



# Jamplate: Exploring LLM-Enhanced Templates for Idea Reflection

Xiaotong (Tone) Xu

xt@ucsd.edu

University of California, San Diego

La Jolla, California, USA

Jiayu Yin

jiayuy11@uci.edu

University of California, Irvine

Irvine, California, USA

Catherine Gu

catheg3@uci.edu

University of California, Irvine

Irvine, California, USA

Jenny Mar

jemar@ucsd.edu

University of California, San Diego

La Jolla, California, USA

Sydney Zhang

syz001@ucsd.edu

University of California, San Diego

La Jolla, California, USA

Jane L. E

je@ucsd.edu

University of California, San Diego

La Jolla, California, USA

Steven P. Dow

spdow@ucsd.edu

University of California, San Diego

La Jolla, California, USA

## ACM Reference Format:

Xiaotong (Tone) Xu, Jiayu Yin, Catherine Gu, Jenny Mar, Sydney Zhang, Jane L. E, and Steven P. Dow. 2024. Jamplate: Exploring LLM-Enhanced Templates for Idea Reflection. In *29th International Conference on Intelligent User Interfaces (IUI '24), March 18–21, 2024, Greenville, SC, USA*. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3640543.3645196>

## 1 INTRODUCTION

From generating creative writing [42], to producing never-before-seen imagery [56, 65], to helping people write code and perform data analysis, large language models (LLMs) have shown the potential to transform creative and complex work [30, 42, 43, 54, 62]. However, to date, the most common paradigm for interacting with LLMs involves text-based chat (i.e., ChatGPT[2], Bard[27], and Bing Chat[49]) which requires users to write specific prompts and organize the information that gets returned. The chat-based paradigm for LLMs creates its own set of challenges: prompting becomes its own skill that users must comprehend to generate appropriate responses from the LLM, and LLM outputs can be long-winded, unorganized, and error-prone.

For complex tasks, users often struggle with “prompt engineering”, the process of authoring text “prompts” that can be interpreted and understood by a generative AI model [25]. Writing effective LLM prompts can require significant trial-and-error and expert strategies, such as creating a chain of prompts that work together [69, 70, 72]. To overcome the challenges of chat-based LLMs, recent research has led to novel interfaces for assisting in the construction of effective prompts [14, 34, 69]. One strategy has been to integrate LLMs directly into users’ existing workspaces to provide in-situ assistance, such as within text editors [26, 42], coding IDEs [52, 62], and digital canvases [22, 51], to name a few. An implicit goal is to reduce the gap between a user’s work context and the information provided by these powerful models.

Beyond simply comprehending LLM outputs, creators need to engage in sensemaking and critical thinking to integrate the provided information into their creative work. LLMs are not only hard to prompt, their chat-based outputs are often long-winded and force users to read and scroll through paragraphs of text to find useful content [64]. To deal with lengthy chat responses, other research

## ABSTRACT

Advances in AI, particularly large language models (LLMs), can transform creative work. When developing a new idea, LLMs can help designers gather information, find competitors, and generate alternatives. However, LLM responses tend to be long-winded or contain inaccuracies, placing a burden on users to carefully synthesize information. In our formative studies with 52 students and five instructors, we find that novice designers typically lack guidance on how to compose prompts, reflect critically on LLM responses, and extract key information to help shape an idea. Building on these insights, we explore an alternative approach for interacting with LLMs, not via chat, but rather through structured templates. Collaborative design templates are a well-established strategy for helping novices think, organize information, and reflect on creative work. Developed as a digital whiteboard plugin, Jamplate integrates LLM capabilities into design templates, streamlining the collection and organization of user-generated content and LLM responses within the template structure. In a preliminary study with 8 novice designers, participants expressed that Jamplate’s reflective questions and in-situ guidance improved their ability to think critically and improve ideas more effectively. We discuss the potential of designing LLM-enhanced templates to instigate critical reflection.

## CCS CONCEPTS

• Human-centered computing → User studies; User interface design; Natural language interfaces.

## KEYWORDS

design process, design template, LLM interaction, large language model interaction,



This work is licensed under a [Creative Commons Attribution International 4.0 License](https://creativecommons.org/licenses/by/4.0/).

IUI '24, March 18–21, 2024, Greenville, SC, USA

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0508-3/24/03

<https://doi.org/10.1145/3640543.3645196>

has explored alternative paradigms to interact with LLMs, such as creating node-link diagrams [35], hierarchical representations [9, 64], or multi-modal interactions [41, 65].

Furthermore, LLMs are known to hallucinate [33]. While this ambiguity can serve as inspiration for creativity [14], designers also need information grounded in reality to effectively assess an idea's constraints, competition, and potential. Even if the accuracy of LLM responses improves, the results are not delivered in a way that induces meaningful reflection, especially for novices who have not yet had enough opportunities to exercise their creativity thinking muscles.

To better understand the challenges of using chat-based LLMs to investigate and iterate on ideas, we ran formative studies with  $n=5$  design instructors and  $n=52$  students. Instructors were interviewed about their perspectives on the use of LLMs in design education. They were generally supportive of students using LLMs as a learning tool, as long as students could demonstrate critical thinking around how they use the results. Students were asked to use ChatGPT to conduct background research on their group design project concepts in a class assignment. Students reported finding the information provided by LLMs useful for their ideation processes. However, through independent qualitative coding of students' reflection reports, we found little evidence of critical reflection regarding the LLM-generated information or meaningful application for evolving their ideas.

To build on these insights and address challenges with the text-chat paradigm of LLMs, we explored the potential impact of integrating LLMs into *design templates*. Design templates are a known strategy for organizing and reflecting on information related to a creative process [21, 50, 53]. Integrating LLMs into interactive design templates has several potential advantages: (1) templates provide context around the type of information the user is seeking, and therefore, can compose contextually specific LLM prompts without burdening the user; (2) long LLM outputs can be parsed and spatially arranged within templates, making the information easier to comprehend; and (3) the template's structure not only scaffolds how people should approach a problem, but it can also offer cues that help users reflect more deeply on the information relayed by LLMs [44, 60].

To explore these potential advantages, we created a prototype called Jamplate as a plugin for FigJam (a digital whiteboard canvas) to assist users in iterating on their ideas. Jamplate leverages the existing spatial arrangement of a template to improve both the accuracy of LLM prompts and the presentation of LLM responses. Additionally, within each cell of a template, users can request contextual guidance to help them reflect on the LLM responses, such as questioning the information's veracity or considering how it can shape their own ideas. After reflection, users are always free to edit the LLM-provided content.

To demonstrate LLM-infused templates, we selected two commonly used design templates: Five Whys (5Ys) and Competitive Analysis (CA). Both templates scaffold users to discover and organize information from the real world to shape an idea, giving us insights on how Jamplate can provide assistance across different stages of ideation. As an example, Jamplate's integration with the CA template allows users to request more competitors, add dimensions for analysis, and summarize key comparative details within

each cell of the matrix. In a preliminary study, we asked 8 design students to use Jamplate to work through a small design process using Jamplate's 5Ys and CA templates. The majority of participants found that LLM-generated reflective questions and in-situ guidance helped them think more critically and improve their ideas. From the preliminary study results, we discuss the implications for interfacing with LLMs and facilitating creative processes using interactive templates.

To summarize, we make the following contributions in this paper:

- Two **formative studies** aimed at understanding perspectives on LLM use in design education across instructors (through expert interviews) and students (through an ideation assignment leveraging LLMs).
- An LLM-enhanced interactive template plugin **prototype**—Jamplate—informed by the formative study insights.
- A **preliminary user evaluation** showing the potentials and limitations on using LLM-empowered interactive design templates.

## 2 RELATED WORK

### 2.1 Supporting Critical Reflection in Creativity

Creativity is one of the most salient traits that indicate human intelligence [40, 63]. It is often regarded as a personality trait, but can be broken down into processes [11, 12]. Schön [57] introduced the concept of "reflection-in-action" suggesting that creativity can be learned. In particular, he highlights the importance of regular reflection for professional designers to think critically as new information becomes available [57].

Fleck and Fitzpatrick [23] delineated different "levels of reflection" in their framework, suggesting techniques that support different levels of reflective behavior. Within this framework, the R1 reflection level—indicative of foundational reflective thinking—involves revisiting knowledge with explanations, and providing justifications or explanations for knowledge. The R2 reflection level: "Exploring Relationships", emphasizes discerning connections between pieces of knowledge, promoting reflection by the use of techniques enabling the "seeing of things from multiple perspectives." While R2 reflection aims at forging links and adopting alternative views, the R3 reflection level gravitates towards transformative reflection, modifying one's initial standpoint to incorporate newly explored perspectives. The ultimate reflective level, R4, refers to critical reflection about wider implications, entailing consideration of aspects beyond the immediate context, and incorporating wider world implications. These levels of reflection inspire our work on how to better facilitate designers to be reflective.

Critical thinking is defined as "reasonable reflective thinking" [20], and often times creativity and critical thinking are co-taught or inter-related [17, 59]. Critical thinking can be linked to an integral set of diverse skills—the ability to analyze facts, generate and organize ideas, defend opinions, make comparisons, draw inferences, evaluate arguments, solve problems [8, 40], as well as conduct reasonable reflective thinking focused on deciding what to believe or do [19, 31]. Critical thinking happens at the higher-order levels, such as analysis, synthesis, and evaluation in Bloom's taxonomy [16, 38]. These include being able to compare one's idea with

others', summarize lessons, and make decisions based on evidence [16].

