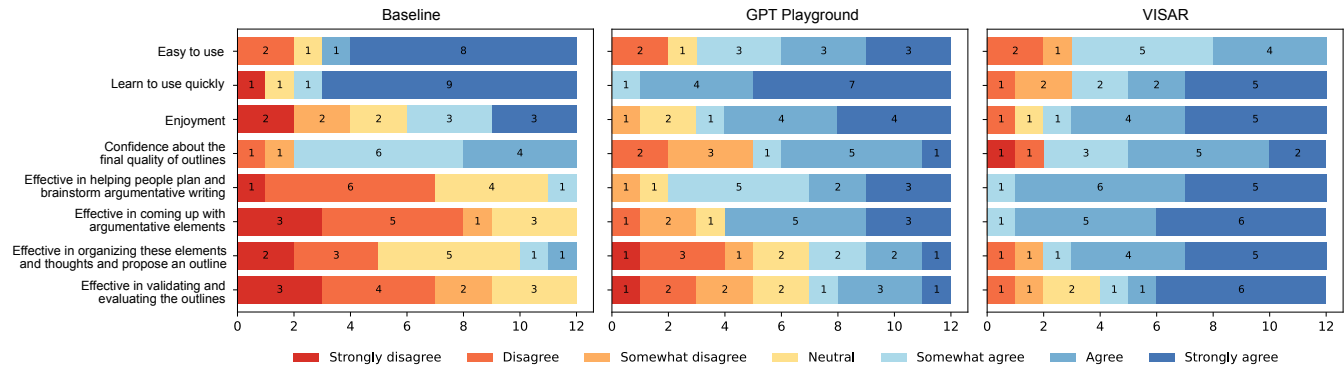


Figure 8: Results from the post-study questionnaire comparing user ratings of three conditions

thinking step, except through guided by machines”. Confronted with this familiar AI-guided ideation process, the participants could “reflect and reassure their own thinking” (P2), “articulate their thoughts” (P8) and “better understand their own arguments” (P8).

However, some participants worried that this ideation assistance could “deprive our (their) control of the writing directions” (P7) and make them “less creative and initiative in thinking about the writing” (P9). Besides, this hierarchical assistance is limited to help on the non-linear ideation part of the planning. For example, P4 found that “this assistance did not help me (her) understand how I (she) could link or compare two aspects included in my (her) outline”.

KF2: The synchronized text and visual views provide complementary assistance for writing planning. All participants reported that text and visual editors helped their writing planning in different ways, and the color coding of text and node helped them easily correspond the two modalities. The visual editor helped them understand the clear logical structure of their outline and draft, as P12 said, “it is pretty useful because when we (they) want to write like a good argumentative essay, we (they) wanted to be logically clear”. Also, the visual editor eased their work to edit and organize their outline. For example, P11 said “we may keep adjusting the outline like making your argument followed by an example or led by another counter argument, without this tool (visual editor), it will be pretty hard for us to logically put plenty of arguments together”. In addition, visual editor could also help them quickly “navigate to a concrete part of outline” (P3) and “find the missing part” (P7). The participants also found the visual editor useful for supporting their “two-stage cognitive process including planning and reviewing” (P1), where they first “draft and think how things can go hierarchically and then see whether I (they) can add more nodes or move a node that should be fit better under another category” (P4).

On the other hand, the participants found the text editor to be helpful and easy to use for reviewing and modifying the concrete draft of a discussion point after deciding on its position in the outline. For example, P2 said “I feel like this argument should go off this branch and then it would just like make sense to do it from the tree part versus doing it from the editor and then I went on the (text) editor once I like made that first initial branch and saw okay this is where it’s gone and then I was able to edit from there”. Compared to the visual editor, the text editor provided the participants with

specific contexts so that they can better “validate the cohesiveness of their arguments” (P5). Besides, some participants preferred to use the text editor for outline sketching because “it aligns with what I (they) normally did as drafting an essay” (P7).

KF3: Argumentative sparks help users strengthen their writing. The participants found that the argumentative spark features help them enhance the persuasiveness of their writing. Specifically, these argumentative sparks help them identify opportunities to improve a particular argument. For example, P6 said “It was really cool to see like you can generate counter arguments for this section, because I feel like I wasn’t focusing enough on counter arguments. So that is helpful to look at like potential logical fallacies”. Besides, the participants reported that the recommendations of counterarguments and evidences were helpful for them to improve the objectivity of their draft and articulate their thoughts. For example, P8 said “I feel like it’s very helpful, especially in like an argumentative writing setting, because it remains relatively objective when it talks about things. Yeah, especially useful when you have an idea in mind that you don’t really know how to articulate it”.

However, the participants also found that the argumentative sparks were sometimes not sufficiently concrete. For example, P3 said “I think the only issue I I was coming up, I asked for supporting evidence for one of my points. And instead of giving me like supporting evidence, the only thing that the the bubble button said was one possible evidence showing that there is scientific research to prove that this is right. So it didn’t really tell me what the scientific research was if that makes sense. So it didn’t give me the point, it just won’t be worth a point.”.

KF4: Draft prototyping helps users improve plans but it cannot be used directly in the final product. Most of the participants thought that the rapid prototype feature could help them understand and improve their planning within a concrete context, although the generated draft tended to suffer from issues such as robotic tones, repeating content, poor transition, and monotonous language style across different generations. An important benefit mentioned by many participants was that the draft prototyping allowed them to preview what the implementations of their plans could look like as they developed the plan. This helped the participants go back to the planning stage to expand their thought and come up

with more specific points, which was otherwise difficult to achieve at the high-level planning stage. For example, P5 said “I initially started with like really basic things. Then I expanded them a little bit more when I saw those generations, because it reminded me of some more specific responses and more detailed evidences that I didn’t think of when drafting the outline. I think it was pretty key, pretty important”. Besides, the participants thought that this feature could dramatically save their effort and time in creating a prototype draft to validate their ideas and support further revision. For example, P11 said “Overall, while I may be picky about how things are phrased, I like being able to directly edit on the rough draft like it, based on how I think ultimately is best. That may be more helpful and saves time to write like three sentences versus doing it on my own. It takes me to think directly about how to phrase those three sentences and improve my arguments”. The automatic prototyping also acts as a “milestone” that provided the participants with a sense of achievement, as P6 commented, “It kind of made you feel like you made a lot of progress. You’re like, I only have like four nodes, but here’s a whole starting essay”.

Challenges, suggestions and additional application scenarios. The participants faced usability challenges related to interaction and display. For instance, P5 and P8 reported difficulties connecting two nodes in the visual editor due to the small size of the connection handles. Additionally, the initial position of the AI-generated nodes and the space between them were sometimes problematic (P3, P5) due to an implementation bug, particularly when participants incorporated a large number of AI-recommended options into their plan simultaneously. Moreover, the participants observed that the order of the paragraphs in the draft sometime fails to update when they adjusted the order of the corresponding visual nodes (P5, P10).

In addition to interaction challenges, the participants noted that VISAR had a steeper learning curve compared to the systems under the other two conditions because they had to learn how to use the new features. In some cases, the use of VISAR may also increase their cognitive load, as they had to “read the long generated draft to make sure it aligns with my (their) intent” (P3).

The participants offered valuable recommendations to improve and expand the current system. P9 proposed adding a feature to generate visual mind maps from users’ existing drafts, assisting users in analyzing their draft’s logical structure. P1 suggested extending the use of argumentative sparks to entire outlines (in addition to individual discussion points) to improve overall persuasiveness. P3 recommended that VISAR include citations for the suggested supporting evidence, which would require technological advances in LLMs. Lastly, P6 believed that VISAR should offer more intuitive controls, such as allowing users to drag paragraphs to reorder them and update the visual mind map accordingly.

Regarding potential application scenarios for VISAR beyond argumentative writing planning, many participants suggested that VISAR could assist with brainstorming and organizing ideas for lighter writing tasks that require quick and logical drafts. Additionally, P10 mentioned that VISAR could be used during the revision stage to facilitate critical evaluations of drafts by providing argumentative sparks. However, some participants expressed caution about using VISAR’s output directly in writing for more serious purposes, such as semester papers (P1) or personal statements (P9),

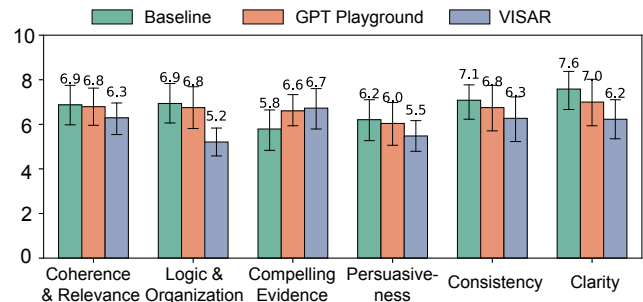


Figure 9: Average expert scores on each evaluative dimension across varying conditions.

due to ethical concerns about plagiarism and authorship (P6), and doubts about VISAR’s ability to accurately represent their personal experiences and knowledge (P8).