Professional designers often participate in activities that help them with these types of critical thinking in the creative process, such as running Five Whys exercises to find out the root causes of a problem or conducting competitive analyses to refine a solution [24, 29]. There are even templates designed around these activities that aim at facilitating critical thinking for designers at various levels, including novices, to achieve these goals or perform these tasks in a more structured manner [21, 50]. In our study, we picked a Five Whys template and a Competitive Analysis template as they are representative of different idea development stages (navigating problem and solution space respectively according to the Double-Diamond Model) [11].

## 2.2 Leveraging Templates as Cognitive Scaffolds

Templates are one type of cognitive tool, defined as an instrument that supports learners' cognitive processes [66]. They can be used in any form, e.g., computers, intelligent agents, etc. [13, 39]. These tools can help novices think more independently or complete more cognitively-demanding tasks. Templates provide this cognitive support in two ways: (1) Breaking down the work into smaller, more manageable tasks. For example, showing the intermediate steps can help people formulate thoughts. Moreover, the templates also help break down otherwise unstructured information into a more structured representation (schema) [21, 50]. For example, planning for sub goals can help people stay organized better than working with an abstract goal or without clear guidance [15]. (2) Providing a spatial arrangement that better organizes the information. This is especially helpful for novices when initially thinking about a complex problem. [47, 48, 55]. On top of the information structure, these templates also have spatial layout and visual design so that the hierarchy and relationships between pieces of information can be clearly understood [21, 50]. Showing hierarchy and relationships of information is important in information design because it helps manage the cognitive load for users [55].

However, the static nature of templates can still present some challenges for novices. For example, the need to gather, chunk, and move large amounts of unstructured information matching the spatial hierarchy of templates can be quite tedious [45]. Recent work has helped support parts of this process, such as helping users digest unstructured information by organizing information into the right part [48]. We are interested in exploring whether integrating LLMs can enhance the existing templates and help users comprehend LLM responses more easily and critically.

## 2.3 Integrating LLMs into Templates

Many creativity support tools have incorporated the generative power of LLMs. For example, in the ideation space, Idea Machine uses generated texts to help expand, rewrite, and combine ideas [14]. Wordcraft generates text snippets for co-writing stories from Google's LLM LAMBDA and allows writers to request the writing with elaboration, replacement, or rewrite more specific feedback [71]. SAGA and CreativeBot support collaborative storytelling between multiple humans or children-robot using GPT-3 [18, 58].

PID	Department	Gender	PBL Teaching Experience
P1	Cognitive Science	M	3 years
P2	Cognitive/Computer Science	M	23 years
P3	Communication	F	5 times
P4	Computer Science	F	3 years
P5	Business	M	20 years

Table 1: We recruited five college design instructors from various department affiliations, all of whom have years of experience teaching project-based learning courses.

PromptMaker [34] allows non-ML experts, like designers, to prototype machine learning functionalities. Multimodal approaches have also been explored combining visuals with LLM-generated output to complete more complex tasks. PopBlends uses GPT-3 to find the most associated trait from pop culture to a product and blend concepts to create memes [68]. Stylette uses natural language to style webpages [37].

Experienced designers may have started embracing LLMs and integrating them into their creative process, but we know little about how instructors and students approach LLMs for design education. This motivates our formative studies that inform the design of our prototype where we embed LLM prompting and responses within the structural scaffolding of a template.

## 3 FORMATIVE STUDIES

To better understand how to best integrate LLMs more broadly into creative processes and design education, our formative studies aim to understand: (1) How do instructors think about the use of LLMs in their design classes? (2) How do students use LLMs (ChatGPT in particular) without much scaffolding to reflect on their creative ideas?

### 3.1 Method

We aimed to understand the potential and practices of embracing LLMs and integrate them into current design education both from the perspective of instructors and students. We first conducted an expert interview study with n=5 university professors in HCI/Design who have experience teaching project-based learning (PBL) courses to understand educators' perspective (Table 1). We interviewed instructors on what they believed to constitute great project ideas and their attitudes towards students' increasing use of LLMs in course assignments. The interviews lasted between 30 minutes to an hour. We then conducted a qualitative study with n=52 design students on their experience using ChatGPT to assist their brainstorming process in a project-based class. Students in a start-up course had a total of two weeks to complete an assignment, where they brainstormed ten start-up ideas, conducted web searches on them, selected three ideas to further explore with ChatGPT, and finally wrote a reflection report on their experience using ChatGPT to improve their ideas.

Using the reflexive thematic analysis method [6], we summarized the major topics of interests from instructors, how students prompted ChatGPT to research their brainstormed project ideas,

Category	Topic	Definition	Sample Prompt	#
Competition (298)	Novelty	Ask how novel is their idea and how to do to make their ideas more novel, best way to do something; something that is unique, never done before, new idea	“How novel is this idea?”	165
	Similarity	Ask for similar features, ideas, to get inspiration, looking at competitors	“Could you list any existing similar ideas to my idea?”	77
	Competitor	Look for other competitors or features they have, differences among them	“Who would be the biggest competitors of my idea?”	44
	Business	Look for business models, financials, management, marketing	“How should I advertise this application?”	12
Impact (127)	Impact	Identify the potential design impact	“How big will the impact be for my idea?”	60
	Usefulness	Ask for the best way to do something, or be helpful	“How useful is this idea?”	44
	Successful	Ask how successful their ideas are or how to do become more successful, encompassing novelty, feasible, usefulness	“How successful do you think this idea will be?”	23
Concern (110)	Feedback	Ask for opinion between two options, yes/no, is this good? Ask for verification, positives and negatives	“What would a potential user say about my idea and how likely is he or she going to be interested in this developed application if it becomes a startup?”	56
	Issue	Ask for problems with design or idea, problem	“Are there any possible issues you can think of?”	54
Ideation (105)	Idea Expansion	Ask ChatGPT to expand on their ideas, features, or building off their ideas, according to the instructions by participants	“What are some gaps of features in these apps that I could add to my idea?”	97
	Feature	Discuss details within one specific feature	“What are the most important features to consider for my idea?”	8
Stakeholder (95)	Potential User	Ask about potential users, includes users that are clearly stated or unknown	“Where to find potential users for my idea?”	75
	Need Analysis	Ask questions before ideation, focusing on finding the underlying issue	“What are some reasons people have trouble reading long passages on their phones?”	20
Fact (36)	Factual Information	Description of something, factual answer, or something one would get out of a traditional web search	“What do your eyes do when you space out when reading?”	36
Feasibility (33)	Feasibility	Within time, material, monetary, technological constraints	“Can I use AI to help develop sounds for music production?”	33

**Table 2: The research team came up with seven categories and 15 sub-categories to code 674 prompts generated by 52 students. A total of 804 codes were assigned.**

and analyzed the students' reflections on their experiences. We analyzed how students used ChatGPT from the prompts and what students described to be useful in their reflection reports. In total, n=52 students generated 674 prompts. Two coders first merged the prompts using overlapping keywords, then categorized similar ones into the same category (Table 2). This resulted in a final set of codes including 15 topics within 7 higher level categories. Two coders then independently coded the prompts, reaching an initial concordance rate of 81.1%, and later discussed the inconsistent coding to reach a full agreement. In total, we assigned 804 codes to 674 prompts.

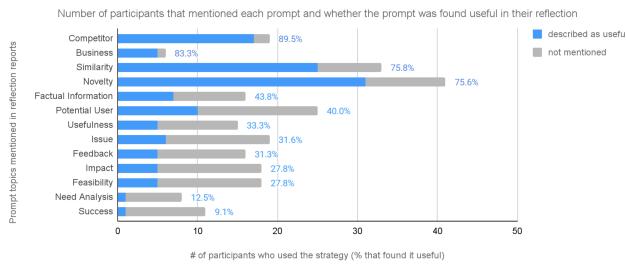
## 3.2 Findings

**3.2.1 LLMs were Useful for Learning New Information and Assessing Ideas.** In contrast to the idea that creativity is essential for novel ideas, in the pedagogical context, all professors mentioned that creativity was not as important as being able to think critically

about the problem. All professors also mentioned that often when students struggled to deliver meaningful projects, it was a result of lacking the necessary understanding of the root cause of the problem. Some potential reasons they mentioned included that students may be lacking in their understanding of background knowledge (P1), previous and existing ideas (P2), social responsibility (P3), and real user needs (P4, P5).

From analyzing students' chat history with ChatGPT for their idea iteration assignment, a considerable amount (298 out of 674) of prompts were coded as understanding the competitive advantages of students' ideas (“Competition” in Fig. 1). Students mostly used ChatGPT to ensure that their ideas were unique and asked for more information on how to improve their existing idea. The next most common type of code (127 out of 674) involved students asking for further information about the market space and how likely their ideas were to be successful as a start-up (“Impact” in Fig. 1).

Though these were common prompts, we wanted to further understand whether students actually found the responses useful. We analyzed students' reflection reports, and found that overall prompts related to understanding competitive advantages, such as those that included words like "competitor", "novelty", or "similarity", were indeed deemed more useful than all other categories combined (Fig. 1).