5.3.3 External Evaluation of Participants’ Drafts. While our main focus was on the prewriting and planning phases of argumentative writing, we also investigated the quality and characteristics of the drafts generated by our participants. It’s important to note that participants were not required to complete a full draft on their assigned topic in each condition; their primary task was to form a robust outline (as discussed in Section 5.2.1). Yet, evaluating and comparing the quality of these provisional drafts across different conditions can offer insights into the advantages and potential problems of transitioning from plans to drafts with each interface.

We have two experts in argumentative writing evaluate the quality of each draft collected in our studies. Both experts were Writing & Rhetoric instructors at a private U.S. university with extensive professional experience in practicing, teaching, and assessing argumentative writing.

The expert evaluators were given a set of assessment criteria adapted from the official GRE website¹⁶, Re³ [83] and Adaptive Learning system [73], where they rated each submission on the following six dimensions on a scale of 1 to 10. They also provided short feedback on the strengths and weaknesses of each outline. During the evaluation process, the evaluators were unaware of the conditions under which the drafts were produced. The evaluative dimensions were shown below:

- (1) Coherence and Relevance: identify aspects of the argument that are pertinent to the assigned task and examine them with insightful analysis.
- (2) Logical organization: develop ideas and arrange them in a logical sequence that supports the overall argument.
- (3) Compelling and thorough evidence: present strong, comprehensive evidence that substantiates the arguments made.
- (4) Persuasiveness: evaluate the effectiveness of the argument in persuading the target audience.
- (5) Consistency: ensure that the argument is factually consistent and free of contradictory details.

¹⁶<https://www.ets.org/gre/test-takers/general-test/prepare/content/analytical-writing/scoring.html#accordion-06cf390d3c-item-60c4f150ea>

- (6) Clarity: assess whether the argument is easily understandable and presented with clear language and structure.

5.3.4 Quantitative Results. A summary of the ratings on each evaluative dimension across conditions is shown in Figure 9. The scores indicated that though drafts from VISAR condition have passing quality, they still fall short from excellence.

In particular, a one-way repeated measures ANOVA was conducted to investigate the influence of three conditions on the six measures. Our findings revealed a significant main effect on the score of Logic & Organization from the use of VISAR ($F_{2,46} = 6.36, p < .01$). Post-hoc pairwise t-tests revealed that the score on Logic & Organization of VISAR ($M = 5.2$) is significantly lower than both the baseline ($M = 6.9, p = .02$) and GPT playground ($M = 6.8, p < .01$). The difference between the baseline and GPT playground was not found to be statistically significant. We found no significant difference among ratings on all other dimensions among the conditions.

5.3.5 Qualitative Results. To understand the characteristics of drafts produced in each condition, two researchers conducted qualitative coding [50] on evaluators’ open-ended comments on the drafts. Specifically, each researcher independently annotated expert comments using either *vivo* codes or self-defined codes. Then they sat together to discuss, came to a consensus on changes to the codebook, and relabeled comments independently using the newly formed unified code book. The Cohen’s kappa score between the coding results from the two researchers in the second round was 0.66, indicating substantial agreement [36].

In terms of strengths and weaknesses, evaluators found that user-written drafts from baseline condition tends to have clear and cogent main arguments. They usually contained logical and coherent argumentation structures and “*there is a clear clue how the draft was expanded from and follows the outline*”. However, those drafts tends to lack details such as specific examples, supporting evidences or counterarguments, which could cause “*the argumentation stays on a surface level*”.

In the case of the GPT Playground condition, evaluators observed that these drafts were filled with a large number of details that gave the arguments an appearance of solidity. However, they also noted that such drafts could be poorly structured, having an unclear main argument (or even contradictory arguments) and an unsatisfactory flow of reasoning. In addition, some examples and evidence provided in the drafts did not adequately support the immediate arguments, causing the argumentation to veer off topic and become disjointed in certain instances. This could be partly attributed to participants frequently invoking “auto completion” to understand how an outline item could be expanded. Due to time constraints, they often exerted minimal effort to modify the generated content, reorganize the flow, or remove unsuitable completions.

Finally, evaluators found that the drafts from the VISAR condition were abundant in specific content drawn from a variety of perspectives and at different depth levels. However, they noted these drafts often suffered from issues like repetitive or self-conflicting

content, and unclear transitions and relationships between paragraphs and sentences. For example, evaluator reported issues like “*The product felt like a series of mini-essays collated together without any transitions or framing. This makes it difficult to remember the writer’s main argument and affects the essay’s overall effectiveness*”. Nonetheless, evaluators thought that some of these drafts “*have great potential of becoming a compelling essay if authors organize the content in a better way*”.

These qualitative findings suggest that while a generative AI model can aid in creating draft and enable writers to anticipate how the outlines may unfold, humans still need to supervise the progression from plan to draft. It is essential to ensure that transitions and logical relationships between writing elements are valid and clear. A major limitation of the current VISAR system is that it does not sufficiently consider the high-level context of the article when generating content. Future AI-augmented writing systems could incorporate external memory repositories¹⁷ (e.g. vector database) or more effective prompting strategies (e.g. self-reflection [63] or tree-of-thoughts [84]) to enhance the coherence and relevance among the generated content.

6 DISCUSSION

6.1 Interaction Modalities and Strategies for Collaborating with LLMs

The “default” modality for interacting with LLMs is through natural language conversations (e.g., ChatGPT). However, this mode of interaction poses challenges to human-AI collaboration and limits the potential applications of LLMs. From the perspective of the AI model, models encounter difficulties in accessing the specific task context and user behavior history, which are often disconnected from the conversational context, making it difficult to understand the implicit task specifications and user intent [7, 41, 42]. This can result in inaccurate or irrelevant responses to user requests.

On the user side, unlike using cognitively friendly tools such as visual diagramming, users may have trouble expressing their intent or needs in textual form [78] because the interaction is “too open” without many indicators for the possible actions at the current state, the system’s capabilities, and the most useful types of input to provide for the system. Furthermore, users must transition to a conversational environment to interact with LLMs, potentially imposing additional cognitive load. These challenges can be particularly significant in complex real-life scenarios where tasks may have multiple sub-parts and users utilize different tools or modules to address them.

VISAR addresses these challenges by simultaneously enabling direct manipulation and conversational agent interfaces across different modalities, allowing users to interact with AI models in a modality that best supports their current task in a unified workspace [64]. For instance, direct manipulation, as an interaction modality, offers advantages such as intuitiveness, implicit representation of intent, awareness of the interaction context, and low cognitive load [27]. Therefore, we employ it as an interaction strategy for tasks where users might otherwise encounter difficulties in articulating their needs and corresponding context (e.g., argumentative sparks and

¹⁷For example, we can incorporate external memory using LLM application framework like LangChain: <https://python.langchain.com/docs/modules/memory/>

hierarchical goal recommendation). Moreover, we use direct manipulation to facilitate user interaction with AI models in visual outlines, following the implications and recommendations of previous work [81, 86]. Conversely, we employ conversational agent interfaces to enable users to explicitly convey and incrementally refine their requirements for draft generation (e.g., refining with instruction, addressing logical fallacies), where users need more flexibility and expressiveness that are difficult to provide in a direct manipulation environment. But in such scenarios, the availability of concrete drafts allow users to refer to them in their natural language instructions, alleviating the challenges in the lack of context when the user interacts with a conversational interface in earlier stages of the process.

6.2 Generative AI for Rapid Prototyping in Creative Tasks

Creative tasks often involve a non-linear process for generating new ideas and transforming them into tangible results. Based on the cognitive theory of creativity, the process includes key stages such as preparation, illumination, evaluation, and implementation [34]. Individuals may cycle through these stages repeatedly, revisiting earlier phases to refine and iterate their ideas. Rapidly drafting prototypes throughout this process can be beneficial, as it allows individuals to reflect on their concepts, gain a clearer understanding of their alignment with the ultimate goal, and draw inspiration from the surrounding context. However, prototyping can be time consuming and labor intensive, as discussed earlier regarding argumentative writing.

Generative AI models have the potential to enhance the creative process with their ability to rapidly generate parallel prototypes. For example, VISAR shows the feasibility and value of quickly producing drafts based on argumentative outlines. These drafts offer a tangible intermediate representation for both humans and models to understand the writing context, enabling models to assist individuals in evaluating their existing ideas, clarifying ambiguities in human-AI communication, and identifying potential gaps in the working scenario. The writers can create and organize different parallel argumentation plans using the visual outline feature, generate concrete drafts that implement these outlines with little effort, and compare the drafts in order to understand the strengths and weaknesses of different argumentation plans and iteratively improve them.