**Figure 1: Number of participants that mentioned each prompt (and the percentage who felt the strategy was useful) in their reflection report. The data rows are ranked by the percentages of people who felt the prompt strategy was useful.**

**3.2.2 The Importance (and Lack) of Critical Thinking About LLM-generated Content When Using LLMs in Learning.** During our interviews, all professors emphasized the importance of students demonstrating critical thinking when utilizing LLM-generated content in their process. Some showed more openness to adopting such tools and even designed assignments specifically instructing students to use AI tools (P2, P4). Others were most concerned about the accountability of students (P1, P3), or how over-reliance on these tools would impede the students' own learning (P5). But as long as students used AI in ways that benefited their learning (meaning not taking the information directly from the generated answer), the professors were open to them using these tools during assignments.

LLMs showed promising results when helping users reflect on their ideas. Most of the participants chose to expand on their ideas after interacting with ChatGPT, suggesting that ChatGPT gave them inspiration to further expand on their ideas, providing additional features that they thought were useful enough to be adopted into their final idea. Most of the ideas were improved (including "Expand Feature", "Position Idea", "Polish Language", "Elaborate Details", "Transform Idea") (Table 3) after using ChatGPT and only a few ideas (n=19) were preserved by the participants as their original ideas without any changes. We saw that ChatGPT generally affirmed or encouraged participants' original ideas, often through giving advice on how to improve or add to their idea, and almost never provided negative feedback or critique. This potentially influenced students' tendencies to keep the same direction of their original ideas and only make minor improvements instead of drastically changing the direction of their ideas.

Instead of iterating, a few students instead chose not to further improve upon their existing ideas—resulting in either making no changes (n=19), or completely changing directions (n=1). We looked specifically at those entries and compared the prompts and reflections of each student to see if there were any significant findings

that would cause this behavior. We saw that most of the participants who preserved their ideas received advice that they deemed useful from ChatGPT (as communicated in their reflection). Some potential reasons for these decisions to not incorporate the advice into their original ideas may be due to falsely believing their ideas were viable, or just lacking motivation on the assignment.

Even though ChatGPT responses influenced the ideas in these ways, few students evaluated the content critically explaining why is it good to integrate parts of the responses in their iterated ideas.

**3.2.3 The Limitations of LLMs for Supporting Critical Thinking.** We identified ten issues that students encountered when using ChatGPT for idea improvement (Table 4). Some of the problems encountered can be easily tackled by participants: We coded issues as "Uninformed" if ChatGPT lacked up-to-date information. In response, three participants mentioned the use of manual verification to confirm ChatGPT's response ("Verification") and two participants mentioned the use of human discretion to judge the results ("Discernment"). For the "Superficial" problem where ChatGPT generates generic and surface-level answers, four participants mentioned they asked more questions to refine ChatGPT's answers ("Refinement"), two participants asked more specific prompts to get more precise information ("Specificity"), and one chose to add forcing prompts to make ChatGPT respond appropriately ("Manipulation"). When faced with the situation of ChatGPT being "Indecisive", two participants chose to add prompts to force ChatGPT to make a decision ("Manipulation"), and one used human discernment to parse through the information themselves and make the decision on their own ("Discernment").

Other problems were harder for students to manually address without other resources: For the limitations of ChatGPT being "Repetitive" in its answers and giving "Impractical" advice, none of the student mentioned any strategies indicating that it was a problem that could be manually solved by users, but is related to the competence of ChatGPT. Similarly, for the limitation of giving "Irrelevant" and "Uncreative" information, only one person mentioned a potential strategy ("Specificity"), whereas most students were unable to generate strategies to better manage these aspects of the ChatGPT responses.

Idea Change	Definition	#
Expand Feature	Participant adds new features and ideas or refine existing features.	65
Polish Language	Participant refines and improves the language or presentation of their ideas.	52
Elaborate Details	Participant adds factual details, such as providing more background information or highlighting potential benefits of their ideas.	42
Target User	Participant focuses on a more concrete target participant or participant scenario for their ideas. It helps to define and contextualize the idea within specific participant contexts.	37
Implement Idea	Participant adds more realistic and logistical implementation details or actionable plans to consider when implementing their ideas.	24
Position Idea	Participant suggests marketing strategies or articulates the differences between their idea and existing ideas. It helps them position their idea in the market.	20
Preserve Idea	Participant decides to make little to no change to their idea.	19
Transform Idea	Participant drastically changes their idea.	1

**Table 3: Distribution of the types of idea changes that occurred**

Perceived Limitation	# of Participants Mentioned this Problem	% of Participants Mentioned this Problem	Strategies to Overcome Limitations	# of Participants who Utilized Strategies to Combat this Limitation
Impractical	2	3.9%	NA	0
Inaccurate	3	5.9%	Verification (1)	1
Biased	3	5.9%	Verification (1)	1
Indecisive	3	5.9%	Manipulation(2); Discernment (1)	3
Irrelevant	6	11.8%	Manipulation(1)	1
Repetitive	6	11.8%	NA	0
Uninformed	7	13.7%	Verification (3); Discernment (2)	5
Uncreative	8	15.7%	Specificity (1)	1
Superficial	9	17.3%	Refinement (4); Specificity (2); Manipulation(1)	7

Table 4: Mapping of each limitation mentioned in student reflection to the specific strategies mentioned to address the limitation, sorted by the percentage of students that mentioned the limitation.

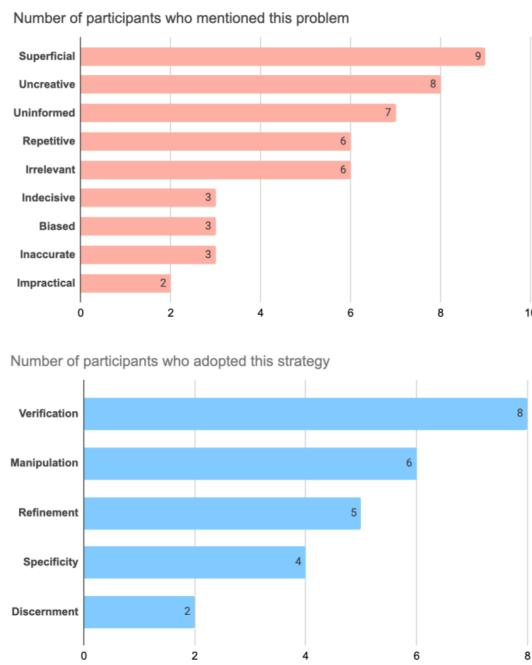


Figure 2: Problems (left) and strategies (right) students mentioned encountering and using in their reflection report, sorted by the number of participants that mentioned each element.

### 3.3 Design Takeaways

In summary, our formative studies suggested that ChatGPT was effective in providing background information and competitive advantages to users' ideas, but users did not show much critical thinking of their problems or the generated responses. We foresee the potential mutual benefits of integrating LLMs into learning templates for novices: (1) LLMs could make it easier for users to locate relevant information needed for a template (2) templates can hide away the messy details of prompt engineering and provide a better

way of presenting information than the typical chat interaction, and (3) LLM-enhanced templates could also instigate meta-cognitive behaviors by embedding questions that help learners reflect on the accuracy and value of information that goes beyond what static templates or ChatGPT can do alone.

To better understand these potential benefits, we wanted to explore how to do this integration for a small subset of representative templates. We purposefully looked for existing templates that encouraged users' critical thinking and enabled them to assess the competitive advantages of their own ideas. From the example template gallery across related design platforms [21, 50], we selected two common templates—"Five Whys (5Ys)" and "Competitive/Competitor Analysis (CA)"—as representative design probes to explore in-situ LLM assistance for reflective thinking. Both templates: (1) offer common frameworks for gathering and synthesizing external information from the world, (2) require critical thinking (e.g., comparing and contrasting information) in order to help people reflect and improve their own ideas, and (3) provide different spatial arrangements and levels of complexity which helps to illustrate the potential benefits of leveraging the templates' layouts to conduct precise prompt engineering and to present results in an organized manner. The 5Ys template encourages individuals to think critically and delve deep into the root causes of an issue and usually offers a linear approach. The CA template is two-dimensional and organizes data to more easily relate one's own idea to those of potential competitors.

Based on findings from the formative study (Section 3), we identified several design guidelines for designing our plugin prototype within the FigJam digital canvas:

- DG1 **Template Integration:** Leverage the existing layout of expert-designed templates as cognitive scaffolds to help users comprehend LLM responses easily
- DG2 **Precision Prompting:** Offload the prompt engineering task by more precisely prompting LLMs, based on the workflow status and the template's context.
- DG3 **Reflection Catalyst:** Provide a range of reflection opportunities within the workflow to assist novices in critically thinking about their ideas and the information in context.

## 4 JAMPLATE SYSTEM

Here we describe how the above design goals (Section 3.3) are translated into a feature framework for LLM-enhanced templates.

- (1) **Template Integration:** We leverage the existing template design and map the internal information structure of LLM responses to the template's information hierarchy, linking it to aspects such as the nodes' types, colors, and locations in the templates. This method prevents any single node or component from having overwhelmingly long text, thus enabling users to more easily digest shorter information snippets.
- (2) **Precision Prompting:** The prompts are tailored to the user's current step in the workflow, which is detected by the template component they are interacting with. Jemplate's design and prompts necessitate intermittent user interaction to initiate additional LLM response requests. This approach aims to segment prompt engineering tasks based on the template workflow, rather than requiring extensive information to fill all components at once.
- (3) **Reflection Catalyst:** We provide in-situ assistance to stimulate thought on incorporating elements of LLM responses into users' ideas, fostering reflection during this process. Furthermore, reflection questions embedded in the template support novices in critically examining their thoughts and the contextual information.