However, the use of generative AI during the prototyping stage of creative projects may present challenges. One concern is that when humans do not create the prototypes themselves, they may need to invest extra effort to review and understand the generated content. As P3 noted, they *had to read the lengthy draft to ensure it aligned with their original intent*. Additionally, machine-generated prototypes could unintentionally introduce bias or constrain users' "blue sky" creative thinking, highlighting research opportunities to better enable users to explore a wide range of potential ideas and integrate their own insights into the drafting process.

6.3 Visual Programming for Bridging User Workflow and Model Thought Process

Visual programming interfaces have proven effective in facilitating user workflow across various application domains (e.g. collaborative work [69, 74], task planning [25], interactive learning [9]), helping to clarify relationships and dependencies between concepts and enabling users to understand and navigate complex information [52]. Visual representations also allow users to see the big picture, identify patterns, connections, and gaps in their ideas, which enhance brainstorming, problem-solving, and decision-making.

VISAR demonstrates the promise of integrating user workflow with LLM's thought process features within a visual programming interface. Users can create organized writing outlines by defining goals and their connections at different abstraction levels using visual representations of arguments, discussion points, and evidence, as well as the relationships among them. Simultaneously, by addressing and implementing these goals in a top-down sequence, the model can follow each logical step proposed by the user, progressively building an understanding of the current writing context. This enables the model to generate content that aligns with the abstraction levels of the user's thought process.

6.4 Ethical Considerations for AI Writing Assistance

While generative AI models have been found useful many subtasks of the writing process, designers of human-AI collaborative writing tools must carefully consider ethical concerns. First, models could introduce bias, stereotypes, and misinformation [7, 22], resulting in factually flawed or prejudiced outcomes. Furthermore, models may unintentionally generate content that closely resembles or reproduces existing sources from their training data, raising concerns about plagiarism or copyright infringement, even if the user is not intentionally copying from existing works. Additionally, directly incorporating AI-generated content into one's writing could call into question the writer's intellectual integrity and authorship of the written product.

VISAR addresses these potential ethical concerns by primarily using generative AI models to assist writers during the prewriting and planning stages, generating early prototypes only for sensemaking and comprehension purposes. The design of VISAR emphasizes user control and autonomy, ensuring that the human writer takes the lead in key creative aspects of the writing process. To better mitigate the biases and stereotypes in generated contents, future work can be done to combine LLM's capability of self-critique and reflection [12, 63] with external strategies such as augmenting LLM prompts with relevant factual knowledge from credible sources like Wikipedia, red flagging potential bias using independent bias detection model [56], and pre-emptively informing users about potential biases in model-generated contents.

To further address the ethical concerns associated with generative AI models, it is essential to establish guidelines and best practices for their use in creative work. This may include transparently revealing the use of AI-generated content, encouraging critical assessment of AI-generated suggestions, and implementing strategies to identify and mitigate biases in output. The new human-AI collaborative writing paradigm introduced by VISAR opens up

new challenges and opportunities regarding these aspects, which we would love to further investigate with the rest of the research community as future work.

7 LIMITATION & FUTURE WORK

The current version of VISAR has several technical constraints. First, the fixed prompt templates used for drafting, recommendation, and argumentative sparks generation could be sub-optimal, potentially yielding similar model responses despite distinct input contexts (e.g., analogous language styles across various generations). Second, as P7 pointed out, the system currently does not account for transitions between generated discussion points in prototype drafts, which may result in reduced cohesiveness in full drafts of prototypes. Third, the system supports only a limited range of argument and relation types, lacking support for certain specific argument components, such as warrants and qualifiers, and specific styles of argumentation, such as bottom-up deductive argumentation [26]. Lastly, our system can only support top-down tree-structured argumentation so far, however, argumentation could have more complex structures such as directed graph.

In addition to VISAR’s implementation limitations, the GPT 3.5 model used in VISAR may also generate factual inaccuracies, inconsistent responses, or contradictory statements in the prototype drafts [7]. This may expose users to confusing and misleading information, and may have adverse second-order effects on how users conceive of research and the practices and methods involved in finding data and evidence. In the design of VISAR, we emphasize that the generated draft is only a “prototype” for facilitating iterative prewriting and planning processes, mitigating the adverse impact of this problem.

Regarding the study design, there are several potential threats to the validity of our findings. Our tasks used argumentative writing topics from the GRE test. Although these tasks are considered adequate representations of general argumentative writing tasks without requiring domain-specific knowledge, it would be valuable to further explore VISAR’s use in more diverse real-world writing tasks through a deployment study. Expanding the demographics of the participants in future deployment studies will also help assess the ecological validity of VISAR.

It is also worth mentioning that the current VISAR system is not specifically designed for any particular user group. As a result, the argumentative components it supports (such as node and relation types) are relatively generic. In the future, we plan to explore the potential application of VISAR in educational settings. For example, we would like to explore how VISAR can help K-12 students to learn argumentative writing. We will also investigate the tool’s efficacy in providing personalized guidance, enhancing the structure and coherence of arguments, and fostering critical thinking skills. In addition, we will employ argument mining based models [37, 71] to automatically validate the logical soundness of arguments generated by LLMs and notify learners about any potential issues in those outputs. Lastly, we will explore how to accommodate additional argumentation structures other than tree-based one in the future version of VISAR.

Moreover, we plan to extend VISAR to accommodate collaborative writing planning [74, 75, 85]. Our research will focus on enhancing the system’s functionality to support seamless collaboration among multiple users. We will develop shared visual programming spaces that allow users to jointly construct and refine their writing plans. Additionally, we aim to investigate the impact of these collaborative features on group argumentative writing quality and individual engagement. By incorporating effective collaborative elements into VISAR, we seek to create a more engaging and interactive platform that promotes peer learning, diverse perspectives, and the co-creation of well-reasoned arguments.

8 CONCLUSION

This paper presented VISAR, an AI-enabled tool designed to support the prewriting and planning stages of argumentative writing. VISAR provides: (1) hierarchical writing goal recommendations to facilitate writers’ ideation processes, (2) synchronized text and visual editors for easy planning and organization of outlines, (3) argumentative sparks that supply writers with supporting evidence, logical fallacies, and counterarguments, and (4) rapid draft prototyping, enabling writers to understand and improve their outlines by reflecting on draft implementations. A user study with 12 participants demonstrated that writers can effectively use VISAR to create argumentative writing plans, and the system was perceived as useful, usable, and likable.

ACKNOWLEDGMENTS

This work was supported in part by an AnalytiXIN Faculty Fellowship, an NVIDIA Academic Hardware Grant, a Google Cloud Research Credit Award, a Google Research Scholar Award, and an Asia Research Collaboration Grant from Notre Dame International. Any opinions, findings or recommendations expressed here are those of the authors and do not necessarily reflect views of the sponsors. We would like to thank John Behrens, Sherry Tongshuang Wu, Hua Shen, and Ting-Hao (Kenneth) Huang for useful discussions.

REFERENCES

- [1] Nader Akoury, Shufan Wang, Josh Whiting, Stephen Hood, Nanyun Peng, and Mohit Iyyer. 2020. Storium: A dataset and evaluation platform for machine-in-the-loop story generation. *arXiv preprint arXiv:2010.01717* (2020).
- [2] Abdelbaset Almasri, Adel Ahmed, Naser Almasri, Yousef S Abu Sultan, Ahmed Y Mahmoud, Ihab S Zaqout, Alaa N Akkila, and Samy S Abu-Naser. 2019. Intelligent tutoring systems survey for the period 2000-2018. (2019).
- [3] Rosa Maria Ballardini, Kan He, and Teemu Roos. 2019. AI-generated content: authorship and inventorship in the age of artificial intelligence. *Online Distribution of Content in the EU* (2019), 117–135.
- [4] Advait Bhat, Saaket Agashe, Parth Oberoi, Niharika Mohile, Ravi Jangir, and Anirudha Joshi. 2023. Interacting with Next-Phrase Suggestions: How Suggestion Systems Aid and Influence the Cognitive Processes of Writing. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) (IUI '23). Association for Computing Machinery, New York, NY, USA, 436–452. <https://doi.org/10.1145/3581641.3584060>
- [5] J Anthony Blair. 2011. *Groundwork in the theory of argumentation: Selected papers of J. Anthony Blair*. Vol. 21. Springer Science & Business Media.
- [6] Wayne C Booth, Gregory G Colomb, and Joseph M Williams. 2003. *The craft of research*. University of Chicago press.
- [7] Ali Borji. 2023. A categorical archive of chatgpt failures. *arXiv preprint arXiv:2302.03494* (2023).
- [8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [9] Brian Broll, Akos Lédeczi, Peter Volgyesi, Janos Sallai, Miklos Maroti, Alexia Carrillo, Stephanie L Weeden-Wright, Chris Vanags, Joshua D Swartz, and Melvin Lu.