We applied these features to two widely used templates for ideation: Five Whys (5Ys) (Section 4.1) and Competitive Analysis (CA) (Section 4.2).

The research team first scraped<sup>1</sup> JSON objects from the platform's default 5Ys and CA template, labeled the key components and their parameters for simplicity, then mapped the JSON objects by type, color, and relative coordinates, deciding on the interaction associated with each based on their function in the workflow.

Jemplate was built using the FigJam Plugin API<sup>2</sup>, Firebase for storing data, Python Flask for a back-end server, and OpenAI's API to access a large language model (GPT-4) [1]. The system first sends requests from the FigJam interface to the Flask server, and then the server prompts the OpenAI API (sample prompts available at Appendix) and receives and processes responses before returning them to the Jemplate front end. All user interactions in the interface and the queries were recorded in the database. Jemplate also persistently monitors the users' changes to the template content and always uses up-to-date information when prompting the LLM.

### 4.1 Integrating LLMs into the Five Whys Template

The Five Whys (5Ys) template is a method for scaffolding a root cause analysis around a particular problem [24]. By repeatedly asking the question "Why" (typically five times), the nature of the problem as well as its solution becomes clearer. This iterative interrogative technique is designed to explore the cause-and-effect relationships underlying a particular problem. For example, the process begins with an initial problem for users to further explore

(e.g., Many people find it hard to meet their fitness goal.) Typically, the default question simply asks "Why?" and then the user provides an answer that largely depends on the user's existing knowledge and interpretation. Jemplate enhances this approach by offering reflective questions based on the users' responses, prompting them to think more deeply about the underlying causes of the problems (Fig. 3).

First, users input an initial problem into the top-left red box in the 5Ys template (Fig. 3a). Then, Jemplate analyzes the problem and presents a reflective question to the users in the first blue box (Fig. 3b). Users then respond to the question in the red box that follows (Fig. 3c). Once done, users can click the "Get Next GPT Prompt" button (Fig. 3d). This cycle repeats five times, leading users through guided inquiry and critical thinking. After their final response, Jemplate amalgamates all the user-provided thoughts to deduce the root cause of the initial problem (Fig. 3e). Users retain the autonomy to review, adapt, or reject LLM responses at each step in the process of analyzing the root cause.

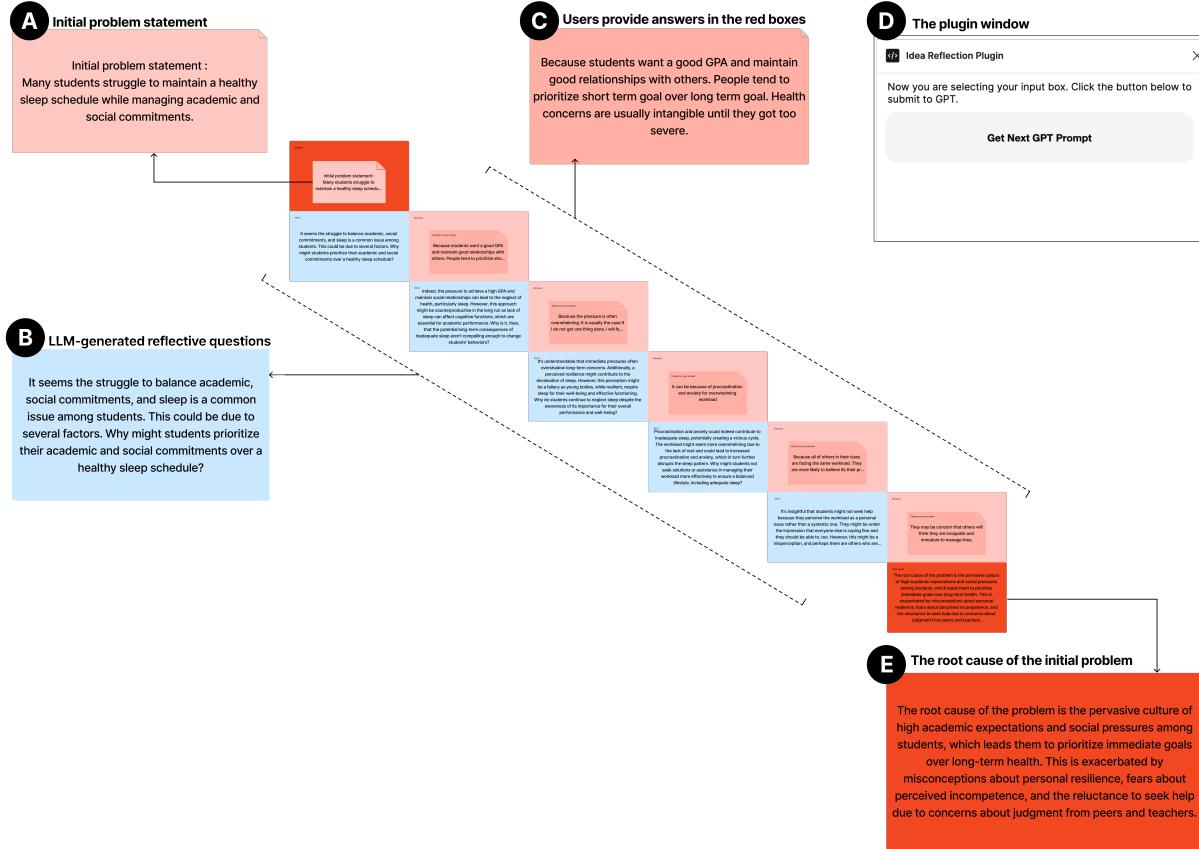
**4.1.1 Template Integration (DG1).** The layout of the 5Ys template showcases a one-directional flow in a staircase-like shape, beginning with the top-left dark red box stating the initial problem. From there, Jemplate fetches this input and poses a "Why" question in the subsequent blue box, prompting users to reflect and articulate their thoughts in the following light red box. This interaction, where the system offers reflective queries based on user input and then waits for their response, repeats five times, facilitating deeper exploration into the problem's underlying causes. After this iterative process, the system deduces the root cause of the original problem, presenting its conclusion in the final dark red box, based on both the initial problem and the five user responses. This spatial organization, beyond its visual appeal, offers an intuitive progression, guiding users through their reflective journey while consistently signposting their advancement.

**4.1.2 Precision Prompting (DG2).** Leveraging the template's structure and information enables more precise prompting from the LLMs. Thus, all the "Why" questions posed by the LLMs are derived from the context of the template, specifically users' input in the red boxes. Such LLM-generated questions spur users to reflect upon the initial problem while keeping all prior answers in consideration, broadening and deepening their exploration. For all content entered into a red box—whether it is the primary problem statement or subsequent answers (Fig. 3a,c)—LLMs generate a subsequent "Why" question response in the blue boxes. These responses: (1) summarize users' input while also pointing out potential shortcomings, offering reflections, examples, or alternative perspectives, and then (2) lead with a "Why" question, guiding users to think about the (other) factors leading to the problem (Fig. 3b). During this interaction, the LLM will avoid merely rephrasing users' inputs. It will ask 5 questions in total, waiting for users' responses after each one. After the user fields and responds to five such questions, the LLM concisely summarizes the root problem based on the five responses and the initial issue (Fig. 3e). The overall goal of these LLM responses is to foster critical thinking about the root cause of the primary problem.

<sup>1</sup><https://www.figma.com/developers/api#get-file-nodes-endpoint>

<sup>2</sup><https://www.figma.com/plugin-docs/api/api-reference/>

## Five Whys Template



**Figure 3: The 5Ys template includes (a) the top-left red box encompassing the initial problem statement, (b) the reflective questions in the blue boxes generated by LLMs by feeding in the content from above red boxes, (c) users type out a response to the LLM generated questions in the red boxes. Once done, users can click (d) the “Get Next GPT Prompt” button, which will send the content from all prior red boxes to the LLM API to generate another question in the next open blue box, and so on, repeating five times. After their final response, Jemplate amalgamates all the user-provided thoughts to deduce (e) a root cause of the initial problem, which the user can further edit, if desired.**

**4.1.3 Reflection Catalyst (DG3).** In the context of the 5Ys template, LLM reflection questions were crafted to facilitate the user’s critical thinking in relation to the initial problem. Inspired by Fleck and Fitzpatrick [23]’s work, we construct our reflection questions along the various reflection levels. Through analytical summaries of user responses—like highlighting potential shortcomings or suggesting alternate interpretations—the LLM questions steer users towards revisiting their prior insights critically (R1 level) and considering the problem from varied perspectives (R2 level). For instance, if a user answers a question about why students struggle to maintain healthy sleep schedules by saying, “*Because students want good GPAs and to maintain good relationships with others. People tend to prioritize short-term goals over long-term goals. Health concerns become tangible only when they’re severe,*” an LLM reflective question can respond with, “*While high GPA aspirations and social ties can overshadow sleep, neglecting sleep might hinder cognitive function,*

*crucial for academic success. Why then do potential long-term sleep deprivation repercussions fail to sway student behavior?*”

## 4.2 Integrating LLMs into the Competitive Analysis Template

Competitive analysis is a strategic method used to evaluate a new concept against the strengths and weaknesses of potential competitors within the market landscape. This analysis explores opportunities and threats, alongside insights on potential competitors. A traditional Competitor Analysis (CA) template gives an overview of a user’s new idea and compares it with a small number of competitors along a fixed set of comparative dimensions (e.g. the “Unique Value Proposition”, “Advantages”, and “Disadvantages”) (Fig. 4a). Jemplate supports users in comparing and contrasting their own ideas with competitors’ ideas by automatically populating each cell of the template with information gleaned from the LLM pipeline.

Moreover, Jamplate allows users to review and modify each cell and provides reflective questions to help users think critically about the information and how it relates to their idea. Users can click the “Add Competitor” button to request more competitors (Fig. 4b), click the “Add New Comparing Dimension” button to create more columns for comparison, (Fig. 4c), or click on any other sticky notes to see an embedded reflection question to help users consider the veracity and implications of the LLM-generated information on their own idea (Fig. 4d).

**4.2.1 Template Integration (DG1).** The CA template employs a structured spatial layout to systematically present and compare a user’s idea with potential competitors. The first column presents users’ ideas and brief descriptions of three competitor ideas, followed by columns detailing unique value propositions, advantages, and disadvantages of the user’s idea and the competitors (Fig. 4a).

Users can easily expand their analysis by adding new competitors at the bottom of the first column, or introducing comparing dimensions at the end of the first row.

**4.2.2 Precision Prompting (DG2).** Jamplate’s CA template integration employs precise contextual prompting by taking users’ brainstormed ideas as input, prompting the LLM to identify three relevant competitors. Jamplate subsequently integrates this information into a table-like template layout. The prompts ensure that the competitors generated by LLM are authentic and highly relevant to the user’s original idea (see Appendix). This eliminates the need for users to search for competitors manually or keep returning to ChatGPT for each part of the comparison analysis. The prompts are also customized based on the context of its position in the table, e.g., advantages, etc.

**4.2.3 Reflection Catalyst (DG3).** Jamplate’s floating plugin window (Fig. 5) provides contextual questions that aim to help users reflect

**Competitive Analysis Template**

**A** Content in template: overview, unique value, advantages, and disadvantages of one's ideas and competitors

**B** Reflection prompts in plugin

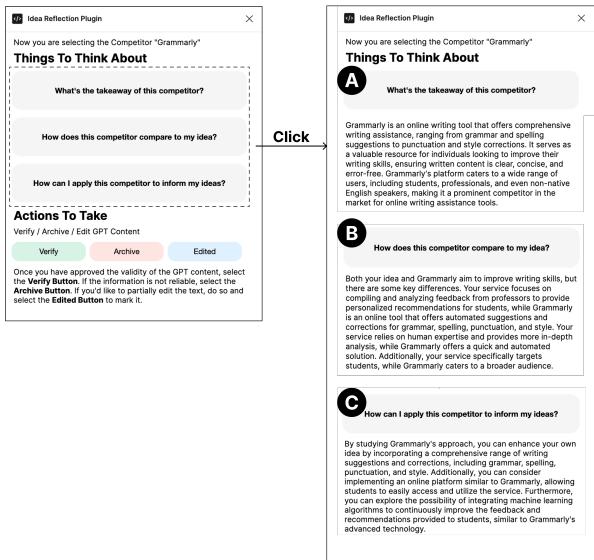
**C** Add New Dimension

**D** Click

**D** Click

**Figure 4:** The CA template includes: (a) information in the template table: overview, unique value, advantages, and disadvantages of one’s own ideas and three competitors, and enables users to request (b) more competitors, (c) add more dimensions for comparison analysis, and once once any sticky note is clicked, (d) pop up reflection prompts in Jamplate comprising summarizing insights in both horizontal and vertical directions, incorporating competitors’ concepts to refine one’s ideas, and verifying the LLM-generated content within the template.

### CA Template Plugin



**Figure 5: The reflection questions in the CA Template Plugin include: (a) What's the takeaway of this competitor/column?, (b) How does this (competitor) compare to my idea?, and (c) How can I apply this competitor to inform my ideas? Users can click on the question and see an LLM generated answer intended to help users reflect.**

on the veracity of LLM information and to reconsider their own idea when doing the competitive analysis.

Based on Fleck and Fitzpatrick [23]’s framework, Jamplate aims to help people to reflect on their knowledge with explanation and justification (R1 level), to “see more things” from multiple viewpoints (R2 level), and to transform their thoughts during the idea comparison process (R3 level). For example, the feature where users can add competitors helps people discover competing ideas (R2) and compare them to their own idea along several dimensions (R1).

While the traditional template only provides three comparative dimensions for contrasting one’s idea with competitors, Jamplate allows users to add dimensions, supporting Fleck’s notion of seeing new perspectives (R2) [23]. Users can add these comparative dimensions as they discover new information and develop their competitive analysis. Jamplate first prompts the LLM to ask for other ways to compare the competitors (e.g., privacy setting, cost, UI design) based on the attributes of their idea and the competitors. Users can select one or more of the new dimensions and this automatically adds new columns and fills in information for each competitor.

Furthermore, Jamplate offers a feature where users can click anywhere on the template to see an embedded reflection question to help them refine and enhance their own ideas. For example, clicking on a competitor in the matrix might provide the following reflection: “How does (the competitor) compare to your idea?”, or clicking on a comparative dimension asks users to reflect on “How

does your idea distinguish itself in terms of (this dimension)?” Upon selecting these questions, LLM-generated responses are promptly displayed in the Jamplate plugin window (Fig. 5b,c). These reflective, in-situ questions aim to enable users to juxtapose their ideas with competitors on specific dimensions in an intuitive manner. This facilitates both seeing more things (R2) and, possibly, transforming one’s ideas (R3). Ultimately, Jamplate aims to help users improve their original idea after completing the CA exercise.

## 5 USER STUDY

To delve into the potential advantages of spatial arrangement and metacognitive scaffolding in conjunction with contextual LLM assistance, we conducted a qualitative user study (n=8). This study aims to evaluate the efficacy and usability of Jamplate, our in-situ LLM plugin, in aiding idea reflection in the context of ideation, and to understand the participants’ experience and feedback for improvement. We were interested in how the contextual LLM-driven scaffolding influences user engagement and reflection in these templates and to ascertain how users integrate the LLM responses into their critical thinking processes. We hypothesized that the in-situ guidance offered by Jamplate, particularly when integrated into the templates, would be considered valuable by users, fostering a deeper, more reflective ideation experience.

### 5.1 Participants

We recruited a group of 8 participants (20 to 22 years old; 6 females) from varying backgrounds and varying experience with both design thinking as well as design tools. Regarding design thinking experience, 4 of them had never previously engaged in any design thinking activities, 3 had a little exposure, and 1 was comfortable with the process. As for the tool familiarity, 4 were unaware of Figma/FigJam, with 2 having used it minimally, 1 having used it many times, and 1 having only seen it being used.

### 5.2 Study Protocol

The user study sessions were conducted remotely via Zoom. Participants were introduced to the background and purpose of the study first, and they were asked to complete a demographic survey and read the consent form. After the researcher obtained their consent to this study, participants were introduced to the basic usage of FigJam, a digital whiteboard tool. Before delving into the tasks, the researcher briefly introduced the 5Ys and CA exercises, guided participants through Jamplate’s features, and then engaged them in specific tasks to gauge their experiences.

**5.2.1 Introduction and Familiarization with FigJam (~10 minutes).** Upon completion of the survey, participants were asked about their familiarity with digital whiteboard tools like FigJam. Those unfamiliar with these platforms were given a brief tutorial of FigJam introducing them to basic functionalities such as creating sticky notes, organizing the information in sticky notes, and navigating around and zooming in and out of the canvas. For those who were already familiar with FigJam, we quickly walked them through these basic functionalities as a refresher.

**5.2.2 5Ys Exercise (~20 minutes).** Participants were introduced to the 5Ys exercise and its purpose of uncovering the root cause of

problems. Participants were presented with a choice of three problem statements and completing the 5Ys exercise in Jamplate. The exercise prompted participants to critically think about their chosen problem, guided by the LLM-populated reflective questions. The goal was for the participant to trace back to the root cause of a given problem, culminating in their own synthesized conclusion on it.

**5.2.3 CA Exercise (~30 minutes).** Following the 5Ys exercise, participants had a deeper understanding of the problem and the cause. They were asked to brainstorm and write down solutions for their dissected problem. They were then introduced to the CA exercise. They were presented with a CA template and instructed to utilize it to compare their solution with LLM-populated competitors. Jamplate's functionalities were highlighted, and participants were encouraged to utilize Jamplate's reflective questions for a richer analysis. Through guided tasks, they were asked to identify the similarities between their ideas and competitors', learnings from the CA, opportunities they could use to improve their ideas, and finally, write down the improved version of their initial ideas. The end goal for them was to refine their initial brainstormed solutions using the insights gained from the comparative analysis.

**5.2.4 Post-Survey and Interview (~30 minutes).** Upon completion of the tasks, participants were directed to fill out feedback surveys on their experiences with both the 5Ys and CA templates using Jamplate compared to how they would normally perform without it. These surveys comprised of questions evaluating the quality of system responses, the influence of the tool on participants' critical thinking processes, the cognitive load of the task, the quality of their new problem statement (or new solution ideas), and the usability of Jamplate [3, 10, 28, 36]. This was followed by an in-depth interview, where participants were probed about their overall experiences, their thought processes, the impact of system-provided questions, the influence of spatial layout, and any challenges they encountered. This provided an understanding of their interaction with the tools and gathered insights into the nuances of their ideation process. The feedback collected from the survey and the interview was pivotal in assessing the tool's potential and identifying areas for future enhancement.

### 5.3 Study Results

Overall, participants reported that Jamplate elicited critical thinking and helped them develop their ideas (Fig. 6). Participants were able to reflect on their ideas throughout the process.