2017. A visual programming environment for learning distributed programming. In *Proceedings of the 2017 ACM SIGCSE technical symposium on computer science education*. 81–86.
- [10] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- [11] Tuhin Chakrabarty, Vishakh Padmakumar, and He He. 2022. Help me write a poem: Instruction Tuning as a Vehicle for Collaborative Poetry Writing. *arXiv preprint arXiv:2210.13669* (2022).
- [12] Xinyun Chen, Maxwell Lin, Nathanael Schärli, and Denny Zhou. 2023. Teaching large language models to self-debug. *arXiv preprint arXiv:2304.05128* (2023).
- [13] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: sketching stories with generative pretrained language models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [14] Andy Coenen, Luke Davis, Daphne Ippolito, Emily Reif, and Ann Yuan. 2021. Wordcraft: a human-ai collaborative editor for story writing. *arXiv preprint arXiv:2107.07430* (2021).
- [15] Hai Dang, Karim Benharrak, Florian Lehmann, and Daniel Buschek. 2022. Beyond Text Generation: Supporting Writers with Continuous Automatic Text Summaries. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–13.
- [16] Frans H van Eemeren, Rob Grootendorst, and Tjark Krugier. 1987. *Handbook of argumentation theory: A critical survey of classical backgrounds and modern studies*. De Gruyter.
- [17] Ralph P Ferretti and Steve Graham. 2019. Argumentative writing: Theory, assessment, and instruction. *Reading and Writing* 32 (2019), 1345–1357.
- [18] Ralph P Ferretti and William E Lewis. 2018. Argumentative writing. *Best practices in writing instruction* 135 (2018).
- [19] Linda Flower and John R Hayes. 1981. A cognitive process theory of writing. *College composition and communication* 32, 4 (1981), 365–387.
- [20] John T. Gage. 2006. *The shape of reason: Argumentative writing in college*. Pearson Education.
- [21] Jie Gao, Yuchen Guo, Gionnieve Lim, Tianqin Zhang, Zheng Zhang, Toby Jia-Jun Li, and Simon Tangi Perrault. 2023. CollabCoder: A GPT-Powered Workflow for Collaborative Qualitative Analysis. *arXiv:2304.07366 [cs.HC]*
- [22] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for science writing using language models. In *Designing Interactive Systems Conference*. 1002–1019.
- [23] Allan Gollins and Dedre Gentner. 2016. A framework for a cognitive theory of writing. In *Cognitive processes in writing*. Routledge, 51–72.
- [24] William Grabe and Robert B Kaplan. 2014. *Theory and practice of writing: An applied linguistic perspective*. Routledge.
- [25] Mary M Hayhoe, Anurag Shrivastava, Ryan Mruzcek, and Jeff B Pelz. 2003. Visual memory and motor planning in a natural task. *Journal of vision* 3, 1 (2003), 6–6.
- [26] David Hitchcock and Bart Verheij. 2006. *Arguing on the Toulmin model*. Vol. 10. Springer.
- [27] Edwin L Hutchins, James D Hollan, and Donald A Norman. 1985. Direct manipulation interfaces. *Human-computer interaction* 1, 4 (1985), 311–338.
- [28] Daphne Ippolito, Ann Yuan, Andy Coenen, and Sehmon Burnam. 2022. Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers. *arXiv preprint arXiv:2211.05030* (2022).
- [29] Ralph H Johnson. 2012. *Manifest rationality: A pragmatic theory of argument*. Routledge.
- [30] Philip N Johnson-Laird. 1999. Deductive reasoning. *Annual review of psychology* 50, 1 (1999), 109–135.
- [31] Jean Kaddour, Joshua Harris, Maximilian Mozes, Herbie Bradley, Roberta Raileanu, and Robert McHardy. 2023. Challenges and Applications of Large Language Models. *arXiv preprint arXiv:2307.10169* (2023).
- [32] Daniel Kahneman. 2011. *Thinking, fast and slow*. macmillan.
- [33] Katri Kleemola, Heidi Hytinen, and Auli Toom. 2022. The Challenge of Position-Taking in Novice Higher Education Students' Argumentative Writing. In *Frontiers in Education*.
- [34] Aaron Kozbelt, Ronald A Beghetto, and Mark A Runco. 2010. Theories of creativity. (2010).
- [35] Max Kreminski and Chris Martens. 2022. Unmet creativity support needs in computationally supported creative writing. In *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*. 74–82.
- [36] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.
- [37] John Lawrence and Chris Reed. 2020. Argument mining: A survey. *Computational Linguistics* 45, 4 (2020), 765–818.
- [38] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research methods in human-computer interaction*. Morgan Kaufmann.
- [39] Mina Lee, Percy Liang, and Qian Yang. 2022. Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [40] Yoonjoo Lee, Tae Soo Kim, Minsuk Chang, and Juho Kim. 2022. Interactive Children's Story Rewriting Through Parent-Children Interaction. In *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*. 62–71.
- [41] Toby Jia-Jun Li, Jingya Chen, Haijun Xia, Tom M. Mitchell, and Brad A. Myers. 2020. Multi-Modal Repairs of Conversational Breakdowns in Task-Oriented Dialogs. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST 2020)*. ACM. <https://doi.org/10.1145/3379337.3415820>
- [42] Toby Jia-Jun Li, Marissa Radensky, Justin Jia, Kirielle Singarajah, Tom M. Mitchell, and Brad A. Myers. 2019. PUMICE: A Multi-Modal Agent that Learns Concepts and Conditionals from Natural Language and Demonstrations. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST 2019)*. ACM. <https://doi.org/10.1145/3332165.3347899>
- [43] Eva Lindgren and Kirk PH Sullivan. 2006. Writing and the analysis of revision: An overview. In *Computer key-stroke logging and writing*. Brill, 31–44.
- [44] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *Comput. Surveys* 55, 9 (2023), 1–35.
- [45] Arthur O Lovejoy. 2011. *The great chain of being: A study of the history of an idea*. Transaction Publishers.
- [46] Astute Digital Solutions Ltd. 2023. Academic phrasebook - A guide for writing research papers. <https://www.ref-n-write.com/academic-phrases-handbook/>
- [47] Yuwen Lu, Chengzhi Zhang, Iris Zhang, and Toby Jia-Jun Li. 2022. Bridging the Gap Between UX Practitioners' Work Practices and AI-Enabled Design Support Tools. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI EA '22). Association for Computing Machinery, New York, NY, USA, Article 268, 7 pages. <https://doi.org/10.1145/3491101.3519809>
- [48] Andrea A Lunsford and John J Ruzskiewicz. 2016. *Everything's an Argument*. Bedford/St. Martin's.
- [49] Aran Lunzer and Kasper Hornbæk. 2008. Subjunctive Interfaces: Extending Applications to Support Parallel Setup, Viewing and Control of Alternative Scenarios. *ACM Trans. Comput.-Hum. Interact.* 14, 4, Article 17 (jan 2008), 44 pages. <https://doi.org/10.1145/1314683.1314685>
- [50] Anne Marguerite McAlister, Dennis M Lee, Katherine M Ehler, Rachel Louis Kajfez, Courtney June Faber, and Marian S Kennedy. 2017. Qualitative coding: An approach to assess inter-rater reliability. In *2017 ASEE annual conference & exposition*.
- [51] Felix Müller-Wienbergen, Oliver Müller, Stefan Seidel, and Jörg Becker. 2011. Leaving the beaten tracks in creative work—A design theory for systems that support convergent and divergent thinking. *Journal of the Association for Information Systems* 12, 11 (2011), 2.
- [52] Brad A Myers. 1986. Visual programming, programming by example, and program visualization: a taxonomy. *ACM sigchi bulletin* 17, 4 (1986), 59–66.
- [53] Emmanuel Sampath Nelson. 1998. *Theorizing composition: A critical sourcebook of theory and scholarship in contemporary composition studies*. Greenwood Publishing Group.
- [54] Bettina Pedemonte. 2007. How can the relationship between argumentation and proof be analysed? *Educational studies in mathematics* 66, 1 (2007), 23–41.
- [55] Christof Rapp. 2002. Aristotle's rhetoric. (2002).
- [56] Shaina Raza, Deepak John Reji, and Chen Ding. 2022. Dbias: detecting biases and ensuring fairness in news articles. *International Journal of Data Science and Analytics* (2022), 1–21.
- [57] Laria Reynolds and Kyle McDonell. 2021. Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7.
- [58] Ramon Ruiz-Dolz, Joaquin Taverner, Stella Heras, Ana Garcia-Fornes, and Vicente Botti. 2022. A Qualitative Analysis of the Persuasive Properties of Argumentation Schemes. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. 1–11.
- [59] Mark A Runco and Garrett J Jaeger. 2012. The standard definition of creativity. *Creativity research journal* 24, 1 (2012), 92–96.
- [60] Thomas P Ryan and JP Morgan. 2007. Modern experimental design. *Journal of Statistical Theory and Practice* 1, 3-4 (2007), 501–506.
- [61] Abulhair Saparov and He He. 2022. Language models are greedy reasoners: A systematic formal analysis of chain-of-thought. *arXiv preprint arXiv:2210.01240* (2022).
- [62] Baruch B Schwarz, Lauren B Resnick, and Michael J Baker. 2017. *Dialogue, argumentation and education: History, theory and practice*. Cambridge University Press.
- [63] Noah Shinn, Beck Labash, and Ashwin Gopinath. 2023. Reflexion: an autonomous agent with dynamic memory and self-reflection. *arXiv preprint arXiv:2303.11366* (2023).
- [64] Ben Shneiderman and Pattie Maes. 1997. Direct Manipulation vs. Interface Agents. *Interactions* 4, 6 (nov 1997), 42–61. <https://doi.org/10.1145/267505.267514>

- [65] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2022. Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence. *ACM Transactions on Computer-Human Interaction* (2022).
- [66] Hannah Snyder. 2019. Literature review as a research methodology: An overview and guidelines. *Journal of business research* 104 (2019), 333–339.
- [67] Sangho Suh, Jian Zhao, and Edith Law. 2022. Codetoon: Story ideation, auto comic generation, and structure mapping for code-driven storytelling. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–16.
- [68] Mirac Suzgun, Nathan Scales, Nathanael Schärli, Sebastian Gehrmann, Yi Tay, Hyung Won Chung, Aakanksha Chowdhery, Quoc V Le, Ed H Chi, Denny Zhou, et al. 2022. Challenging BIG-Bench tasks and whether chain-of-thought can solve them. *arXiv preprint arXiv:2210.09261* (2022).
- [69] Keith D Swenson. 1993. A visual language to describe collaborative work. In *Proceedings 1993 IEEE Symposium on Visual Languages*. IEEE, 298–303.
- [70] Frans H Van Eemeren, Rob Grootendorst, Ralph H Johnson, Christian Plantin, and Charles A Willard. 2013. *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Routledge.
- [71] Henning Wachsmuth, Nona Naderi, Yufang Hou, Yonatan Bilu, Vinodkumar Prabhakaran, Tim Alberdingk Thijm, Graeme Hirst, and Benno Stein. 2017. Computational argumentation quality assessment in natural language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. 176–187.
- [72] Thiemo Wambsgans, Tobias Kueng, Matthias Soellner, and Jan Marco Leimeister. 2021. ArgueTutor: An adaptive dialog-based learning system for argumentation skills. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–13.
- [73] Thiemo Wambsgans, Christina Niklaus, Matthias Cetto, Matthias Söllner, Siegfried Handschuh, and Jan Marco Leimeister. 2020. AL: An adaptive learning support system for argumentation skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [74] Dakuo Wang, Judith S Olson, Jingwen Zhang, Trung Nguyen, and Gary M Olson. 2015. DocuViz: visualizing collaborative writing. In *Proceedings of the 33rd Annual ACM conference on human factors in computing systems*. 1865–1874.
- [75] Dakuo Wang, Haodan Tan, and Tun Lu. 2017. Why users do not want to write together when they are writing together: Users' rationales for today's collaborative writing practices. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–18.
- [76] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. 2022. Chain of thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903* (2022).
- [77] Arnold Wentzel. 2017. *A guide to argumentative research writing and thinking: Overcoming challenges*. Routledge.
- [78] Kirsten N. Whitley. 1997. Visual programming languages and the empirical evidence for and against. *Journal of Visual Languages & Computing* 8, 1 (1997), 109–142.
- [79] Joseph M. Williams. 2012. *Three Modules on Clear Writing Style: An Introduction to the Craft of Argument*. SOAR: Student Observation and Research.
- [80] Marisol Wong-Villacres, Margarita Ortiz, Vanessa Echeverría, and Katherine Chiluiza. 2015. A tabletop system to promote argumentation in computer science students. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*. 325–330.
- [81] Meng Xia, Qian Zhu, Xingbo Wang, Fei Nie, Huamin Qu, and Xiaojuan Ma. 2022. Persua: A visual interactive system to enhance the persuasiveness of arguments in online discussion. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–30.
- [82] Daijin Yang, Yanpeng Zhou, Zhiyuan Zhang, Toby Jia-Jun Li, and Ray LC. 2022. AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In *Joint Proceedings of the ACM IUI Workshops*, Vol. 10.
- [83] Kevin Yang, Nanyun Peng, Yuandong Tian, and Dan Klein. 2022. Re3: Generating longer stories with recursive reprompting and revision. *arXiv preprint arXiv:2210.06774* (2022).
- [84] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L Griffiths, Yuan Cao, and Karthik Narasimhan. 2023. Tree of thoughts: Deliberate problem solving with large language models. *arXiv preprint arXiv:2305.10601* (2023).
- [85] Soobin Yim, Dakuo Wang, Judith Olson, Viet Vu, and Mark Warschauer. 2017. Synchronous collaborative writing in the classroom: undergraduates' collaboration practices and their impact on writing style, quality, and quantity. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 468–479.
- [86] Mark Young, Danielle Argiro, and Steven Kubica. 1995. Cantata: Visual programming environment for the Khoros system. *ACM SIGGRAPH Computer Graphics* 29, 2 (1995), 22–24.
- [87] Zheng Zhang, Ying Xu, Yanhao Wang, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, and Toby Jia-Jun Li. 2022. Storybuddy: A human-ai collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. In *Proceedings of the 2022 CHI Conference on Human*

	Topic	Task description
Topic 1	Governments should not fund any scientific research whose consequences are unclear	Write a response in which you discuss the extent to which you agree or disagree with the statement and explain your reasoning for the position you take. In developing and supporting your position, you should consider ways in which the statement might or might not hold true and explain how these considerations shape your position
Topic 2	People’s attitudes are determined more by their immediate situation or surroundings than by society as a whole	
Topic 3	Educational institutions should actively encourage their students to choose fields of study that will prepare them for lucrative careers	

Table 1: Argumentative topics used in our user study, adopted from the sample topics of GRE issue writing

Task	Prompt template	Few-shot examples
Key aspect generation	<ul style="list-style-type: none"> ● system role: A helpful writing assistant that aims to help writers come up with high level aspects or topics that they can think of to support their argument ● user input: Please list key aspects that are worth discussing to support the argument: [selected argument] 	<p>Key aspects of the argument “<i>We should advocate for the expansion of engineering course offerings in colleges</i>”:</p> <ul style="list-style-type: none"> ● Growing demand for engineers ● Diverse career opportunities ● Promoting interdisciplinary learning ● Fostering innovation and creativity ● Enhancing STEM education
Discussion point generation	<ul style="list-style-type: none"> ● system role: A helpful writing assistant that aims to come up with pertinent discussion points based on a specified aspect to reinforce the given argument ● user input: Please list key discussion points worth including in the discussion to support argument [selected argument] from the aspect of [selected aspect] 	<p>Discussion points of the argument “<i>We should advocate for the expansion of engineering course offerings in colleges</i>” from the perspective of “<i>Growing demand for engineers</i>”:</p> <ul style="list-style-type: none"> ● Recent increased job position number in Engineering ● National interest in key engineering area ● Varying engineering job types ● Labor scarcity in the engineering job market
Counter argument generation	<ul style="list-style-type: none"> ● system role: A helpful writing assistant specializing in argumentative essay tutoring and generating counterarguments for the given statement ● user input: Please list potential counterarguments that can challenge the argument [selected argument] 	<p>Counterarguments of the statement “<i>Some scholars and researchers claim that there are negative impacts of technology on a child’s developing mind. According to one research study, scholars claimed that “moderate evidence also suggests that early exposure to purely entertainment content, and media violence in particular, is negatively associated with cognitive skills and academic achievement”</i>”:</p> <ul style="list-style-type: none"> ● <i>Evidence of Positive Impacts:</i> While some studies suggest negative impacts of technology on children’s cognitive skills and academic achievement, there are also studies that demonstrate positive impacts, such as improved visual-spatial skills, problem-solving abilities, and creativity. ● <i>Importance of Parental Involvement:</i> The negative impacts of technology can be mitigated by parental involvement and guidance. Parents can set limits on screen time and select age-appropriate content, and monitor their children’s technology use to ensure they are engaging in positive and educational activities. ● <i>Individual Differences:</i> Not all children are affected the same way by technology, and the impact on their cognitive skills and academic achievement may depend on individual factors, such as age, gender, socioeconomic status, and learning style. ● <i>Importance of Context:</i> The negative impacts of technology on cognitive skills and academic achievement may be dependent on the context in which it is used. For example, technology use in the classroom may have different effects than technology use at home.