**5.3.1 Jamplate Elicits Critical Thinking and Reflection.** Jamplate was reported to both help narrow down and expand thinking. P6 and P8 expressed that Jamplate's reflections helped them think more deeply about the proposed problem. When P6 faced the problem of addressing students' unhealthy sleep schedules ("Many students struggle to maintain a healthy sleep schedule while managing academic and social commitments."), they mentioned how Jamplate guided their thinking to be more in-depth.

*"I actually think it's very useful because it's always step by step, and also break[s] down the problem in a deeper way. It goes from a superficial level to a deeper level: from the beginning, it's just academic and social*

commitment, and then it goes to the psychological aspect (about personal resilience, and fears about perceived incompetence)." —P6

P5 mentioned that Jamplate broke down the problem framing process and helped them think more comprehensively and concretely about the problem.

*"It's kind of a narrowing down process, and breaking down the question into different aspects. For the first one or two 'Whys', I'm just putting down my overall thinking about the question. But then, after reading the prompt, I was able to think more about the internal versus external factors, and more specific, like motivations and prioritizing, those kind of (factors)." —P5*

P5's assigned initial problem was: "Students constantly struggle to manage their time effectively and keep up with coursework," and the LLM responses inspired them to consider the external pressure (e.g., responsibilities), and internal factors (e.g., individual time management skills) of the initial problem.

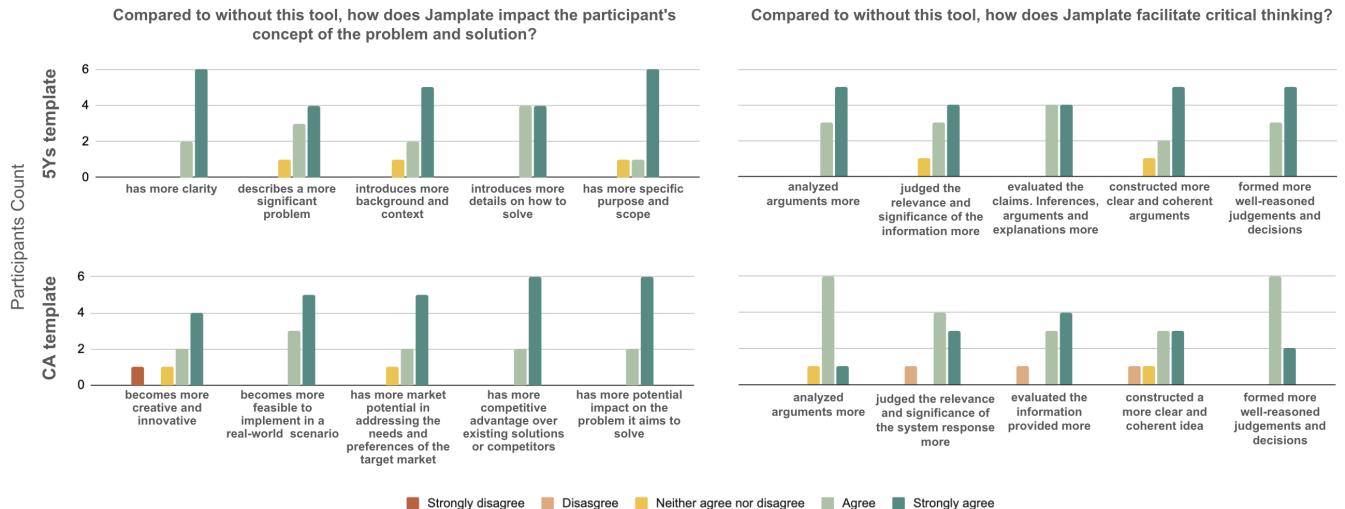
**5.3.2 Reflection Encouraged by Jamplate Helps Develop Users' Ideas.** Jamplate was consistently reported as building on and elaborating users' ideas, rather than creating new conversations. Several (3) participants mentioned that the questions guided them to think more concretely about their present idea. For example, P2's initial answer to the problem of students' unhealthy sleep schedule was students lacking discipline. Then Jamplate asked "Why did students lack discipline?" and P2 perceived it as a kind of affirmation. Upon answering with "having poor exercise and exposure to blue light can affect sleep quality," Jamplate asked the question—"Why do students have that?"—again which made them be concrete about their answers.

*"Those questions (generated from LLMs) were reflected and elaborated on my responses. It built off my response and it's forcing me to get specific in regards to the problem. The more general the answer, the more specific it wants the question to be." —P2*

P1 described that it helped to confirm their idea and used specific words that they did not think of (but close to what they described), enabling them to reframe their thoughts.

*"[The prompted question from Jamplate] used specific words in the questions that I didn't use in my previous responses. It helps me think about specific aspects of the problem and address it and reframe the solution." —P1*

When P1 answered the initial problem "Students constantly struggle to manage their time effectively and keep up with coursework", they had an answer that "(students) have a lot of commitments on their plates and may procrastinate." The LLM responded "Why might students have so many commitments that they're finding it difficult to prioritize their coursework?", the phrasing "prioritize their coursework" inspired P1 and guided them to think about these factors. This both helped make their previously rather vague idea of "procrastinating" more concrete, and also gave them additional directions to consider.



**Figure 6: Participants' survey responses after experiencing Jamplate for both completing the 5Ys and CA exercises.** These graphs show participants' responses to questions relating to two groups: (left) their evaluation of the quality of their final problem (in 5Ys, top) and solution (in CA, bottom), and (right) how well Jamplate facilitated their critical thinking during these ideation stages.

Similarly, P3 also reported that they would have pursued a different angle, but the system's responses pointed out an idea that they had not initially considered and in fact was rather dissimilar from what they had proposed in the previous conversations. P3 was also assigned to the problem around students struggling to maintain healthy sleep schedules.

*"I would have gone towards a different path with this whole initial issue of sleep. I think I would have gone into time management, responsibility and other stuff. It helped me think about the things that I wasn't initially thinking about, like the school policies (such as scheduling of exams or classes). It helped me think about that and how that could be an impact on the initial problem of sleep management." —P3*

Most participants also expressed that Jamplate saved them time in gathering information in areas such as the takeaway and disadvantages of other competitors. This allowed participants to focus their time on thinking more deeply about their ideas and to consider how they would use any additional information to elaborate on their idea.

P5 mentioned how they used the information to think deeply and elaborate on their ideas when doing competitive analysis.

*"There are two factors that actually help me to analyze and think about the question. The first one is the competitor's advantages because I feel like this part helped me to get the key competitiveness of each competitor and help[ed] me to think about how to integrate their ideas into my own idea. And then the second factor was the 'main takeaway' provided by the plugin (in-situ reflective questions) because I feel like that information help me to get the gist of each*

*competitor and help me to think about, how can I solve this problem from different perspectives." —P5*

P5's initial idea was about helping individuals find the best lifestyle and teaching them necessary time management skills. Inspired from the competitors' advantages, P5 refined their idea by incorporating features like visualizing the time usage for every task, offering a reward system for staying focused, and managing time into their final idea.

**5.3.3 Jamplate's Spatial Layout Helps Users Structure Process and Relate Ideas.** We saw evidence that the spatial layout of Jamplate implies the hierarchy and relationships between pieces of information, did in fact provide unique advantages over solely chat-based LLMs.

The chatbot interaction paradigm is not only linear but also open-ended without a clear expectation of the upcoming conversation structure. In contrast, several (3) participants pointed out that the boxes and colors in Jamplate templates suggested where to fill-in and what to expect for the whole interaction experience. P4 mentioned that "the structure (a red box after a blue box) keeps going and is *really clear about which box I should answer*". Given the interaction in the 5Ys template is rather linear, we expected the spatial layout to have relatively minimal impact (among template options). However, participants expressed otherwise, as P1 suggests, the staircase layout in the template visually symbolized the answer going deeper.

*"Instead of just a straight line with different questions on the same level, it gave me a visual indication or a sign that the questions getting deeper as I going down the steps." —P1*

For the CA template, spatial layout is even more important because the templates contain more information and dimensions than

the 5Ys templates. The structure of the CA templates easily broke down long paragraphs information. All participants appreciated the grid layout and how they were able to easily compare, contrast, and stay organized because of this format. Many (5) participants mentioned that the organization of the template layout was useful for structuring how they thought about advantages and disadvantages of existing solutions and how these can be applied to their own ideas.

“I liked it because it was pretty [visually] organized. You can see different rows and columns. The information was all very concise, which I like. You can see the different bullet points so that you can get the critical information very easily, and then just compare the information with each other ” —P5

**5.3.4 Jamplate Limitations.** The current status of Jamplate only supports building on preset templates. P8 brought up that they would like to adjust the template so the interaction can be more customized rather than restricting to only five question exchanges in the 5Ys template. P4 and P7 all suggested that there might be an indefinite number of answers, and the template should be more flexible rather than honing in on a singular, fixed number of answers. The information presented in Jamplate plugin for CA template sometimes were expressed to be repetitive, and users desired more unique information and tailored summary on how to leverage that information.

## 6 DISCUSSION AND FUTURE WORK

Jamplate is an LLM-empowered FigJam plugin aimed at enhancing design templates on collaborative digital whiteboards. Jamplate utilizes the internal structure of the design templates as cognitive and contextual scaffolds to help template designers prompt LLMs, designate presentations of LLM responses, and enable template users to grasp the content and build on LLM responses more easily. Based on our study results, we discuss the potential of LLM-empowered dynamic templates in the following several directions.