<p>Logical fallacy generation</p>	<ul style="list-style-type: none"> • system role: A helpful writing assistant focusing on argumentative essay tutoring and trying to suggest logical weaknesses in the given statement • user input: Please list potential logical weaknesses in the argument [selected argument] 	<p>Logical fallacies in the statement “<i>The seriousness of a punishment should match the seriousness of the crime. Right now, the punishment for drunk driving may simply be a fine. But drunk driving is a very serious crime that can kill innocent people. So the death penalty should be the punishment for drunk driving</i>”:</p> <ul style="list-style-type: none"> • <i>Slippery Slope Fallacy:</i> The argument assumes that the punishment for drunk driving should be escalated all the way to the death penalty, without considering other proportional punishments that could be implemented between a fine and the death penalty. This is a slippery slope fallacy. • <i>False Analogy Fallacy:</i> The argument equates drunk driving, which is a serious crime that can result in innocent deaths, with other crimes that are punishable by the death penalty, such as murder. This is a false analogy fallacy, as drunk driving and murder are not equivalent in terms of their severity, intent, or harm caused. <p>Logical fallacies in the statement “<i>I know the exam is graded based on performance, but you should give me an A. My cat has been sick, my car broke down, and I’ve had a cold, so it was really hard for me to study!</i>”:</p> <ul style="list-style-type: none"> • <i>Appeal to Pity Fallacy:</i> The argument attempts to persuade the grader by appealing to pity, by suggesting that the student’s unfortunate circumstances should override their actual performance on the exam. • <i>False Cause Fallacy:</i> The argument implies that the student’s poor performance on the exam is caused by their external circumstances, such as a sick cat or a broken car, without providing any evidence to support this claim. This is a false cause fallacy, as there may be other factors that contributed to the student’s poor performance, such as lack of preparation or understanding of the material. <p>Logical fallacies in the statement “<i>Caldwell Hall is in bad shape. Either we tear it down and put up a new building, or we continue to risk students’ safety. Obviously we shouldn’t risk anyone’s safety, so we must tear the building down</i>”:</p> <ul style="list-style-type: none"> • <i>False Dilemma Fallacy:</i> The argument presents a false dilemma by suggesting that there are only two options: tearing down Caldwell Hall and putting up a new building or risking students’ safety. This ignores the possibility of other solutions, such as renovating the existing building or relocating students to a different building. • <i>Slippery Slope Fallacy:</i> The argument assumes that if we do not tear down Caldwell Hall, then we are automatically risking students’ safety. This is a slippery slope fallacy, as there may be other ways to ensure student safety without tearing down the building, such as implementing safety measures or conducting regular inspections. • <i>Hasty Generalization Fallacy:</i> The argument assumes that the state of Caldwell Hall is representative of all buildings on campus, or that all old buildings are in bad shape and pose a risk to students. This is a hasty generalization fallacy, as the state of one building does not necessarily represent the state of all buildings on campus.
-----------------------------------	---	---

<p>Supporting evidence generation</p>	<ul style="list-style-type: none"> ● system role: A helpful writing assistant focusing on argumentative essay tutoring and trying to suggest supporting evidence types in the given statement ● user input: Please list potential supporting evidences that can back the argument: [selected argument]. You can think from the following aspects: sharing professional experience (ethos), arousing audience's emotion (pathos), providing facts and strict logical reasoning (logos) and presenting concrete practical examples (example) 	<p>Supporting evidence types in the statement “<i>Renewable energy has the potential to significantly benefit people’s lives in many ways. By reducing reliance on fossil fuels and transitioning to cleaner sources of energy, we can improve air quality and reduce the negative health effects associated with pollution. Additionally, renewable energy can create new job opportunities and boost local economies, particularly in rural areas where wind and solar energy projects can be developed</i>”:</p> <ul style="list-style-type: none"> ● <i>Statistical data (logos):</i> Data from credible sources such as the International Energy Agency and National Renewable Energy Laboratory can provide statistical evidence of the potential benefits of renewable energy, such as reductions in air pollution and increases in job creation and economic growth. ● <i>Expert opinion (ethos):</i> Opinions of experts, such as researchers and environmental scientists, can provide credibility to the argument that renewable energy can have significant benefits for people’s lives. ● <i>Case studies (example):</i> Examples of successful renewable energy projects, particularly in rural areas, can provide concrete evidence of the potential benefits of renewable energy. ● <i>Surveys and polls (logos):</i> Surveys and polls can provide evidence of public opinion and support for renewable energy, as well as demonstrate the potential for consumer demand for renewable energy products and services.
<p>Draft of key aspect</p>	<ul style="list-style-type: none"> ● system role: A helpful writing assistant focusing on argumentative essay tutoring. You are trying to write a starting sentence of the paragraph that support user’s argument from a particular perspective ● user input: Write a starting sentence for the paragraph that elaborates on the argument [selected argument] from the perspective of [key aspect] 	<p>N/A</p>
<p>Draft of discussion point</p>	<ul style="list-style-type: none"> ● system role: A helpful writing assistant focusing on argumentative essay tutoring. You are trying to elaborate on a particular given discussion point to support my argument. ● user input: Please write a paragraph that elaborates on my argument [selected argument] by considering the following discussion point [selected discussion point] 	<p>N/A</p>

<p>Draft of counter argument</p>	<ul style="list-style-type: none"> • system role: A helpful writing assistant focusing on argumentative essay tutoring. You are trying to argue against an argument by considering a provided counter argument • user input: Please write a paragraph that argues against the argument [selected argument] by considering the following counter argument [counter argument] from the perspective of [key aspect] 	<p>N/A</p>
<p>Draft of supporting evidence</p>	<ul style="list-style-type: none"> • system role: A helpful writing assistant focusing on argumentative essay tutoring. You are trying to support an argument by considering a provided supporting argument • user input: Please write a paragraph that supports the argument: [selected argument] by realizing the following kind of supporting evidence [supporting evidence type] 	<p>N/A</p>
<p>Settlement of logical fallacy</p>	<ul style="list-style-type: none"> • system role: A helpful writing assistant. You are trying to fix the mentioned logical weaknesses in my argument • user input: I just made an argument: [selected argument]. I know this argument has the following logical weaknesses: [logical fallacies]. Rewrite the argument to fix the logical weaknesses 	<p>N/A</p>

Table 2: Prompt templates used in goal recommendation, argumentative sparks and draft generation. According to the ChatGPT API specification, the "system role" determines the role that the model needs to assume during the current session, while the "user input" provides the model with the prompt that the user has inputted.

Topic2 [Attitudes and environment]:

People's attitudes are determined more by their immediate situation or surroundings than by society as a whole.

Write a response in which you discuss the extent to which you agree or disagree with the statement and explain your reasoning for the position you take. In developing and supporting your position, you should consider ways in which the statement might or might not hold true and explain how these considerations shape your position.

I disagree with the statement that people attitudes are determined more by their immediate situation or surroundings than by society as a whole. The two of these concepts go together, as what we are taught by society influences our reaction to immediate situations. For example, what you are taught from your parents and teachers when you are a child could determine how you handle situations you are in later in life. For example, I go to the store because I am hungry and need food. If I were to react based off of my attitudes of my immediate situation, I would be inclined to simply take the bag of chips and start eating because I am starving and want food. I have been taught by society that this is not right. The proper and appropriate thing to do would be to pick out the chips, pay for them, and then consume them. Therefore, society has taught me what is socially and morally acceptable and influences my decisions and attitudes more than my immediate situation. This can be applied to beliefs as well. Oftentimes, stereotypes that we are taught by society as children plays into our attitudes and beliefs as adults. This can have negative impacts, as seen in today's social climate. Regardless, what we are taught by society shapes our beliefs and attitudes more than our immediate situation. One could argue against this saying that as humans we have the capability to exercise free will and make our own decisions based off of the situation presented to us. In this view, what society has taught us does not matter because we can do whatever it is we want. For example, some people would go to the store and steal a bag of chips. Whether because they want to, or can't afford it. They know it is wrong. It is widely accepted by society that stealing is wrong. And yet some people exercise their free will to act off of their attitudes of the immediate situation: they are hungry, they steal the chips, their hunger is immediately satisfied. In their view, the benefit of stealing the chips was greater than the cost, therefore prompting their attitudes to be shaped around their immediate situation rather than the views of society. But they had to develop this view from somewhere, probably a lesson from society (although a more negative lesson). Therefore, regardless of a positive or negative attitude, our attitudes are developed based off of society as opposed to our immediate situation because human thought is developed from somewhere, most likely society.