### 6.1 Scaffolding Interactions with LLMs to Instigate Reflection

Reflection is critical in the creative process, especially for iterating on a design [4, 5, 23, 57]. We designed Jamplate to provide better scaffolding around interactions with LLMs. Information is laid out spatially, and users can click to see reflective questions in order to interrogate each piece of information. Jamplate intentionally structures the workflow to prevent users from requesting information on all template components at one time, only allowing the user to proceed step by step. In this way, it nudges users to think independently, be explicit about their intentions, and articulate the current status of their thoughts.

Jamplate embeds a variety of reflection opportunities that can be invoked in a number of ways, supporting users to reflect on their own inputs, as well as the LLM responses. For example, Jamplate asks reflective questions after summarizing what users have described as a problem when the intention is finding out the root cause for the user. Jamplate also provokes further reflection through creating questions on how to apply specific content to users’ own work,

encouraging users to mark how they processed the information, and showcasing several examples on how to instigate reflection through the interactions with LLMs. We envision many other directions worth exploring in this large design space. Future work may consider designing for reflection based on the type of content (users’ own ideas versus LLM responses), the timing and granularity (immediate review of information versus a delayed but more meta review of the creative process), to name but two.

### 6.2 Adapting to Diverse Templates and Workflows

Jamplate has three main features—(1) precise prompting based on the internal information structure of the template, (2) presenting information integrated within the spatial organization and visual representation of the template, and (3) embedding hints to encourage reflection throughout the ideation process. Our initial insights suggest that these patterns can be useful for a wide range of templates and workflow. Previous work has attempted to automatically gain insights from the spatial arrangement of information [32, 46, 61]. Besides spatial arrangement, users can manipulate many more parameters in digital whiteboards, such as the components’ order, shape, color, typeface, etc.

Future work might look to capture the information hierarchy from a static template and map it to a potential set of interactive components systematically to implement precise prompting, information presentation, corresponding reflection cues with a diverse set of templates, or workflows. However, careful design may be needed to customize the integration for each template. Key challenges include understanding enough of the work context in order to compose useful LLM prompts, and mapping out where particular information should be laid out spatially. The designer should carefully consider which aspects are best facilitated by LLMs to best encourage reflection and boost productivity.

### 6.3 Exploring Different Modalities and Representations

We have demonstrated Jamplate focused on text-based interactions for a single individual working within a design template. However, we see potential to further explore additional modalities such as audio or imagery [41], which will likely be part of future GPTs, like Gemini [65]. Further flexibility in modalities would be especially interesting to explore in the context of collaborations across synchronous and asynchronous teams.

Jamplate currently focuses on text-based input, as it is the most common way that designers communicate information asynchronously. However, we could also imagine instances, such as workshops, where a team of designers could collaborate verbally on digital whiteboard. Previous studies have extended the effort of transcribing designers’ conversations and show relevant information on a physical screen [67]. With the uprising of remote meetings and improving the accuracy of speech-to-text technologies, there have also been an increase in tools for capturing meeting conversations [7], which currently focus on summarizing text. We see a potential to output artifacts with richer structure and media, such as templates for cognitive exercises and group activities on a collaborative digital canvas. Expanding beyond templates, we see

opportunities to explore further structures like moodboards, storyboards, prototypes, etc. which can embed rich media of modalities ranging from voice to images, videos, hyperlinks, etc.

## 6.4 Evaluating the Efficacy of Jamplate Compared to Alternatives

Existing digital canvas platforms like FigJam and Miro are actively working on their own versions of embedded AI assistance [22, 51]. Their tools offer functionalities like synthesizing, brainstorming, and expanding on sticky notes. However, a common limitation of several features is that they do not combine the existing features of digital artifacts beyond text content on the online canvas. Thus, this makes them not too different from separately using a static template, with a chat-based LLM platforms on the side, and manually integrating content between platforms. An example of simple combination is the organization of sticky notes by their higher-level topic, which has been supported by existing AI assistance [51]. Jamplate leverages the internal information structure from the design template, which is an abundant resources in the online design community, and enables more dynamic template interactions. We propose that future designs of LLM applications should leverage the domain expertise of existing resources more, and overcome the limitations of chat-based text interactions.

## 7 CONCLUSION

In summary, we introduced a system named Jamplate that showcases in-situ assistance in two enhanced templates—Five Whys and Competitive Analysis—for ideators to think critically about design problems and solutions in a digital canvas environment (FigJam). The choices of these templates and design features of Jamplate is motivated by interviews with five design PBL university professors and a qualitative study with fifty university design students. We found the usefulness of using ChatGPT in thinking about ideas and the need for critical thinking of LLM responses in education. Through an exploratory study with eight users, Jamplate was reportedly favorable and effective in helping users iterate ideas. We discussed several future study opportunities to leverage templates as cognitive scaffolds and LLMs as generative engines to instigate reflective thinking in the creative process.

## ACKNOWLEDGMENTS

We would like to thank the Design Lab faculty members for being interviewed, and students and instructional team for the introductory design class at UCSD. We would like to thank David Min for initiating the early prototype for this project. We would also like to thank our anonymous reviewers for their constructive and insightful reviews.

## REFERENCES

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).
- [2] Open AI. 2022. ChatGPT. <https://chat.openai.com/>
- [3] Aaron Bangor, Philip T Kortum, and James T Miller. 2008. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction* 24, 6 (2008), 574–594.
- [4] Eric PS Baumer. 2015. Reflective informatics: conceptual dimensions for designing technologies of reflection. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 585–594.
- [5] Marit Bentvelzen, Paweł W. Woźniak, Pia S.F. Herbes, Evropi Stefanidi, and Jasmin Niess. 2022. Revisiting Reflection in HCI: Four Design Resources for Technologies That Support Reflection. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 1, Article 2 (mar 2022), 27 pages.
- [6] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis*. American Psychological Association.
- [7] Jenna Butler, Sonia Jaffe, Nancy Baym, Mary Czerwinski, Shamsi Iqbal, Kate Nowak, Sean Rintel, Abigail Sellen, Mihaela Vorvoreanu, Najeeb G. Abdulhamid, Judith Amores, Reid Andersen, Kagonya Awori, Maxamed Ahmed, danah boyd, James Brand, Georg Buscher, Dean Carignan, Martin Chan, Adam Coleman, Scott Counts, Madeleine Daep, Adam Fourney, Daniel G. Goldstein, Andy Gordon, Aaron L Hafkner, Javier Hernandez, Jake Hofman, Jenny Lay-Flurrie, Vera Liao, Siân Lindley, Sathish Manivannan, Charlton McIlwain, Subigya Nepal, Jennifer Neville, Stephanie Nyairo, Jacki O'Neill, Victor Poznanski, Gonzalo Ramos, Nagu Rangan, Lacey Rosedale, David Rothschild, Tara Safavi, Advait Sarkar, Ava Scott, Chirag Shah, Neha Parikh Shah, Teny Shapiro, Ryland Shaw, Auste Simkute, Jina Suh, Siddharth Suri, Ioana Tanase, Lev Tankelievitch, Mengting Wan, Ryen W. White, Longqi Yang, Brent Hecht, and Jaime Teevan. 2023. *Microsoft New Future of Work Report 2023*. Technical Report MSR-TR-2023-34. Microsoft. <https://www.microsoft.com/en-us/research/publication/microsoft-new-future-of-work-report-2023/>
- [8] Paul Chance. 1986. *Thinking in the Classroom: A Survey of Programs*. ERIC.
- [9] Kuo-En Chang, Yao-Ting Sung, and Ine-Dai Chen. 2002. The effect of concept mapping to enhance text comprehension and summarization. *The Journal of Experimental Education* 71, 1 (2002), 5–23.
- [10] Robert G Cooper and Elko J Kleinschmidt. 1987. New products: what separates winners from losers? *Journal of product innovation management* 4, 3 (1987), 169–184.
- [11] Design Council. [n. d.]. The Double Diamond. ([n. d.]). <https://www.designcouncil.org.uk/our-resources/the-double-diamond/>
- [12] Arthur Cropley. 2016. The myths of heaven-sent creativity: Toward a perhaps less democratic but more down-to-earth understanding. *Creativity Research Journal* 28, 3 (2016), 238–246.
- [13] Ton de Jong. 2006. Scaffolds for Scientific Discovery. In *Handling complexity in learning environments: Theory and research*, Jan Elen and Richard E Clark (Eds.). Emerald Group Publishing, Chapter 6, 107–128.
- [14] Giulia Di Fede, Davide Rocchesso, Steven P Dow, and Salvatore Andolina. 2022. The Idea Machine: LLM-based Expansion, Rewriting, Combination, and Suggestion of Ideas. In *Proceedings of the 14th Conference on Creativity and Cognition*. 623–627.
- [15] Melissa C Duffy and Roger Azevedo. 2015. Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior* 52 (2015), 338–348.
- [16] Robert Duron, Barbara Limbach, and Wendy Waugh. 2006. Critical thinking framework for any discipline. *International Journal of teaching and learning in higher education* 17, 2 (2006), 160–166.
- [17] Fabian Eggers, Kathi J Lovelace, and Frederik Kraft. 2017. Fostering creativity through critical thinking: The case of business start-up simulations. *Creativity and Innovation Management* 26, 3 (2017), 266–276.
- [18] Maha Elgarf, Sabha Zojaji, Gabriel Skantze, and Christopher Peters. 2022. CreativeBot: a Creative Storyteller robot to stimulate creativity in children. In *Proceedings of the 2022 International Conference on Multimodal Interaction*. 540–548.
- [19] Catherine D Ennis, Juanita Ross, and Ang Chen. 1992. The role of value orientations in curricular decision making: A rationale for teachers' goals and expectations. *Research quarterly for exercise and sport* 63, 1 (1992), 38–47.
- [20] Robert Ennis. 2011. Critical thinking: Reflection and perspective Part II. *Inquiry: Critical thinking across the Disciplines* 26, 2 (2011), 5–19.
- [21] Figma. 2023. Figma template gallery. <https://www.figma.com/templates/> [“Online; accessed 9-October-2023”].
- [22] Figma. 2023. Jambot. <https://www.figma.com/community/widget/127448146448630971> [“Online; accessed 9-October-2023”].
- [23] Rowanne Fleck and Geraldine Fitzpatrick. 2010. Reflecting on Reflection: Framing a Design Landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction* (Brisbane, Australia) (OZCHI '10). Association for Computing Machinery, New York, NY, USA, 216–223.
- [24] Wikimedia Foundation. 2023. Five whys. [https://en.wikipedia.org/wiki/Five\\_whys](https://en.wikipedia.org/wiki/Five_whys)
- [25] Wikimedia Foundation. 2024. Prompt Engineering. [https://en.wikipedia.org/wiki/Prompt\\_engineering](https://en.wikipedia.org/wiki/Prompt_engineering)
- [26] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for science writing using language models. In *Designing interactive systems conference*. 1002–1019.
- [27] Google. 2023. Bard. <https://bard.google.com/chat>