Table 3: An example draft prototype generated by *plain text editor* (i.e., baseline) from a user study session

Topic1 [Scientific research funding]:

Governments should not fund any scientific research whose consequences are unclear.

Write a response in which you discuss your views on the policy and explain your reasoning for the position you take. In developing and supporting your position, you should consider the possible consequences of implementing the policy and explain how these consequences shape your position.

Governments should not fund any scientific research whose consequences are unclear. From an ethical standpoint, it is imperative that governments exercise caution when funding scientific research with uncertain consequences, as the potential risks to human life and the environment cannot be overlooked. Research findings are an essential part of scientific research, and they play a crucial role in shaping public policy and decision-making. However, it is important to recognize that research findings are not always clear-cut, and their interpretation can be subject to bias or misinterpretation. This is particularly true when dealing with scientific research that has uncertain consequences. In such cases, it is essential that governments exercise caution when interpreting research findings and consider the potential risks to human life and the environment. For instance, if a study finds that a particular chemical compound has potential therapeutic benefits, but also carries a significant risk of toxicity, the government must weigh the potential benefits against the risks before deciding to fund further research. Ultimately, it is the responsibility of the government to ensure that scientific research is conducted ethically and in a way that minimizes potential harm to both humans and the environment. The discussion point "New Node" is an important consideration when it comes to funding scientific research with uncertain consequences. The development of new technologies and scientific breakthroughs can bring about significant advancements for society, but they can also pose potential risks to human life and the environment. Therefore, it is crucial for governments to exercise caution when providing funding for such research. The potential risks must be carefully evaluated and weighed against the potential benefits. Additionally, ethical considerations must also be taken into account, as the consequences of scientific research can have far-reaching implications for future generations. Ultimately, it is the responsibility of governments to ensure that the funding of scientific research is conducted in a responsible and ethical manner, with a focus on minimizing potential risks and maximizing benefits. When it comes to scientific research, it is crucial to consider ethical principles before funding any project. Governments must exercise caution to ensure that the research aligns with ethical standards, especially when the consequences of the research are uncertain. Ethical principles require that the potential risks to human life and the environment are taken into account before any research is conducted. It is essential that the research does not harm human life, violate human rights, or negatively impact the environment. Governments must, therefore, ensure that the research is conducted with the utmost care and responsibility to prevent any adverse effects. In conclusion, ethical principles should be a top priority when funding scientific research, and the potential risks must be considered to ensure that the research does not have any negative consequences. Case studies provide compelling evidence for the argument that governments must exercise caution when funding scientific research with uncertain consequences. For instance, the case of thalidomide, a drug developed in the 1950s to alleviate morning sickness in pregnant women, highlights the disastrous consequences of a lack of caution in scientific research. Despite insufficient testing, the drug was widely distributed, resulting in thousands of birth defects and deaths. Similarly, the case of genetically modified organisms (GMOs) illustrates the potential risks to the environment and human health that can result from insufficient testing and regulation. These examples demonstrate the importance of government oversight and caution in funding scientific research, particularly in cases where the consequences are uncertain or potentially harmful.

Looking back at precedents of similar decisions in the past, it becomes clear that governments should not fund any scientific research whose consequences are unclear. When it comes to funding scientific research, ethical considerations should always be at the forefront. While looking back at precedents of similar decisions in the past can be helpful in making informed decisions, it is important to remember that science and technology are constantly evolving. This means that the consequences of research may not always be entirely clear, even if we have historical examples to draw from. Therefore, it is crucial that governments carefully consider the ethical implications of any research they choose to fund. This includes considering the potential risks and benefits of the research, as well as the potential impact on vulnerable populations or the environment. Ultimately, funding scientific research should be done with the goal of advancing knowledge and improving society, but never at the cost of ethical principles. The potential for unintended consequences of scientific research that is not fully understood is a crucial factor to consider when deciding whether or not to fund a particular research project. Looking back at precedents of similar decisions in the past, it is evident that many scientific discoveries have led to unforeseen negative consequences. For instance, the development of nuclear technology led to the creation of atomic bombs, which have caused widespread destruction and loss of life. Similarly, the widespread use of DDT as a pesticide led to environmental damage and health concerns. Therefore, it is essential to consider the potential risks and benefits of any scientific research before investing in it. Governments should prioritize funding research that has clear and well-understood consequences, rather than taking a gamble on research that could have unintended and harmful outcomes. When it comes to funding scientific research, it is crucial to consider the potential consequences. If the outcomes of the research are unclear, it is difficult to determine the potential benefits or harms that may result. In these cases, the negative consequences of funding research with unclear consequences can be significant. For example, funding research that could lead to the development of dangerous weapons or technologies could have devastating consequences. Additionally, funding research that could harm the environment or public health without fully understanding the risks could also have severe negative consequences. Therefore, it is important for governments to carefully evaluate the potential consequences of funding scientific research before committing resources to it. By looking back at precedents of similar decisions in the past, we can see the importance of making informed decisions based on the potential consequences of the research. **[continue in next page»]**

[»continue for last page] While it is important to consider precedents of similar decisions in the past, it is also important to consider the potential for innovation. The counter argument suggests that funding scientific research with unclear consequences can lead to innovation. Many scientific breakthroughs and advancements have come from research that was initially deemed uncertain or risky. If governments only funded research with clear and immediate results, they would miss out on potentially groundbreaking discoveries. Additionally, it is difficult to predict the long-term consequences of any scientific research, so refusing to fund research with unclear consequences could limit progress and innovation in the future. Therefore, it is important to weigh the potential risks and benefits of funding scientific research and not solely rely on precedent as a deciding factor.

From the perspective of potential risks of scientific research, it is imperative that governments exercise caution when funding research whose consequences are unclear. When it comes to scientific research, it is essential to consider the potential risks associated with it, especially the potential harm to human health or the environment. Governments must exercise caution when funding research whose consequences are unclear, as the consequences can be catastrophic. The potential risks associated with scientific research can be numerous, including the release of hazardous chemicals or the creation of new diseases. In such cases, the consequences can be devastating, and the government must take measures to ensure that the risks are minimized. Therefore, it is crucial to evaluate the benefits and potential risks of the research before funding it, as it can have far-reaching consequences for the health and well-being of the population and the environment. Ultimately, it is the responsibility of the government to ensure that the research is conducted safely and that the risks are minimized to the greatest extent possible. It is undeniable that scientific research has the potential to be misused by individuals or groups with malicious intent. This misuse could take many forms, from the development of new weapons to the creation of dangerous new viruses. Given this reality, it is essential that governments exercise caution when funding research whose consequences are unclear. In particular, they must carefully consider the potential risks associated with any research they choose to support, and take steps to mitigate those risks as much as possible. This may involve setting strict ethical and safety standards for researchers, or imposing regulations on the use of certain technologies. Ultimately, the goal should be to promote scientific progress while minimizing the potential for harm, and this requires a careful balancing of risks and benefits at all stages of the research process. The difficulty of predicting the long-term consequences of scientific research is a major concern when it comes to funding scientific research. The potential risks of scientific research cannot be fully understood until the research has been conducted and the consequences become apparent. This is especially true for research that involves new technologies or untested hypotheses. Governments must exercise caution when funding research whose consequences are unclear, as the long-term effects of such research may be unpredictable and potentially harmful. Therefore, it is crucial that governments take a proactive approach in assessing the potential risks of scientific research before funding it, and ensure that researchers are conducting their studies in a responsible and ethical manner. By doing so, governments can help to minimize potential risks and ensure that the benefits of scientific research outweigh the potential harms.

Table 4: An example draft prototype generated by *GPT Playground* from a user study session

Topic3 [Career-oriented education]:

Educational institutions should actively encourage their students to choose fields of study that will prepare them for lucrative careers.

Write a response in which you discuss the extent to which you agree or disagree with the claim. In developing and supporting your position, be sure to address the most compelling reasons and/or examples that could be used to challenge your position.

Educational institutions should actively encourage their students to choose fields of study that will prepare them for lucrative careers. Career development is a crucial aspect of a student's education, and educational institutions should actively encourage their students to choose fields of study that will prepare them for lucrative careers. By doing so, these institutions can help students achieve their financial goals and secure a stable future. Encouraging students to choose fields that offer high-paying jobs can also help address the issue of student debt. Students who graduate with high-paying jobs are better equipped to pay off their loans and avoid the financial burden that comes with student debt. Additionally, by promoting lucrative careers, educational institutions can contribute to the overall economic growth of a country. When students are trained in high-demand fields, they are more likely to find employment and contribute to the growth of various industries. Therefore, it is imperative that educational institutions prioritize career development and encourage students to consider fields that offer financial stability and growth opportunities.

One of the primary advantages of encouraging students to pursue fields of study that prepare them for lucrative careers is the economic benefits that students can enjoy throughout their lives. Encouraging students to pursue fields of study that prepare them for lucrative careers not only benefits the students themselves but also their families. With a higher income, students can provide better financial support for their families, which can lead to an improved quality of life. For example, they can afford better housing, healthcare, and education for their children. Additionally, students can also contribute to their communities by investing in local businesses and philanthropic endeavors. By pursuing a career that offers a higher salary, students can not only secure their own financial stability but also positively impact the lives of those around them. Therefore, it is essential to encourage students to consider the economic benefits of their chosen field of study, not only for their own future but also for their families and communities. When students pursue fields of study that prepare them for high-paying careers, they are more likely to be able to pay off their student loans and other debts in a timely manner. This is because they have the financial resources to do so, which can relieve a significant amount of stress and anxiety. The ability to pay off student loans and debt also allows students to invest in their futures by saving for retirement, purchasing a home, or starting a business. Furthermore, having a high-paying career can provide financial security and stability, which can lead to a better quality of life overall. Therefore, encouraging students to pursue fields of study that prepare them for lucrative careers can provide economic benefits that can last a lifetime. Higher earning potential and career growth opportunities are two significant benefits of pursuing fields of study that prepare students for lucrative careers. By investing in education and training in high-demand fields such as computer science, engineering, finance, or healthcare, students can increase their earning potential significantly. These fields offer not only high starting salaries but also numerous opportunities for career advancement and growth. Students who pursue these careers can expect to earn more throughout their lives, which can lead to a more comfortable lifestyle and financial stability. Furthermore, the skills and knowledge gained in these careers can open doors to leadership positions and entrepreneurship, which can lead to even greater financial rewards. In short, encouraging students to pursue fields of study that prepare them for lucrative careers can provide them with lifelong economic benefits and opportunities for professional growth. **[continue in next page»]**

[»continue for last page] When it comes to choosing a field of study, practicality should be a key consideration for students looking to maximize their educational investment. Career counseling services provided by educational institutions play a crucial role in helping students make practical and informed decisions when it comes to choosing a field of study. These services provide students with valuable information about the job market, industry trends, and the skills and qualifications required to succeed in various fields. By working with career counselors, students can assess their strengths and weaknesses, identify potential career paths that align with their interests and goals, and develop a plan to achieve their desired career outcomes. Such guidance helps students make informed decisions about their educational investments, ensuring that they choose a field of study that not only aligns with their interests but also offers them the best opportunities for career growth and financial stability. Therefore, students should take advantage of career counseling services provided by educational institutions to make practical and informed decisions when choosing their field of study. The availability of internships and job placement programs should be a crucial factor when students are selecting their field of study. In today's competitive job market, acquiring relevant work experience is just as important as earning a degree. By choosing a practical field of study that offers ample internship opportunities and job placement programs, students can gain hands-on experience and develop a network of professional contacts that can help them secure employment after graduation. Additionally, internships can provide students with a glimpse into the day-to-day operations of their chosen industry, allowing them to make more informed decisions about their career path. Therefore, choosing a practical field of study that offers real-world experience can be a smart investment for students seeking to maximize their educational investment. I completely agree with your argument that practicality should be a key consideration for students when choosing a field of study. However, it is not solely the responsibility of the students to ensure that their education aligns with the demands of the job market. Educational institutions also have a responsibility to equip their students with the skills and knowledge necessary to succeed in their chosen careers. This means that universities and colleges should offer programs that are tailored to the needs of the job market and provide opportunities for students to gain practical experience through internships, co-op programs, and other work-integrated learning opportunities. By doing so, educational institutions can help students maximize their educational investment and increase their chances of finding employment after graduation. Ultimately, both students and educational institutions must work together to ensure that graduates are well-prepared for the workforce and can make a positive impact on their chosen industries.

While it is true that the job market can be unpredictable, this is not a valid reason to shift the responsibility of practicality solely onto educational institutions. Students must also take responsibility for their own education and career prospects by researching job trends, networking with professionals in their desired field, and gaining relevant experience through internships or part-time jobs. Additionally, while the job market may be unpredictable, certain industries and fields consistently have high demand and job opportunities. Educational institutions should focus on providing programs and opportunities that align with these in-demand industries and equip students with the necessary skills to succeed in them. By solely relying on educational institutions to tailor their programs to the job market, students may miss out on valuable opportunities and fail to take proactive steps towards their own career success. Therefore, it is important for both students and educational institutions to work together and take responsibility for ensuring practicality in education.

So overall, our conclusion is educational institutions should actively encourage their students to choose fields of study that will prepare them for lucrative careers.

Table 5: An example draft prototype generated by VISAR from a user study session

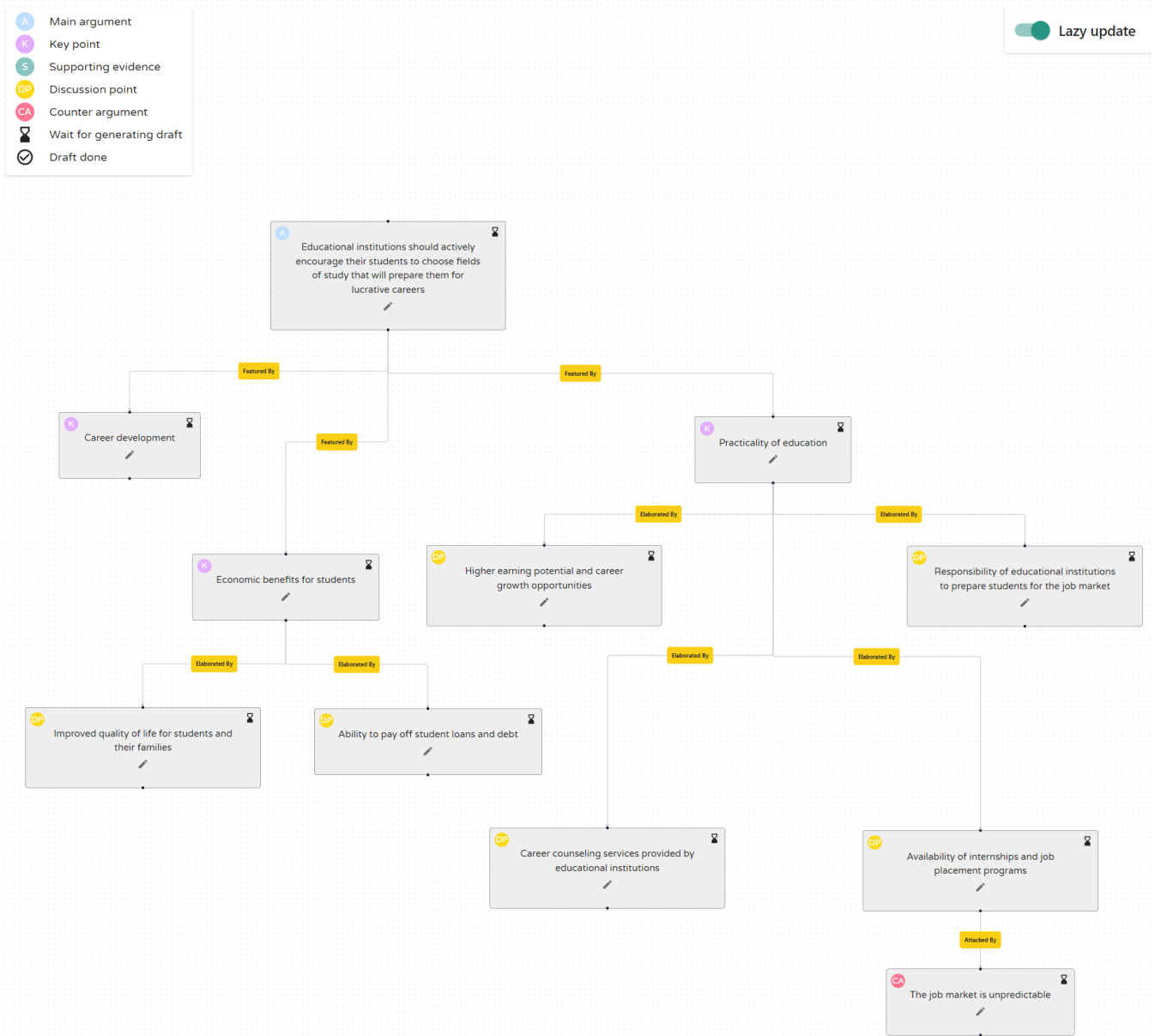


Figure 10: The visual outline of the example draft prototype shown in Table 3