[28] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[29] Jerry Hausman, Gregory Leonard, and J Douglas Zona. 1994. Competitive analysis with differentiated products. *Annales d'Économie et de Statistique* (1994), 159–180.

[30] Arthur Huang, Ying Chao, Efrén de la Mora Velasco, Anil Bilgihan, and Wei Wei. 2022. When artificial intelligence meets the hospitality and tourism industry: an assessment framework to inform theory and management. *Journal of Hospitality and Tourism Insights* 5, 5 (2022), 1080–1100.

[31] William Huitt. 1998. Critical thinking: An overview. *Educational psychology interactive* 3, 6 (1998), 34–50.

[32] Ajit Jain, Andruid Kerne, Nic Lupfer, Gabriel Brittain, Aaron Perrine, Yoonsuck Choe, John Keyser, and Ruihong Huang. 2021. Recognizing creative visual design: multiscale design characteristics in free-form web curation documents. In *Proceedings of the 21st ACM Symposium on Document Engineering*, 1–10.

[33] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *Comput. Surveys* 55, 12 (2023), 1–38.

[34] Ellen Jiang, Kristen Olson, Edwin Toh, Alejandra Molina, Aaron Donsbach, Michael Terry, and Carrie J Cai. 2022. Promptmaker: Prompt-based prototyping with large language models. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–8.

[35] Peiling Jiang, Jude Rayan, Steven P Dow, and Haijun Xia. 2023. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. *arXiv preprint arXiv:2305.11473* (2023).

[36] David Kember, Doris YP Leung, Alice Jones, Alice Yuen Loke, Jan McKay, Kit Sinclair, Harrison Tse, Celia Webb, Frances Kam Yuet Wong, Marian Wong, et al. 2000. Development of a questionnaire to measure the level of reflective thinking. *Assessment & evaluation in higher education* 25, 4 (2000), 381–395.

[37] Tae Soo Kim, DaEun Choi, Yoonseo Choi, and Juho Kim. 2022. Stylette: Styling the web with natural language. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

[38] David R Krathwohl. 2002. A revision of Bloom's taxonomy: An overview. *Theory into practice* 41, 4 (2002), 212–218.

[39] Susanne P Lajoie and Sharon J Derry. 2013. *Computers as cognitive tools*. Routledge.

[40] Joe YF Lau. 2011. *An introduction to critical thinking and creativity: Think more, think better*. John Wiley & Sons.

[41] Susan Lin, Jeremy Warner, JD Zamfirescu-Pereira, Matthew G Lee, Sauhard Jain, Michael Xuelin Huang, Piyawat Lertvittayakumjorn, Shanning Cai, Shumin Zhai, Björn Hartmann, et al. 2024. Rambler: Supporting Writing With Speech via LLM-Assisted Gist Manipulation. *arXiv preprint arXiv:2401.10838* (2024).

[42] Ken Liu. 2023. Sudowrite. <https://www.sudowrite.com/> [Online; accessed 9-October-2023].

[43] Michael Xieyang Liu, Advait Sarkar, Carina Negreanu, Benjamin Zorn, Jack Williams, Neil Toronto, and Andrew D Gordon. 2023. “What It Wants Me To Say”: Bridging the Abstraction Gap Between End-User Programmers and Code-Generating Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–31.

[44] Jennifer A Livingston. 2003. Metacognition: An Overview. (2003).

[45] Andrés Lucero. 2015. Using affinity diagrams to evaluate interactive prototypes. In *Human-Computer Interaction—INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14–18, 2015, Proceedings, Part II* 15. Springer, 231–248.

[46] Nic Lupfer, Andruid Kerne, Rhema Linder, Hannah Fowler, Vijay Rajanna, Matthew Carrasco, and Alyssa Valdez. 2019. Multiscale Design Curation: Supporting Computer Science Students' Iterative and Reflective Creative Processes. In *Proceedings of the 2019 Conference on Creativity and Cognition*. 233–245.

[47] Stephen MacNeil, Zijian Ding, Kexin Quan, Thomas j Parashos, Yajie Sun, and Steven P Dow. 2021. Framing Creative Work: Helping Novices Frame Better Problems through Interactive Scaffolding. In *Creativity and Cognition*. 1–10.

[48] Stephen MacNeil, Ziheng Huang, Kenneth Chen, Zijian Ding, Alexander Yu, Kendall Nakai, and Steven P Dow. 2023. Freeform Templates: Combining Freeform Curation with Structured Templates. In *Proceedings of the 15th Conference on Creativity and Cognition*. 478–488.

[49] Microsoft. 2023. Bing Chat. <https://www.bing.com/chat>

[50] Miro. [n. d.]. Miro Templates Library. <https://miro.com/templates/> [Online; accessed 9-October-2023].

[51] Miro. 2023. Miro AI. <https://miro.com/ai/> [Online; accessed 9-October-2023].

[52] Nhan Nguyen and Sarah Nadi. 2022. An empirical evaluation of GitHub copilot's code suggestions. In *Proceedings of the 19th International Conference on Mining Software Repositories*. 1–5.

[53] Notion. 2024. Notion Templates. <https://www.notion.so/templates> [Online; accessed 27-January-2024].

[54] OpenAI. 2023. ChatGPT plugins. <https://openai.com/blog/chatgpt-plugins> [Online; accessed 9-October-2023].

[55] Fred Paas, Alexander Renkl, and John Sweller. 2004. Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional science* 32, 1/2 (2004), 1–8.

[56] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-shot text-to-image generation. In *International Conference on Machine Learning*. PMLR, 8821–8831.

[57] Donald A Schön. 1987. *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.

[58] Hanieh Shakeri, Carman Neustaedter, and Steve DiPaola. 2021. SAGA: Collaborative Storytelling with GPT-3. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing (Virtual Event, USA) (CSCW '21)*. Association for Computing Machinery, New York, NY, USA, 163–166.

[59] Barry L Shoop and Eugene K Ressler. 2011. Developing the critical thinking, creativity and innovation of undergraduate engineering students. *International Journal of Engineering Education* 27, 5 (2011), 1072.

[60] Ira Shor. 1980. *Critical teaching and everyday life*. Black Rose Books Limited.

[61] Pao Siangliuue, Joel Chan, Steven P Dow, and Krzysztof Z Gajos. 2016. IdeaHound: improving large-scale collaborative ideation with crowd-powered real-time semantic modeling. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 609–624.

[62] Dominik Sobania, Martin Briesch, and Franz Rothlauf. 2022. Choose your programming copilot: A comparison of the program synthesis performance of github copilot and genetic programming. In *Proceedings of the genetic and evolutionary computation conference*. 1019–1027.

[63] Robert J Sternberg. 2019. A theory of adaptive intelligence and its relation to general intelligence. *Journal of Intelligence* 7, 4 (2019), 23.

[64] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. *arXiv preprint arXiv:2305.11483* (2023).

[65] Gemini Team, Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805* (2023).

[66] Wouter Van Joolingen. 1998. Cognitive tools for discovery learning. *International journal of artificial intelligence in education* 10 (1998), 385–397.

[67] Hao-Chuan Wang, Dan Cosley, and Susan R Fussell. 2010. Idea expander: supporting group brainstorming with conversationally triggered visual thinking stimuli. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. 103–106.

[68] Sitong Wang, Savvas Petridis, Taeahn Kwon, Xiaojuan Ma, and Lydia B Chilton. 2023. PopBlends: Strategies for conceptual blending with large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[69] Tongshuang Wu, Ellen Jiang, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, and Carrie J Cai. 2022. Promptchainer: Chaining large language model prompts through visual programming. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–10.

[70] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–22.

[71] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *27th International Conference on Intelligent User Interfaces (Helsinki, Finland) (IUI '22)*. Association for Computing Machinery, New York, NY, USA, 841–852.

[72] JD Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.